НАЦИОНАЛЬНАЯ АКАДЕМИЯ НАУК УКРАИНЫ
ИНСТИТУТ ПРОБЛЕМ РЕГИСТРАЦИИ ИНФОРМАЦИИ
НАН УКРАИНЫ

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ И БЕЗОПАСНОСТЬ

МАТЕРИАЛЫ XVII МЕЖДУНАРОДНОЙ НАУЧНО-ПРАКТИЧЕСКОЙ КОНФЕРЕНЦИИ

ВЫПУСК 17

Киев – 2017

Рекомендовано к печати Ученым советом
Института проблем регистрации информации НАН Украины
(протокол № 3 от 5 декабря 2017 г.)

Информационные технологии и безопасность. Материалы XVII Международной научно-практической конференции ИТБ-2017. – К.: ООО "Инжиниринг", 2017. – 292 с. ISBN 978-966-2344-59-2

В сборник вошли материалы докладов, представленных на XVI Международной научно-практической конференции «Информационные технологии и безопасность» (ИТБ-2017, 30 ноября 2017 года, г. Киев, Украина).

В сборнике представлены статьи, посвященные вопросам кибернетической безопасности критических инфраструктур, моделированию и противодействию информационным операциям, технологиям информационно-аналитических исследований на основе открытых источников информации, онтологическому подходу, семантическим сетям, сценарному анализу при обеспечении информационной поддержки принятия решений, компьютерному моделированию процессов и систем, актуальным проблемам технологического и правового обеспечения информационной и кибернетической безопасности.

Для специалистов в области информационных технологий, информационной безопасности, информационного права а также для аспирантов и студентов старших курсов высшей школы соответствующих специальностей.

ISBN 978-966-2344-59-2

ОРГАНІЗАЦІЯ УПРАВЛІННЯ ГРУПОЮ МОБІЛЬНИХ ТЕХНІЧНИХ ОБ'ЄКТІВ

О.Г. Додонов, О.С. Горбачик, М.Г. Кузнєцова
Інститут проблем реєстрації інформації Національної академії наук України, м. Київ, Україна
dodonov@ipri.kiev.ua, ges@ipri.kiev.ua, margle@ipri.kiev.ua

Вступ. Проблема управління множиною мобільних технічних об'єктів, які мають спільно виконати певне завдання, є актуальною у багатьох сферах сучасного життя. Так, ця проблема виникає у практиці дослідження великих об'єктів, територій, робототехніці тощо. У технічній сфері будь-яка система, яка складається з окремих вузлів (наприклад, група процесорів у багатопроцесорній обчислювальній системі) може розглядатись як об'єкт групового управління (управління колективом).

Зрозуміло, що управління колективом інтелектуальних технічних об'єктів з метою виконання ними певного завдання, з одного боку, потребує розробки методів і алгоритмів управління взаємодією окремих інтелектуальних засобів для досягнення загальної цілі «колективу», а з іншого боку – розробки методів і засобів реалізації цих взаємодій засобами колективу у реальному часі і з урахуванням змін, що відбуваються у середовищі їх функціонування.

Нажаль немає достатньо узагальнених підходів і методик розв'язання задач управління колективом інтелектуальних технічних засобів з автономною системою пересування й навігації, які оснащені «органами» для реалізації конкретних завдань, функціонують в непередбачуваному недетермінованому середовищі, зокрема в умовах короткочасних уражуючих впливів.

Особливості колективного управління. Практика вирішення задач управління множиною технічних об'єктів свідчить, що у разі управління рухом групи рух окремого об'єкта групи неважливий, необхідно визначити характеристики руху усієї сукупності об'єктів, що утворюють складну просторово-часову структуру. У разі ж управління рухом об'єкта у складі групи стають важливими характеристики руху окремого об'єкта і його поведінки і взаємодії із іншими об'єктами групи. Управління спрямоване на забезпечення виконання загальносистемної цілі множиною технічних об'єктів.

Група технічних об’єктів з колективним управлінням може бути представлена як система

$$\Omega = \{O, E, G\},$$

де O – множина (група) технічних об’єктів, що утворюють відповідну просторово-часову структуру при функціонуванні системи,

E – система каналів обміну інформацією,

G – загальносистемна ціль функціонування.

У найбільш загальному вигляді окремий технічний об’єкт групи може бути поданий:

$$O_i = \langle E_{ij}, G_i, A_i, C_i, B_i \rangle,$$

де E_{ij} – матриця, що описує зв’язки об’єкта з іншими об’єктами групи,

G_i – множина цілей об’єкта,

A_i – множина характеристик автономності, самоорганізації об’єкта,

C_i – множина стратегій поведінки,

B_i – база знань i -го об’єкта.

Загальносистемну ціль, у більшості прикладних задач, визначає певний об’єкт (структура) більш високого рівня. Окремий об’єкт групи може пересуватись у просторі, виконувати обмін інформацією з іншими об’єктами групи щодо формування цілі, змін в умовах функціонування і поведінці. Непередбачуваність середовища функціонування і наявність уражуючих впливів потребує наявності у окремих об’єктів з групи (в залежності від ступеню інтелектуальності) змоги самостійно формувати цілі на основі наявної бази знань окремого об’єкта і аналізу ним поточної інформації від інших об’єктів групи і з середовища.

Приналежність до групи вимагає здатності до узгодження своєї поведінки з поведінкою інших об’єктів групи завдяки обміну інформацією з іншими об’єктами групи, зокрема, щодо визначення повноважень, оповіщення про можливості дій, стан середовища, поточний стан інших об’єктів з групи, інформування про виконання чи неможливість виконання свого завдання тощо.

У разі колективного управління якість взаємодії та обміну інформацією характеризується орієнтованістю, селективністю, інтенсивністю, динамічністю, інформативністю та стійкістю взаємодії об’єктів групи.

Стратегії колективного управління. Розрізняють три принципових підходи до вибудови стратегій колективного управління. *Перший підхід* передбачає централізоване формування стратегії – управління виконується деяким «центром» (стаціонарним чи мобільним), що отримує і опрацьовує всю необхідну інформацію.

Другий підхід – процес прийняття рішення й управління переноситься всередину групи і покладається на самі об'єкти – колективне формування стратегії. Між об'єктами управління відбувається активний обмін поточною інформацією та знаннями з баз даних об'єктів, а також результатами аналізу ситуації засобами й алгоритмами, що задіяні в окремих об'єктах. Складність реалізації цього підходу пов'язана з об'єктивними конфліктними ситуаціями, що безумовно повинні виникати і виникають при колективному прийнятті рішень.

Основою розв'язання об'єктивних конфліктних ситуацій при колективному управлінні може стати єдина платформа знань у всіх об'єктів групи, гарантуючи, що всі роблять однакові логічні висновки із схожих передумов.

Третій підхід – кожен об'єкт групи приймає рішення самостійно, обмінюючись інформацією з іншими з групи, спираючись на власний досвід (переваги) – самостійне формування стратегії. Цей підхід реалізується у разі, коли кожен об'єкт групи виконує власну задачу і тим вносить особистий вклад у рішення загально системної задачі групи. Задача окремого об'єкта буде порівняно нескладною, оскільки він вирішує задачу оптимізації лише своїх дій у складі групи, не оптимізуючи дії всієї групи.

Важливою перевагою децентралізованих стратегій групового управління є підняття в цілому живучості групи. Оскільки усі технічні об'єкти рівнозначні у групі, то втрата чи пошкодження будь-якого з них не призводить до втрати працездатності всієї групи. І підвищення живучості групи досягається без додаткових витрат, а лише за рахунок самої децентралізованої організації групового управління [1].

Нажаль стратегії децентралізованого групового управління складно алгоритмізувати і до того ж, вони не гарантують оптимальність рішення групової задачі. Та у разі підвищених вимог до живучості групи технічних об'єктів слід обирати децентралізовані стратегії управління групою.

Задача використання групи технічних об'єктів. У практиці, наприклад при обстеженні та картографуванні великих територій, виникає задача групового застосування технічних об'єктів, яка полягає у тому, щоб опрацювати інформацію з території з мінімальним перекриттям областей, що опрацьовуються окремим технічним об'єктом групи і з мінімальними витратами на переміщення об'єктів.

Група являє собою множину технічних об'єктів, які виконують реєстрацію, передачу, збереження і обробку даних різних типів, а також взаємодіють один з одним через захищені канали зв'язку, утворюючи певну цілісність і єдність для досягнення загальносистемної цілі.

Припустимо, у середовищі функціонує група O з N об'єктів O_i, яка може опрацювати деяку територію площею S. Окремий об'єкт $O_i \in O$ може опрацювати площу $s_i = \pi l^2$, де l – радіус області, яку опрацьовує O_i.

Для того, щоб площа покриття, яку опрацьовує група, була максимальною, необхідно, щоб величина $R = \sum_{i}^{N} \sum_{i=i+1}^{N} F(T_i, T_j)$ була мінімальною. $T_i \in S$ і є точкою, де об'єкт O_i знімає інформацію. Це є цільова точка об'єкта O_i ($i = \overline{1, N}$).

Функція $F = (T_i, T_j)$ визначає, як перетинаються зони опрацювання i-го та j-го об'єктів групи, якщо об'єкт O_i знімає інформацію з точки $T_i \in S$, а об'єкт O_j – з точки $T_j \in S$. Величина R визначає загальну площу перетину зон опрацювання інформації всіма об'єктами, тому, чим менше буде перетинів між зонами опрацювання, тим відповідно буде більшою загальна площа опрацювання інформації всіма об'єктами. Зрозуміло, що має виконуватись умова відсутності пропусків на площі покриття точок опрацювання інформації.

У [2] доведено, що функція перетину зон опрацювання i-го та j-го об'єктів групи

$$F = (T_i, T_j) = \begin{cases} 0, & якщо \quad L_{i,j} \geq 2l, \\ \varphi(\Delta_{i,j}), & якщо \quad L_{i,j} < 2l, \end{cases}$$

де $\Delta_{i,j} = 2l - L_{i,j}$, а $L_{i,j}$ – відстань між точками T_i та T_j . Також в [2] визначені необхідні умови, але, нажаль, не достатні, відсутності пропусків на площині покриття точок опрацювання інформації.

У [2,3] обґрунтовано, що рішення задачі покриття площини точок опрацювання інформації без пропусків групою у складі N технічних об'єктів залежить від «першої області», яку обере будь-який технічний об'єкт, і показано, що доцільно у якості «першої області» обирати ту, що найближче до центру площини, яку мають опрацювати технічні об'єкти $O_i \in O$.

Задачу покриття площини точок опрацювання інформації групою у складі N технічних об'єктів, як показано в [2], можна звести до класичної задачі про призначення.

В умовах наявності організованих уражуючи впливів у середовищі функціонування групи технічних об'єктів задача групового управління стає складною і алгоритми її рішення належать до класу алгоритмів із слабким апріорним і слабким апостеріорним інформаційним забезпеченням.

Висновок. Утворення групи об'єктів з організацією групового управління (колективною поведінкою) дозволяє забезпечити спільне розв'язання сукупності задач, які неможливо вирішити у разі неколективної поведінки. Децентралізована стратегія управління групою підвищує ефективність функціонування і ймовірність досягнення загальносистемної цілі, також виконання задачі окремим об'єктом.

Публікація містить результати досліджень, які проводяться за грантової підтримки Державного фонду фундаментальних досліджень за конкурсним проектом Ф76/109-2017/1.

Література

1. Додонов А.Г., Кузнецова М.Г., Горбачик Е.С. Введение в теорию живучести вычислительных систем. – К.: Наук. думка, 1990. –184 с.

2. Каляев И.А., Гайдук А.Р., Капустин С.Г. Модели и алгоритмы коллективного управления в группах роботов.-М.: ФИЗМАТДИТ, 2009.- 280 с.

3. Капустян С.Г. Метод организации мультиагентного взаимодействия в распределенних системах управления груп пой роботов при решении задач покрытия площади // Искусственный интеллект, 2004.- №3.- С.715-727.

АНАЛІЗ ПУБЛІКАЦІЙНОЇ АКТИВНОСТІ ЗА НАПРЯМКОМ КОМП'ЮТЕРНОЇ БЕЗПЕКИ НА БАЗІ РЕСУРСІВ WEB OF SCIENCE TA SCOPUS

В.Б. Андрущенко, І.В. Балагура
Інститут проблем реєстрації інформації Національної академії наук України, Київ, Україна
valentyna.andrushchenko@gmail.com, balaguraira@gmail.com

На сьогодні питання інформаційної безпеки є чи не найголовнішим і глобальним питанням для всіх країн світу. Питання наукової інформації та її доступності з цього питання є протирічливим. В роботі було проаналізовано масив реферативної та повнотекстової інформації за даним напрямком, що представлено в світових наукометричних ресурсах. Представлено результати щодо хронології та розподілу публікацій за науковими напрямками, публікаційною активністю країн та інтересу тих чи інших видань до цієї проблематики. Також представлено масив публікаційї з України та їх короткий аналіз.

Вступ

Інформаційні технології сьогодні є невід'ємною частиною всіх сфер існування людства – повсякденна комунікація, будь-яка діяльність, робота державних і недержавних установ, підприємств, офісів, закладів медицини та освіти. З кожним днем інформація набуває іншого формату представлення і зберігання. Людство поступово відмовляється від паперових носіїв за рахунок використання електронних пристроїв – це не тільки оптимізація та прискорення обміну, візуалізації та передачі інформації, але і економія природних ресурсів, що є актуальним і пріоритетним в усьому світі питанням.

Інформаційна безпека – є першочерговим аспектом для постійної та неперервної роботи тих чи інших установ і забезпечення стабільності процесів – адже на сьогодні світ повністю залежить від інформаційних технологій та їх стрімкого розвитку.

Тільки протягом 2017 року було зафіксовано 7 глобальних кібератак (1), що були спрямовані на пошкодження не тільки програмних засобів, але й вихід зі строю апаратних засобів, збій та зупинку роботи великих корпорацій, підприємств, державних установ в усьому світі, саботаж виборів, а також витік

конфіденційної інформації. Серед таких атак, що опинилися серйозною загрозою для України (2) є **Petya/NotPetya/Nyetya/Goldeneye** – глобальна кібератака, що відбулася за місяць після удару WannaCry, що порушив роботу в першу чергу медичних закладів і спричинив затримку надання медичної допомоги пацієнтам клінік Великої Британії та втрату значної суми коштів у криптовалюті. Атака **Petya/NotPetya/Nyetya/Goldeneye** порушила роботу і великих українських підприємств, компаній та банків, серед яких Державне підприємство «Антонов», Міжнародний аеропорт «Київ», Ощадбанк, Промінвестбанк, Укрсіббанк, компанії – Київстар, Водафон, телевізійні канали – СТБ, ICTV, АТР та інших.

Питання кібербезпеки сьогодні є одним із пріоритетних не тільки в Україні, але і у світі. Події поточного року довели, що немає країни повністю захищеної від кібератак.

Проблематика дослідження

Наукові роботи з інформаційної безпеки розміщені у спеціалізованих виданнях та на ресурсах, що є доступними тільки вузькому колу читачів/користувачів.

З точки зору наукової роботи, публікації зокрема, важливо відмітити те, що питання кібербезпеки є поняттям, що притаманне не тільки окремим галузям, а може бути елементом дослідження широкого кола напрямків.

Основна задача дослідження полягає у виокремленні публікацій, що пов’язані із терміном – «кібербезпека» («cybersecurity»), визначенні хронології – відповідно збільшення чи зменшення кількості публікацій за останні 15 років, кола наукових напрямків, наукових видань, країн, представлення інформації з цього питання у журнальних публікаціях, тезах конференцій, тощо. Всі дані було отримано за рахунок обробки інформації, що міститься у двох провідних наукометричних ресурсах – Web Of Science (Clarivate Analytics) та Scopus (Elsevier). Ці системи було обрано для виконання дослідження через їх авторитетність. Дані систем передбачають низку суворих вимог до видань, які індексуються цими ресурсами і відповідно мова іде про компетентність і достовірність інформації, що є представлення у відповідних виданнях. Це дозволяє впевнено говорити про науковість інформації, на базі якої сформовано аналітичні результати.

Аналіз публікацій за даними наукометричної бази даних Web of Science

В межах ресурсу Web Of Science було виділено 1454 записи за тегом «cybersecurity». Публікації відповідають тематиці комп'ютерних наук, технічних наук, бізнес економіці, що відповідає категоріям Web of science: Engineering electrical electronic, computer science theory methods, computer science information systems, computer science software engineering, computer science hardware architecture. Найбільшу кількість публікацій в даній тематиці опубліковано співробітниками організацій із США United states department of energy DOE, United states department of defence, University system of Maryland та ін.

Тематика найбільш цитованих наукових праць у відповідності до розподілу Web Of Science:

- Reliability assessment of photovoltaic power systems;
- Review of current status and future perspectives;
- Cyber-Physical System Security for the Electric Power Grid;
- A Reliability Perspective of the Smart Grid.

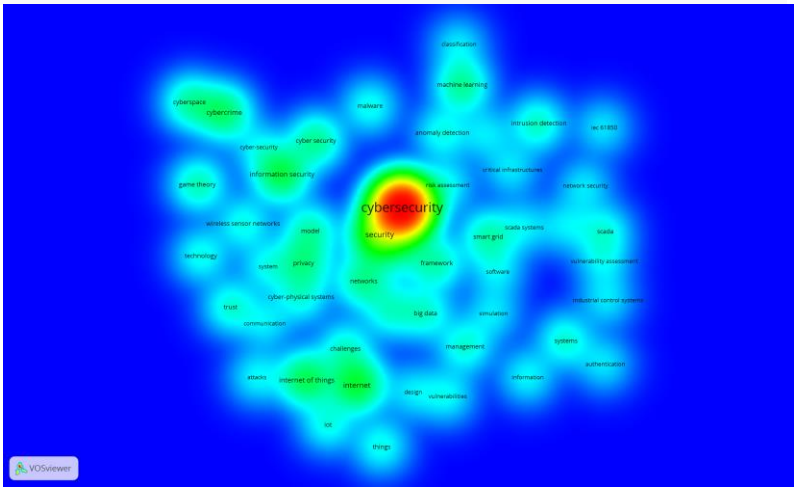


Рисунок1 – Мережа термінів за тегом cybersecurity

Відповідно до цитування робіт, актуальність розглянутої тематики особливо зросла в останні роки (рис. 2)

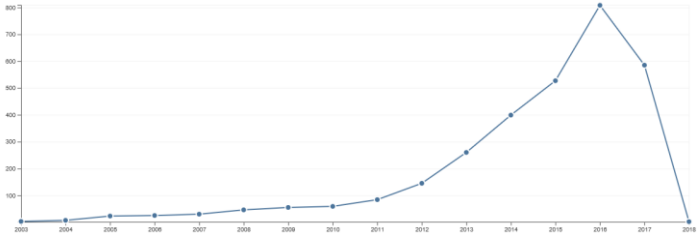


Рисунок 2 – Цитування публікацій, що розміщен у виданнях, які індексуються наукометричним ресурсом Web of Science

Авторами із України за тегом опубліковано 4 роботи, роки публікацій 2015-2017. Терміни, виділені в роботах українських авторів: cybersecurity, cyberthreat, intelligent technique, differential transformation, mathematical model, cyberspace, information protection, security incident, critical aviation information, civil aviation, security control, criticality. Незначна кількість робіт українських авторів не дозволяє визначити експертів у даній галузі та найбільш розвинуті напрями досліджень в Україні. Проте питання кібербезпеки не відрізняються в різних країнах. Тому слід звернути увагу на актуальні напрями, визначені на основі мережі термінів (рис.1): big data, machine learning, internet of things, information secutity, cyber-physical systems and others.

Аналіз публікацій за даними наукометричної бази даних Scopus

В наукомеричному ресурсі Scopus за останні 15 років представлено 2848 публікацій. На рис. 3 показано динаміку кількості публікацій протягом зазначеного періоду.

Характерний зріст уваги науковців до цієї проблематики помітний з 2011 року. Саме останнні ~10 років вважаються набільш актуальним для розгляду через підвищення кіберзагроз (3). Пік публікаційної активності припадає на 2016 рік – рік стрімкого зросту кількості та частоти кібератак, і тому зумовлений нагальним інтересом до цього питання. За аналітичними даними (4) саме у 2016 році більше 4000 шантаж-атак відбулося з початку 2016 року. Поряд з цим у 2016 році значно зріс ринок кібербезпеки – загальноссвітові витрати країн на питання інформаційної безпеки (5).

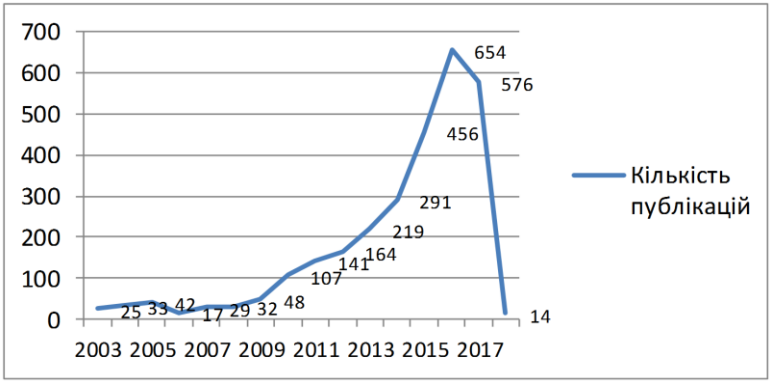


Рисунок 3 – Динаміка кількості публікацій за тегом «cybersecurity» за період 2003-2017 роки

Також дуже важливим є розподіл масиву публкацій в індексованих ресурсом Scopus журналів за науковими напрямками (рис.4). Такий розподіл показує значну кількість – 26 наукових спрямувань (за класифікцією Scopus), науковий інтерес вчених яких пов'язаний із поняттям кібербезпеки.

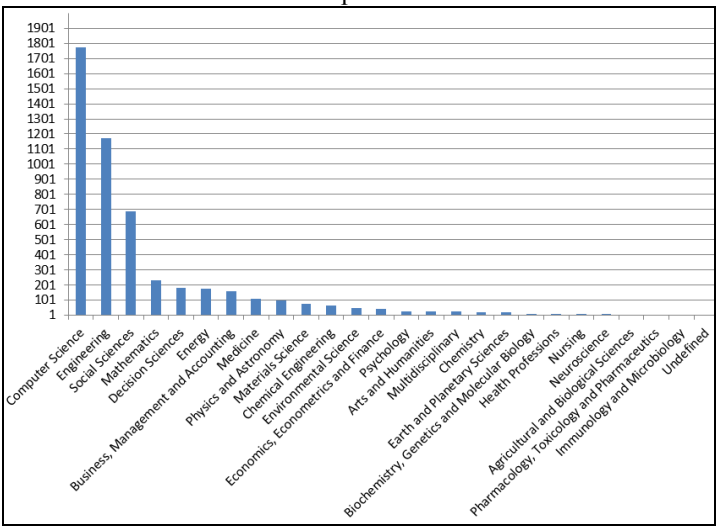


Рисунок 4 – Розподіл кількості публікацій Scopus за тегом «cybersecurity» за науковими напрямами.

З рисунку видно, що спектр наукових напрямів є достатньо широкий, але основний науковий інтерес притаманний 3 науковим напрямам – Computer Science – спрямовані на висвітлення аналітичних даних куберзагроз, побудови програмних та апаратних засобів протидії кібернападам, Engineering – засоби виробництв для захисту та Social Sciences – аналіз поведінкової та психологічної складової, людського фактору в межах даного явища.

Важливою і в той же час протирічливою постає картина розподілу кількості публікацій на країнами. З одного боку цілком закономірним є пріоритет Сполучених Штатів Америки з огляду на підвищену увагу до проблематики кібербезпеки не тільки з точки зору захисту та безпеки інформації, а й стартегічних міркувань. В той же час Великобританія, як основна постраждала країна від кібератаки WannaCry. Третє місце посідає Китай (рис. 5).

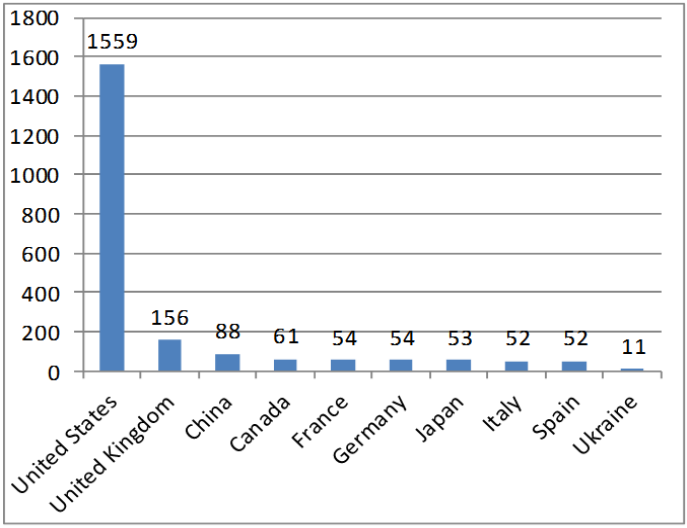


Рисунок 5 – Розподіл кількості публікацій Scopus за тегом «cybersecurity» між країнами світу

Серед видань, що містять найбільшу кількість публікацій за заданою тематикою такі:

1. IEEE Security And Privacy – 97 публікацій.
2. Computer – 49 публікацій.
3. IT Professional – 25 публікацій.
4. Intech – 20 публікацій.

5. Pipeline And Gas Journal – 15 публікацій.
6. IEEE Internet Computing – 12 публікацій.
7. IEEE Transactions On Smart Grid –
8. Aviation Week And Space Technology New York – 11 публікацій.
9. ACM Inroads – 10 публікацій.
10. Computers And Security – 10 публікацій.

Серед загального масиву виокремлено 11 публікацій з України. Такий показник не є ілюстративним. Можна зазначити кілька обгрунтувань низької публікаційної активності за напрямком (за даними Scopus) через малу кількість видань з України, що індексуються наукометричною базою Scopus та афіліацією фахівців з України організаціями інших країн.

В таблиці 1 зазначено інформацію щодо публікацій з України, вказано назву роботи, назву видання та тип публікації, назву організації, яку представляють автори публікації.

Таблиця 1. Публікації з України в наукометричній базі даних Scopus за тегом «cybersecurity»

№ п/п	Назва публікації	Назва видання	Організація	Тип публікації
1.	Cybersecurity of distributed information systems. The minimization of damage caused by errors of operators during group activity	2nd International Conference on Advanced Information and Communication Technologies, AICT 2017 – Proceedings 29 August 2017, Article number 8020071, Pages 83-87	Sumy State University, Sumy Sumy National Agrarian University	Матеріали конференції
2.	A Mathematical Cybersecurity Model of a Computer Network for the Control of Power Supply of Traction Substations	Cybernetics and Systems Analysis Volume 53, Issue 3, 1 May 2017, Pages 476-484	State Economic-Technological University of Transport S. P. Korolyov Zhytomyr Military Institute	Стаття в науковому виданні
3.	Applying the functional effectiveness information index in cybersecurity adaptive expert system of information and communication transport systems	Journal of Theoretical and Applied Information Technology Volume 95, Issue 8, March 2017, Pages 1705-1714 **Open Access**	Faculty of Information Technology, Taras Shevchenko National University of Kyiv Economics Information Systems Department, SHEE "Kyiv National Economic University named after Vadym Hetman"	Стаття в науковому виданні
4.	Method for assessing cybersecurity of distributed computer networks for	Journal of Automation and Information Sciences Volume 49, Issue 7, 2017,	State Economy and Technology University of	Стаття в науковому виданні

	control of electricity consumption of power supply distances	Pages 48-57	Transport	
5.	Designing a decision support system for the weakly formalized problems in the provision of cybersecurity	EasternEuropean Journal of Enterprise Technologies Volume 1, Issue 2-85, 2017, Pages 4-15 **Open Access**	Department of Computer Engineering International, Kazakh-Turkish University named after H. A. Yesevi B. Department of Managing Information Security, European University Department of Information Technology Security, National Aviation University,	Стаття в науковому виданні
6.	Management of information protection based on the integrated implementation of decision support systems	EasternEuropean Journal of Enterprise Technologies Volume 5, Issue 9-89, 2017, Pages 36-42 **Open Access**	Department of Managing Information Security, European University Department of Technical Information Security Tools Department of Information Technology Security, National Aviation University	Стаття в науковому виданні
7.	Critical aviation information systems cybersecurity	Meeting Security Challenges Through Data Analytics and Decision Support2016, Pages 308-316NATO Advanced Research Workshop on Meeting Security Challenges Through Data Analytics and Decision Support, ARW 2015; Aghveran; Armenia; 1 June 2015 through 5 June 2015; Code 125134	IT-Security Academic Department, National Aviation University	Матеріали конференції
8.	Cybersecurity case for FPGA-based NPP instrumentation and control systems	International Conference on Nuclear Engineering, Proceedings, ICONEVolume 5, 2016, Article number V005T15A0272016 24th International Conference on Nuclear Engineering, ICONE 2016; Charlotte; United States; 26 June 2016 - 30 June 2016; Code 124474	National Aerospace University KhAI, Computer Systems and Networks Department Center for Safety Infrastructure-Oriented Research and Analysis, RPC Radiy, Kirovograd	Матеріали конференції
9.	The study of participants' values convergence on the	EasternEuropean Journal of Enterprise Technologies	Department of Information	Стаття в науковому

	example of international scientific project on cyber security	Volume 6, Issue 3-84, 2016, Pages 4-11 **Open Access**	Technology and Software Engineering, Chernihiv National University of Technology Department of Public Administration and Management organization, Chernihiv National University of Technology	виданні
10.	The control technology of integrity and legitimacy of LUT-oriented information object usage by self-recovering digital watermark	CEUR Workshop ProceedingsVolume 1356, 2015, Pages 486-49711th International Conference on ICT in Education, Research and Industrial Applications: Integration, Harmonization and Knowledge Transfer, ICTERI 2015; Lviv; Ukraine; 14 May 2015 - 16 May 2015; Code 112160	Odessa National Polytechnic University	Матеріали конференції
11.	Bilevel programming and applications	Mathematical Problems in Engineering Volume 2015, 2015, Article number 310301 **Open Access**	Departamento de Ingeniería Industrial y de Sistemas, Mexico Department of Social Modeling, Central Economics and Mathematics Institute (CEMI), Russian Academy of Sciences (RAS) Department of Electronics and Computing, Sumy State University	Огляд

Важливо відмітити, що серед публікацій є і статті у наукових виданнях і матеріали конференцій. Також 5 із зазначених публікації є розміщенні у відкритому доступі і доступні світовому науковому загалу.

Висновки

Обрана проблематика та концепція дослідження дозволяє за рахунок аналізу даних авторитетних наукометричних ресурсів, що містять реферативну інформацію щодо публкацій за напрямком кібербезпека, оцінити увагу світового наукового загалу до

проблеми та внесок українських вчених до розв'язку питань комп'ютерної безпеки.

Започатковані дослідження вимагають більш глибоко аналізу спеціалістів не тільки в галузі інормаційних технологій,а й інших галузей знань і може стати підгрунтям для розвитку та вдосконалення політики інформаційної безпеки України.

На сьогодні питання безпеки в усіх галузях знань є не тільки пріоритетом наукових розробок, але й одним із приоритетних напрямків програми грантової підтримки Європейського Союзу «Горизонт 2020». Створення консорціумів, залучення спеціалістів різних наукових напрямків є чреговою сходнкою до пошуку нових шляхів боротьби із кіберзагрозами. Проведений аналіз із подальшим розвитком може стати актуальним інструментом для обраття тематики проектів, пошуку експертів та формквання колаборацій для реалізації результативних досліджень, що зможуть продукувати нові наукові знання на благо безпеки світу.

Перелік посилань:

1. The Biggest Cybersecurity Disasters of 2017, URL: https://www.wired.com/story/2017-biggest-hacks-so-far/ (дата звернення 2 листопада 2017 року)
2. 2017 Cyberattacks on Ukraine, URL: https://en.wikipedia.org/wiki/2017_cyberattacks_on_Ukraine
3. https://news.sap.com/pwc-study-biggest-increase-in-cyberattacks-in-over-10-years/ (дата звернення 30 жовтня 2017 року)
4. 6 Must-Know Cybersecurity Statistics For 2017/ Barkly Blog, URL: https://blog.barkly.com/cyber-security-statistics-2017 (дата звернення 1 листопада 2017 року)
5. Cybersecurity Market Reaches $75 Billion In 2015; Expected To Reach $170 Billion By 2020, URL: https://www.forbes.com/sites/stevemorgan/2015/12/20/cybersecurity-market-reaches-75-billion-in-2015-expected-to-reach-170-billion-by-2020 (дата звернення 3 листопада 2017 року)

ІНТЕРНЕТ РЕЧЕЙ (IOT): ІНТЕГРАЛЬНА БЕЗПЕКА

О.А. Баранов
Науково-дослідний інститут інформатики і права Національної академії правових наук України, м. Київ, Україна

Феномен Інтернету речей (скорочено - IP або Internet of Things - англ., скорочено - IoT), який отримав таку назву понад 15 років тому, потужно увійшов в життя сучасного суспільства, демонструючи вражаючі результати в усе нових і нових сферах людської діяльності. В передових країнах світу активно йде впровадження технологій Інтернету речей. За даними багатьох експертів: до 2025 року до 100 мільярдів пристроїв будуть підключені до мережі Інтернет, а ринок IP буде складати $ 7-19 трильйонів. У 2016 році Китай прийняв п'ятирічну Державну програму з загальним фінансуванням $ 127,5 млрд. Німеччина у 2016 році, як і багато розвинених країн, стала виконувати програму «Індустрія 4.0» – багаторічну стратегічну ініціативу для створення всеосяжного бачення і плану дій щодо IP в промисловому секторі.

Приклади у сфері безпілотних автомобілів безсумнівно свідчать про невідворотність настання ери Інтернету речей: Intel витратить 15,3 мільярда $ на покупку Mobileye (виробника компонентів для безпілотних авто); Qualcomm придбала за 47 млрд. $ NXP, найбільшого постачальника автомобільних чіпів; Google, Uber, Ford, Tesla, Nvidia мають власні програми щодо створення власних безпілотних автомобілів.

На основі проведеного аналізу наукових юридичних і технічних джерел, виявлених характерних властивостей IP та результатів роботи [1] запропонуємо наступну дефініцію терміну: «Інтернет речей» – це с сукупність технологій, що поєднують методи, способи і процеси, а також взаємодіючих технічних систем і комплексів, які складаються з мікропроцесорів, сенсорів, виконавчих пристроїв, систем передачі даних, локальних та/або розподілених обчислювальних ресурсів і програмних засобів, призначених для здійснення суспільних відносин, в тому числі, пов'язаних з наданням послуг або проведенням робіт, на основі використання безлічі різноманітних даних і мережі Інтернет при безпосередній участі або без участі суб'єктів цих відносин (юридичних або фізичних осіб).

За 20-25 років стан речей з технологіями IP буде характеризуватись наступним: поширеністю – IP буде пронизувати

всі сфери людської діяльності; інтелектуальністю – практично всі комплекси IP будуть функціонувати на базі штучного інтелекту; самоорганізацією, динамічністю і гетерогенністю – різнорідні комплекси IP будуть самостійно об'єднуватись з іншими комплексами IP; великими даними – буде мати місце колосальна концентрація різноманітних даних; стрімким розвитком мережі Інтернет – відбудеться колосальне зростання трафіку цифрових даних.

Таким чином, з огляду на те, що комплексі та системи IP будуть безперервно забезпечувати людську діяльність у будь-який сфері, базуючись на використанні комп'ютерних систем та мережі Інтернет, можна дійти висновку, що об'єкти з комплексами і системами IP за визначенням це об'єкти критичної інфраструктури для будь-якої діяльності за участю людини або без її участі.

Розвитку та впровадженню технологій IP на сучасному етапі притаманні такі основні зони ризиків та бар'єрів: техніко-технологічні; безпеки; конфіденційності; сумісності; стандартизації; правового регулювання.

Це обумовлюється низкою певних факторів. Наприклад, для техніко-технологічної сфери буде характерним дотримання наступного принципу, що створює умови для забезпечення постійного розвитку: всі технології, які дозволять надати послугу або здійснити роботи комфортніше, дешевше, швидше, якісніше, безпечніше будуть імплементовані в технології IP. Зазначене призведе до того, що сфера IP буде характеризуватись наступним: архітектура IP – багатозв'язна, багатовимірна, складна, адаптивна, багатофункціональна; ідентифікація – всі складові елементи комплексів та систем IP, об'єкти, якім надаються послуги або для яких виконуються роботи мають бути ідентифіковані; безшовною роботою технологій передачі даних; максимальною ефективністю користування радіочастотним ресурсом; економним енергоспоживанням і збільшенням часу автономності; транскордонним характером взаємодії; підвищеними вимогами до надійності і стійкості.

У великій кількості дослідних робіт і статтях обговорюються питання безпеки IP. Однак, як і в ситуації з визначенням терміну «інтернет речей», консолідованого підходу до визначення дефініції терміну безпеки інтернету речей ще не відбулось, що призводить до дещо хаотичного, безсистемного уявлення сутності безпеки інтернету речей та формування переліку та змісту проблем,

пов'язаних з безпекою, пошуку шляхів вирішення цих проблем тощо.

У 2015 році Федеральна торгова палата (США) опублікувала звіт, в якому підкреслюється наявність цілому розуміння важливості того, що при використанні технологій інтернету речей повинна забезпечуватися розумна (прийнятна) безпека, зміст якої і вимоги до якої ще тільки належить визначити на основі кращих практик компаній [2]. З цього звіту випливає, що в дійсності проблема безпеки інтернету речей поки ще не усвідомлена в повному її масштабі і значенні, а надія на кращі практики компаній є далеко не найкращою державною політикою забезпечення безпеки.

Зі змісту технологій IP, досвіду їх впровадження, вже наявних результатів використання технологій IP, високої динаміки цих процесів стає зрозумілим, що в самому найближчому майбутньому чимала частина людської діяльності буде повністю базуватися на технологіях IP, а значна частка інтересів і потреб людини і суспільства будуть задовольнятися виключно завдяки функціонуванню певних систем і комплексів IP, які будуть виконувати різноманітні роботи і надавати послуги в інтересах людей, іноді навіть без їх участі.

При розгляді проблем безпеки будемо виходити з наступної характеристики:

– IP, як глобальна система, буде складатись з величезної кількості (з десятків і сотень мільйонів) окремих систем і комплексів IP;

– архітектура IP буде являти собою суперскладну, гетерогенну, гібридну, багатозв'язну та багатофункціональну глобальну гіперсистему IP;

– конфігурація окремих частин (різної розмірності) гіперсистеми IP може адаптивно, заздалегідь не передбачуваним чином, змінюватися відповідно до завдань, які будуть вирішуватись системами та комплексами IP;

– для виконання якихось робіт або надання послуг певною системою або комплексом IP може знадобитися заздалегідь непередбачувана, тобто випадкова взаємодія з будь-яким іншим елементом, системою або комплексом гіперсистеми IP;

– безпека будь-якої діяльності людини, суспільства і держави стає залежною від безпечного функціонування IP, на якій ця діяльність базується.

Отже, основна мета забезпечення безпеки гіперсистеми IP і/або її складових – це забезпечення надійності, стійкості і якості здійснення тієї чи іншої діяльності, що базуються на використанні технологій IP.

З огляду на надану вище характеристику та за умови широкого використання комплексів та систем IP потрібно визначати та одночасно забезпечувати наступне: інфраструктурну безпеку (надійність, стійкість, резервування); технологічну безпеку; інформаційну безпеку; кібернетичну безпеку; багаторівневу систему безпеки; безперервність всіх складових безпеки; інтеграцію та взаємодію різноманітних систем безпеки

Таким чином, безпека гіперсистеми IP є інтегральною, що складається з наступних видів безпеки: інфраструктурної безпеки; технологічної безпеки; інформаційної безпеки; кібернетичної безпеки.

З огляду на зазначене, управління системою безпеки гіперсистеми IP і/або її складових потребує розроблення теоретичних та практичних засад створення та функціонування спеціального менеджменту інтегральної безпеки.

В роботі надається характеристика всіх складових інтегральної безпеки IP.

Література

1. Баранов О. «Інтернет речей» як правовий термін / О. Баранов // Юридична Україна. – 2016. – №5-6. – С. 96-103.
2. Internet of Things: Privacy & Security in a Connected. – World Federal Trade Commission (FTC) Staff Report. January 2015 – https://www.ftc.gov/system/files/documents/reports/federal-trade-commission-staff-report-november-2013-workshop-entitled-internet-things-privacy/150127iotrpt.pdf

РАЗРАБОТКА, ОЦЕНКА И ИСПОЛЬЗОВАНИЕ АЛГОРИТМА СЕГМЕНТАЦИИ СЛОВ ДЛЯ СИСТЕМ МОНИТОРИНГА НАЦИОНАЛЬНЫХ ИНТЕРНЕТ-РЕСУРСОВ

Березин Б.[1], Ландэ Д.[1], Павленко О.[2]
[1]Институт проблем регистрации информации НАН Украины, г. Киев, Украина
[2]Университет "Украина", г. Киев, Украина

Актуальность и анализ публикаций. Растущее количество информационных ресурсов, представленных в Интернете, ведет к необходимости развития поисковых систем для доступа к ним. В то же время, растет доля и важность мировых веб-ресурсов, представленных на языках, поиск в которых требует определения границ слов (китайский, японский, тайский и др.), в текстах которых не принято разделение слов пробелами. В работах [1,2] отмечаются такие особенности китайского Интернет-пространства как высокие темпы роста веб-ресурсов и числа пользователей; наличие собственных социальных сетей и поисковой системы Baidu, ориентированной на китайский язык (существенной проблемой является применение латиницы и кириллических кодов) и покрывающей значительную часть веб-ресурсов китайского сегмента Интернет. В этих работах также показаны подходы к построению систем мониторинга национальных Интернет-ресурсов, где актуальной является сегментация слов при формирования индекса поисковой системы.

В ранней работе [3], рассматривается портирование систем поиска и извлечения информации в сегмент ресурсов, представленных на китайском и японском языках. Отмечается важность индексирования китайских символов и проблемы автоматической сегментации. В [4] отмечается, что для решения проблемы сегментации китайского текста используются три основных способа: словарный (обычно с применением алгоритма максимального соответствия), статистический и комбинированный, сочетающий в себе оба предыдущих. При этом, удаётся правильно сегментировать текст более чем на 90%. Работа [5] рассматривает сегментацию слов языка урду (который также не использует разделителей между словами) как важную проблему приложений обработки естественных языков (NLP). Рассматриваются

технология совпадения с наиболее длинными словами из словаря (longest matching); технология максимального совпадения (maximum matching) и методы статистической сегментации.

Данная работа посвящена реализации модели алгоритма сегментации слов (АСС) для формирования индекса поисковой системы. Показаны варианты АСС, предложена реализация моделей алгоритма сегментации, получена оценка качества сегментации. Реализованные АСС используются авторами для создания обобщенной модели предметной области на базе мониторинга ресурсов китайского сегмента Интернет.

Особенности моделей алгоритма сегментации слов. В работах, посвященных сегментации слов, выделяются две основные модели — статистические и использующие словарь (правила, списки слов). Оценка этих двух подходов к сегментации выполнена в работе [6]. Для моделей на основе словаря должен быть доступен предопределенный словарь. При этом отмечается вариант алгоритма с максимальным совпадением (maximal matching), для которого есть модификации - forward maximal matching (FMM) и backward maximal matching (BMM), в зависимости от направления обработки текста. Второй вариант для алгоритма со словарем - это алгоритм, который находит сегментацию с минимальным количеством слов shortest path (SP).

Для моделей на основе словаря предполагается наличие списка слов, каждое из которых связано с оценкой вероятности того, что оно является истинным словом. Пусть

$$W = \{\{w_i, g(w_i)\}_{i=1,\ldots,n}\}$$

будет таким списком, где w_i является кандидатом на слово, а также $g(w_i)$ его функция качества. Алгоритм прямого максимального соответствия FMM обрабатывает текст T для вывода лучшего текущего слова w^*многократно с T=t* для каждого цикла, таким образом

$$\{w^*, t^*\} = \underset{wt=T}{\operatorname{argmax}}\, g(w)$$

с каждым $\{w, g(w)\} \in W$.

Алгоритм сегментации на основе кратчайшего пути [6,7] использует предположение о том, что правильная сегментация должна максимизировать длины всех слов или минимизировать общее количество слов. Для предложения *S* из *m* символов

$\{c_1, c_2, \ldots, c_m\}$ лучшее сегментированное предложение S^* из $\{w_1^*, w_2^*, \ldots, w_{n^*}^*\}$ n^* слов

$$S^* = \underset{w_1 \cdots w_i \cdots w_n = T}{\operatorname{argmin}} n.$$

Эта задача оптимизации преобразуется в задачу нахождения кратчайшего пути для направленного нециклического графа.

Разработка алгоритма сегментации. Для анализа АСС были рассмотрены различные реализации моделей сегментатора. С помощью языка Perl на основе алгоритма максимального соответствия FMM было разработано соответствующее программное обеспечение. Также рассматривалась предложенная в [8] реализация сегментатора на основе алгоритма максимального соответствия. При поиске слов он пытается использовать самое длинное возможне слово. Этот простой алгоритм достаточно эффективен при большом словаре. Особенностью алгоритма являются элементы средств идентификации и извлечения китайских именованных объектов. В реализации алгоритма на Perl описаны такие объекты, как числа, ASCII коды, китайские фамилии и т. п. и предусмотрены процедуры для их извлечения. Более подробно система извлечения китайских именованных объектов рассмотрена в [9], где приводятся несколько видов правил для отнесения слов к именованным объектам. Реализация сегментатора [8] была адаптирована и использована для формирования индекса поисковой системы при работе с новостными, научно-техническими и др. веб-ресурсами китайского сегмента Интернет.

В данной работе, для анализа и усовершенствования модели АСС предложен и разработан алгоритм с поиском кратчайшего пути в графе [10]. Его реализация выполнена на языке Perl. Предложенный алгоритм сегментации состоит из трех частей: Формирование таблицы Слов, Формирование таблицы Шаг-Позиция, Формирование массива сегментированных слов и вывод в файл. В первой части программы создается массив, каждая строка которого соответствует символу входной строки. При нахождении множества слов, на которые может быть разбита входная строка, для каждой входной буквы анализируются возможные подстроки, начинающиеся с данной буквы, длиной от 1 до n. (n – максимально возможная длина слова, зависит от языка. Для китайского, в словаре можно найти слова до 5-6 иероглифов, для русского до 18-20 букв и т.д.). Если для анализируемой подстроки находится соответствующее слово в словаре, то такое слово используется в

разбиении. Построенная таким образом таблица содержит множество слов, на которые может быть разбита входная строка при данном объеме словаря. В таблице 1 показано разбиение, полученное для входной строки IWORKATTHERESEARCHINSTITUTE (в словаре найдены слова и часто используемые сокращения).

Множество слов, на которые разбивается входная строка, может быть представлено направленным графом. Буквы входной строки являются вершинами такого графа, а слова разбиения, получаемые для каждой буквы, представляют ребра, исходящие из соответствующих вершин. На рис. 1 приведено представление разбиения входной строки IWORKATTHERESEARCHINSTITUTE на слова в виде графа. При таком представлении, выбор слов правильного разбиения входной строки может рассматриваться как поиск минимального пути в графе. Различные варианты волнового алгоритма, наряду с другими методами нахождения минимального пути в графе, могут использоваться для решения этой задачи.

Во второй части программы, в соответствии с волновым алгоритмом, с помощью таблицы Слов формируется таблица ШагПозиция, с помощью которой выполняется разметка вершин графа, определяется их расстояние (соответствующие количеству ребер в случае невзвешенного графа) от начальной вершины (первой буквы входной строки). Для этого, на каждом шаге (волне) алгоритма определяется множество соседних, достижимых в данный момент вершин. Т.е., таблица ШагПозиция содержит множество вершин графа, которые достижимы при анализе текущего входного символа (текущей строки таблицы Слов). А также содержит позиции начала слов, на которые разбивается входная строка при сегментации. В целом, программа может рассматриваться как модифицированный волновой алгоритм нахождения минимального пути в графе.

Таблица 1.

Входная строка	**Найденные в словаре слова разбиения**				
i	**i**				
w	**work**				
o	**or**				
r					
k					

a	**a**	**at**			
t					
t	**th**	**the**	**there**	**theres**	
h	**he**	**her**	**here**		
e	**ere**				
r	**re**	**res**	**research**		
e	**es**				
s	**se**	**sea**	**sear**	**search**	
e	**ear**				
a	**a**	**ar**	**arc**	**arch**	
r					
c	**ch**	**chi**	**chin**	**chins**	
h	**hi**				
i	**i**	**in**	**ins**	**inst**	**institute**
n	**ns**				
s	**st**				
t	**ti**	**tit**			
i	**i**	**it**			
t	**tu**	**tut**			
u	**ut**				
t					
e	**e**				

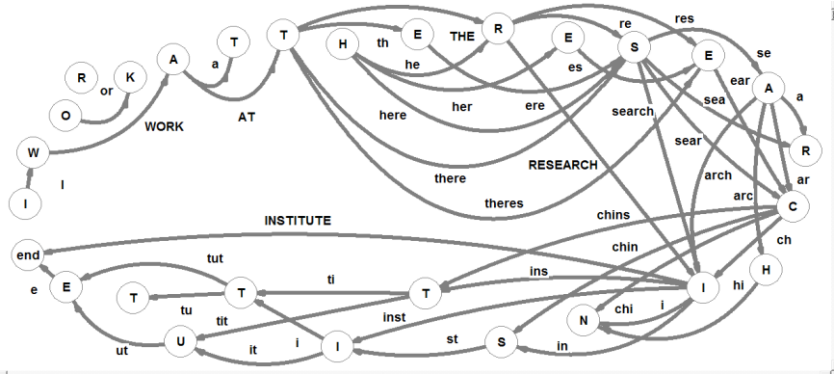


Рисунок 1 − Представление разбиения входной строки IWORKATTHERESEARCHINSTITUTE на слова в виде графа.

На третьем этапе алгоритма, начиная с последней вершины, на основе расстояний до предшествующих вершин, производится выбор ребер в графе, составляющих минимальный путь, т.е. выбор минимального количества слов, на которые может быть разбита, сегментирована входная строка. На этом этапе, при выборе слов, для улучшения качества сегментации, кроме расстояний могут также учитываться другие особенности языков, например частота использования слов в текстах на разных языках и т.п. Т.е., в этой части программы, на основе таблицы ШагПозиция и входной строки символов заполняется массив Сегментированных Слов. Затем слова из сформированного массива Сементированных Слов выводятся в выходной файл программы.

Оценка и использование моделей сегментации. Для оценки качества сегментации и возможности использования реализованных вариантов АСС проводилось тестирование алгоритма максимального соответствия FMM и алгоритма с поиском кратчайшего пути в графе, а также других, эксплуатируемых в сети Интернет сегментаторов. Аналогично работам [11,12], для оценки сегментаторов использовался инструментальный набор для оценки сегментации китайских слов https://web.archive.org/web/20100702191858fw_/
http://projectile.sv.cmu.edu
/research/public/tools/segmentation/eval/index.htm. В состав набора входят программные средства и тестовый корпус. При оценке сегментаторов использовался тестовый корпус из инструментального набора на основе китайских новостных текстов (данные проекта the Chinese Treebank, с китайскими текстами из Xinhua news, Sinorama news magazine, and Hong Kong News [11]). Кроме того, в данной работе был подготовлен тестовый корпус объемом 10 тысяч слов на основе частотного китайского словаря.

Метод оценки, который использовался для тестирования сегментаторов, называется Edit Distance of the Word Separator (расстояние при редактировании разделения слов, EDWS).

Пусть $C_1C_2....C_iC_{i+1}...C_n$. представляет предложение без сегментации. Сегментатор разделяет это предложение на последовательность слов путем вставки разделителя ***S*** между символами, в результате предложение может иметь, например, такой вид: $C_1C_2\boldsymbol{S}C_3C_4C_5\boldsymbol{S}C_6C_7C_8C_9\boldsymbol{S}C_{10}\boldsymbol{S}....C_iC_{i+1}...\boldsymbol{S}C_n$. Разные программы сегментации могут размещать разделители в разных позициях одного и того же предложения. Расстояние при редактировании разделения слов (EDWS) определяет, сколько

операций редактирования (вставки, удаления и совпадения) необходимы для модификации автоматически полученной сегментации текста до стандартной сегментации того же текста, полученной вручную, экспертами. Точность (precision) и полнота (recall) являются показателями, которые используются при оценке большей части алгоритмов извлечения информации. Для оценки сегментации эти показатели могут быть определены так:

Precision= (число совпадений) / (число разделителей в полученном тексте) *(1)*

Recall=(число совпадений) / (число разделителей в стандартном тексте) *(2)*

F=2*Precision*Recall / (Precision+Recall) *(3)*

В таблице 2 в столбце 1 показаны виды сегментаторов, для которых проводилось тестирование. В столбце 2 — тип корпуса (новостной, из инструментального набора, или словарный, отобранный из словаря), использованный при тестировании. В 3-5 столбцах таблицы приведены количества операций редактирования (совпадения, вставки, удаления), которые необходимо было выполнить для модификации текста после сегментатора в стандартный текст. В последних, 6-8 столбцах представлены значения показателей, рассчитанные по формулам (1) - (3) при проведении тестирования. Для алгоритма Е. Петерсона тестирование было проведено с использованием словарей двух размеров - 348982 слов (один из словарей сегментатора Jieba) и 119804 слов (словарь WordList).

Таблица 2 – Результаты тестирования**.**

Вид сегмента-тора	**Тестовый корпус**	**Сов.**	**Вст.**	**Удл.**	**Precision**	**Recall**	**F-1**
1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Jieba	**новостной**	**19581**	**2567**	**1233**	**94.08%**	**88.41%**	**0.91**
	словарный	**9966**	**33**	**365**	**96.47%**	**99.67%**	**0.98**
Online	**новостной**	**21334**	**814**	**1911**	**91.78%**	**96.32%**	**0.94**
	словарный	**9686**	**313**	**3508**	**73.41%**	**96.87%**	**0.84**
Алгоритм с поиском кратчай-шего пути в графе (словарь 348982 слов)	**новостной**	**19313**	**2835**	**1699**	**91.91%**	**87.20%**	**0.89**
	словарный	**9988**	**11**	**464**	**95.56%**	**99.89%**	**0.98**

Алгоритм Петерсона (словарь 348982 слов)	**новостной**	**20673**	**1475**	**1073**	**95.07%**	**93.34%**	**0.94**
	словарный	**9882**	**117**	**2940**	**77.07%**	**98.83%**	**0.87**
Алгоритм Петерсона (словарь 119804 слов)	**новостной**	**20524**	**1624**	**1452**	**93.39%**	**92.67%**	**0.93**
	словарный	**9696**	**303**	**4850**	**66.66%**	**96.97%**	**0.79**
Ir-Segmenter	**новостной**	**21345**	**803**	**3145**	**87.16%**	**96.37%**	**0.92**

Кроме тестирования алгоритмов сегментации на основе китайских текстов, для алгоритма максимального соответствия FMM и алгоритма с поиском кратчайшего пути в графе для оценки разбиения проводилось также тестирование с использованием русских и английских текстов и словарей. Оценивался процент слов, сегментированных с ошибками для массивов данных из различных видов информационных ресурсов - новостных сообщений, технических и художественных текстов. При этом использовались два вида словарей - словарь русского языка и словарь, сформированный путем накопления новостных сообщений.

Реализованные и адаптированные ACC использовались для мониторинга китайских Интернет-ресурсов. Интерфейс и результаты отработки запроса на мониторинг при использовании сегментатора на основе адаптированного алгоритма Е. Петерсона [8] в китайской социальной сети Weibo показаны на рис. 2. Кроме того, в рассматриваемой модели реализованы средства поиска документов в получаемых результатах мониторинга.

Рисунок 2 – Отработка запроса на мониторинг в китайской социальной сети Weibo

Выводы. Показана актуальность задачи сегментации слов при формировании индекса поисковых систем в связи с ростом ресурсов китайского и др. сегментов Интернет. Приведены варианты АСС, которые могут быть использованы для формирования индекса поисковой системы, показана применимость моделей на основе словаря.

- Рассмотрены модели реализации FMM АСС на основе словаря. Предложен алгоритм сегментации с поиском кратчайшего пути в графе и разработано программное обеспечение.

- Получены оценки качества сегментации и результаты использования модели АСС при формировании индекса поисковой системы для мониторинга веб-ресурсов китайского сегмента Интернет, которые показывают возможность использования алгоритма при достаточном объеме словаря.

1. Ландэ Д.В. Обзор особенностей и возможности контент-мониторинга национального сегмента сети Интернет / Д.В. Ландэ, Б.А. Березин, В.А. Додонов // Реєстрація, зберігання і обробка даних, - 2016. - Т. 18, - N 3. - С. 20-38.

2.Ландэ Д., Березин Б., Павленко О. Построение модели информационного сервиса на базе национального сегмента Интернет // Информационные технологии и безопасность. Материалы XVI Международной научно-практической конференции ИТБ-2016. - К.: ИПРИ НАН Украины, 2017. - С. 48-57.

3.Boisen, S. Chinese information extraction and retrieval / S. Boisen, M. Crystal, E. Peterson, R. Weischedel, J. Broglio, J. Callan, M. E. Okurowski // Proceedings of a workshop on held at Vienna, Virginia. Association for Computational Linguistics, - 1996. - P. 109-119.

4.Загибалов Т.Е. Автоматический анализ текстов на китайском языке. Проблема выбора базовой единицы // Труды международной конференции “ Диалог”, - 2005. – С. 31-37.

5.Durrani, N., Hussain, S. Urdu word segmentation // Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics. Association for Computational Linguistics, 2010. –P. 528-536.

6. Zhao H., Utiyama M., Sumita E., Lu B. L. An empirical study on word segmentation for chinese machine translation // International Conference on Intelligent Text Processing and Computational Linguistics. Springer Berlin Heidelberg, 2013. - P. 248-263.

7. Jia Z., Wang P., Zhao H. Graph model for Chinese spell checking // Proceedings of the 7th SIGHAN Workshop on Chinese Language Processing (SIGHAN'13), 2013. - P. 88-92.

8. [Электронный ресурс]. — Режим доступа: http://www.mandarintools.com/segmenter.html. Peterson Erik. — Название с экрана.

9. Peterson Erik. A Chinese named entity extraction system // Proceedings of the 8th Annual Conference of the International Association of Chinese Linguistics, Melbourne, Australia. 1999. – P. 47-58.

10. Ландэ Д.В., Березин Б.А., Павленко О.Ю. Разработка алгоритма сегментации слов для систем мониторинга национальных интернет-ресурсов //Міжнародна науково-практична конференція "Інтелектуальні технології лінгвістичного аналізу": Тези доповідей.- Київ: НАУ, 2017. - С. 11.

11. Fung R., Bigi B. (2015, October). Automatic word segmentation for spoken Cantonese // Oriental COCOSDA held jointly with 2015 Conference on Asian Spoken Language Research and Evaluation (O-COCOSDA/CASLRE), 2015 International Conference IEEE. -P. 196-201.

12. Chea V., Thu Y.K., Ding C., Utiyama M. Khmer word segmentation using conditional random fields // Khmer Natural Language Processing, 2015. – P. 62-69.

ІНФОРМАЦІЙНА ТЕХНОЛОГІЯ СЦЕНАРНОГО АНАЛІЗУ НА БАЗІ ДОСЛІДЖЕННЯ МОДЕЛЕЙ ПРЕДМЕТНИХ ОБЛАСТЕЙ

Бойченко А.В.
Інститут проблем реєстрації інформації НАН України, Київ, Україна

Досліджено алгоритми виділення ключових понять та запропоновано підхід до побудови мережі понять з використанням технології сценаріїв розвитку ситуації. Розроблена технологія онтологічного дослідження графа інформаційного середовища на основі сценарного підходу для опису, моделей і структурно-алгоритмічної організації інформаційної бази для різних ієрархічних рівнів їх подання.

Розроблена технологія розробки та дослідження сценаріїв розвитку ситуації, заснована на онтологіях, отриманих із масиву документів, які описують певну тематичну область. За основу для побудови онтологій пропонується використовувати мережі природньої ієрархії термінів, які базуються на інформаційно-значимих елементах тексту.

Послідовність завдань для реалізації пропонованого підходу може бути визначена таким чином [1,2]:

1) Виявлення базових цілей, що характеризують досліджуваний процес або систему;
2) Виявлення ключових елементів процесу або системи;
3) Виявлення взаємовпливів факторів;
4) Побудова когнітивної моделі;
5) Побудова можливих сценаріїв розвитку;
6) Визначення критеріїв оцінки сценаріїв;
7) Оцінка сценаріїв і виявлення найкращого з точки зору обраних критеріїв.

На вхід моделюючого комплексу надходить текст у вихідному форматі, тобто в популярних форматах документів DOC, OAT, RTF, PDF, HTML. Такі документи крім самого тексту містять також форматування, виноски, «зайві» ділянки тексту (реклама).

Для отримання тексту з документів, поданих на вхід використовуються існуючі програмні бібліотеки на мові Perl: LWP, HTML::Extract, HTML::Parse, HTML::FormatText,

Lingua::EN::Bigram, Lingua::StopWords. Далі визначаються межі слів, речень, абзаців, видаляються числа, розділові знаки та інші символи і послідовності символів, які є словами, але які відіграють важливу роль у реченні. Складність даного етапу полягає в тому, що деякі послідовності символів часто буває неможливо ідентифікувати на цьому етапі аналізу.

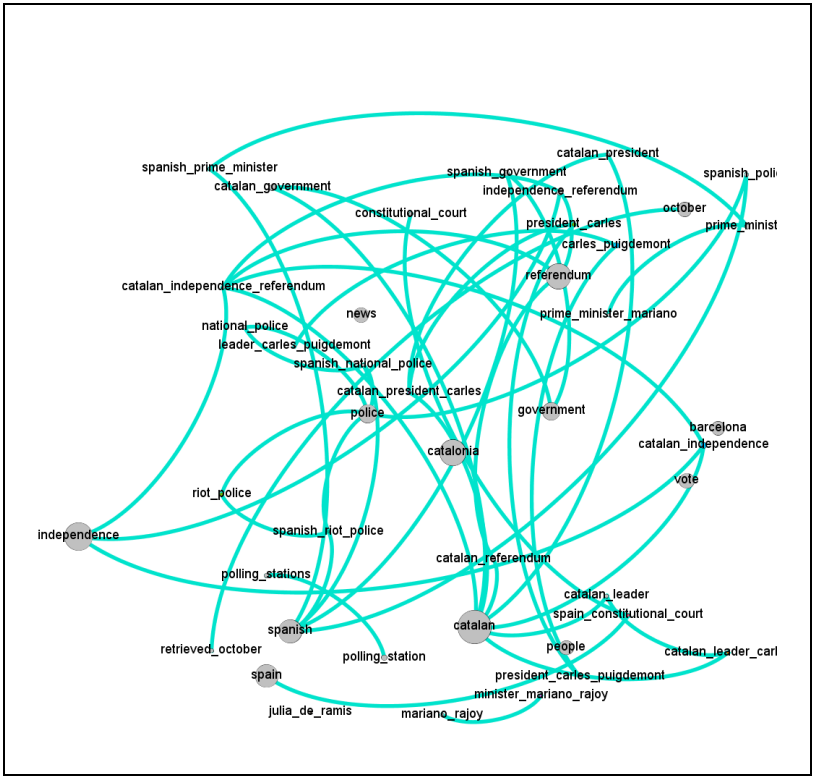


Рисунок 1 – Приклад отриманого графу

Для визначення цінності сценаріїв пропонується використати мурашиний алгоритм (англ. ant colony optimization, ACO) — один з ефективних поліноміальних алгоритмів для знаходження наближених розв'язків завдань пошуку маршрутів на графах.

Запропонована технологія дозволяє в автоматизованому режимі на базі аналізу вхідного пакету документів вирішувати задачу

розробки та дослідження сценаріїв здійснення впливів на об’єкти, які відповідають вибраним ключовим поняттям.

Література

1 Ландэ Д.В.,Снарский А.А. Подход к созданию терминологических онтологий // Онтология проектирования, 2014. - N 2(12). - С. 83-91.

2 Ланде Д.В., Бойченко А.В. Використання моделей предметних областей у задачах сценарного аналізу // Міжнародна науково-практична конференція "Інтелектуальні технології лінгвістичного аналізу": Тези доповідей.

RESEARCH OF THE SIMULATION POLYGON FOR THE PROTECTION OF CRITICAL INFORMATION RESOURCES

Liliia Galata, Bogdan Korniyenko, Alexander Yudin
National Aviation University, Kiev, Ukraine
galataliliya@gmail.com, bogdanko@i.ua, yak333@ukr.net

The article devises and implements a simulation polygon for protecting critical information resources based on the application software GNS3. The main element of computer network security is the Cisco ASA 5520 firewall, which divides the company's network into a demilitarized zone, an internal and an external network. In the demilitarized zone are Web-server and FTP-server configured. To test a simulation polygon for protecting critical information resources, a network scan was performed using the Zenmap utility. Using the hping3 utility, a stress-test of the network was implemented. Realized counteraction to these attacks.

1. Introduction

The main problem that needs to be solved while constructing a cybersecurity polygon which consists of real hardware is the high price of components for building a secure network. Sometimes, even if the company has all necessary equipment, it will be used in its own network and will not be provided for testing, network interconnection, and so on. For small companies, building of network for training purposes, for testing various network equipment, flexible configuration or testing various security policies are almost impossible, because it requires a lot of costs [1-4].

Therefore, it was decided to build a secure computer network based on a special emulator platform, which allows you to virtualize various network equipment and create a real virtual network on their base. So, as a platform for building a secure network, you need to use a full-fledged emulator that allows you to run a complete model of the network device and run the original software. The emulator runs the real Cisco IOS operation system, so it allows to run full-featured network device, whether it's a router or a switch, or a firewall. It means that, this network equipment will have all the features that are available on real equipment. There is also the possibility of launching a multivendor heterogeneous network, which allows you to use not only Cisco network devices, but also Juniper, Microtik, Check Point. Also, in order to build a fully

secured virtual computer network, it is necessary to have the ability to add workstations and servers to the network [5].

2. Emulator platforms

To build a secure virtual network, you can now use one of two most suitable platforms:

UNetLab (Unified Networking Lab, UNL) or GNS3 (Graphical Network Simulator-3). Firstly, both emulators are completely free, unlike the same VIRL or Boson NetSim, and secondly, have a similar functionality with certain features. The main disadvantage of such emulators is the impossibility of using them in the topology of wireless network devices, so you can`t build and test wireless networks.

GNS3 (Graphical Network Simulator).

GNS3 is a graphical network emulator that allows you to simulate a virtual network of network equipment from more than 20 different companies on a local computer, connect a virtual network to the real one, add a computer to the network, support other software for network packets analyzing, for example Wireshark. It also supports the SolarWinds Response Time Viewer utility, which accepts traffic logs and analyzes the network callback time and data volumes. Depending on the hardware platform for GNS3, it is possible to build complex projects consisting of Cisco, Cisco ASA, Juniper routers, and servers with network operation systems. For ease of testing, it supports virtualization software, which allows you to add a virtual machine to the network and make appropriate network tests, including VMware and VirtualBox. GNS3 is a graphics shell for Dynamips, it also has full support for QEMU and Cisco IOU, with all advantages and disadvantages of each technology [6-9].

Dynamips works on most Linux systems, Mac OS X and Windows, that allows the emulation of the routers's hardware part by downloading and interacting with real Cisco IOS images. The main purpose is to test and experiment with different versions of Cisco IOS, checking the configuration before using it on real hardware, and to use as a training laboratory.

Main benefits of Dynamips:

- A real IOS image is used for work, not partial emulation of some of its capabilities.
- Ability to work in hypervisor mode, for load balancing between multiple computers.
- Capture traffic on interfaces using the PCAP library (for example, using Wireshark).

- Connect the emulated network equipment to a real network.

Main drawbacks are:

- High system requirements, as a real IOS image is downloaded into memory.

- Impossibility to emulate Catalyst switches and some router models due to the large number of ASICs in them.

QEMU - free open source software for emulating various platforms hardware. UNetLab as GNS3 uses QEMU to emulate ASA, ASAv, IPS firewalls, and L2-switches, that are not emulated by Dynamips. The advantage of using QEMU in UNetLab is no RAM limitations and the number of connections for QEMU, compared to GNS3, which has a 2 GB RAM limit and 16 connections for various emulated devices.

Cisco IOU (Cisco IOS for UNIX) - Cisco emulator for internal use. It installs over UNIX, may work under Linux (IOS on Linux). The emulator has complete support for L3, L2 devices, with low CPU resources requirements (RAM is required a lot). There are no restrictions on the boards and interfaces. In the settings, you simply specify how much and what you need. It supports both GNS3 and UNetLab.

Emulator UNetLab (Unified Networking Lab, UNL).

UNenLab (Networking Lab Unified, UNL) is a multivendor and multi-user environment for creating and modeling various laboratories that allows you to simulate a virtual network with routers, switches, security devices and connect them to a real network, connect a virtual machine with any operation system or real computer. It allows you to use both Cisco images (Dynamips emulator) and Juniper or QEMU components. In addition, since UNetLab 0.9.54 release, it contains a multi-user functionality. By using the same virtual machine, each authorized user can create their stands independently of each other, and collaborate with a common stand that shares several users at one time. In this case, users have the opportunity to launch a common stand independently of each other.

Main features:

- It is a completely free product.

- It has web interface. So, there is no need to install a custom user client.

- Allows you to simulate virtual networks with switches, routers, security devices, and more. Capabilities are limited only by resources of the computer or the server.

- Almost full support for L2 level switches.

- Can be installed on both physical server and virtual machines VMware Player, VMware Workstation, VMware ESXi with Vsphere vClient, etc.

- The multi-user functionality allows UNetLab to be used as a network laboratory in educational institutions without any restrictions.

Built-in characteristics of platform functionality for emulation of computer networks.

Today, there are three Cisco VIRL, GNS3 and UNetLab emulators, which can be used for building and quality testing a secure computer network. The comparison of the disadvantages and advantages of each of them is given in Table 1. The main disadvantage of these emulators are the impossibility of emulating wireless network devices, as a result, the impossibility of including them in the network topology, which greatly reduces the functionality of the cybersecurity polygon.

Table 1.Comparative characteristics of the emulator's functionality

Characteristics	Cisco VIRL	GNS3	UNetLab
Serial Interfaces support	-	+	+
Advanced Cisco Network Equipment support	+	+	+
Other vendors support	-	+	+
Out-of-Band management	+	-	+
Multiuser platform	+	-	+
Easy to deploy	+	+	-
Network connections limitations	-	+	+
User support	+	+	-
Price	-	+	+
Emulation of wireless network devices	-	-	-

3. Setting of the network equipment

Cisco ASA Firewall Configuration

Cisco ASA is a hardware firewall with stateful session's inspection. ASA is able to work in two modes: routed (router mode, by default) and transparent (transparent firewall when ASA works as a filtration bridge) [10].

In routed mode, each ASA interface configures the IP address, mask, security level, interface name, and also the interface must be forced to

"up", because by default all interfaces are in the "disabled by administrator" mode.

The security level parameter is a number from 0 to 100, which allows you to compare 2 interfaces and determine which of them are more "secure". The parameter is used qualitatively, not quantitatively, that is important only "more or less." By default, the traffic that goes "outside", from the interface with a high level of security to the interface with a lower level of security, is skipped, the session is remembered and back passes only the answers to these sessions. Traffic going "inside" by default is forbidden.

So, firewall occurs between the logical vlan interfaces. As a rule, the level of security of interfaces is selected for the maximum match to the logical topology of the network. The topology represents a security zone and the rules of interaction between them. A classical model has various interfaces with different levels of security. There are no prohibitions about the same level of security for different interfaces, but by default exchange of traffic between such interfaces is forbidden. The "name of the interface" parameter (nameif) allows you to use in the settings not the physical name of the interface, this name can be chosen to mean its purpose (inside, outside, dmz, partner).

Between interfaces with the same level of security there is no firewall, but only routing. Therefore, this approach applies to interfaces that relate to the same logical security zone.

As any router (ASA also uses a routing table for package transfer), the configured interfaces automatically fall into the routing table with the tag "attached" (connected), for only «up» interfaces. Package routing between these networks is automatic.

The networks that for ASA not known should be described additionally. It is necessary to specify the interface for "next-hop", because ASA does not make such search (as opposed to the usual cisco router).

The routing table only gets one route to the destination network, as opposed to classical routers, where up to 16 parallel paths can be used. The default route is specified by hand additionally. If the ASA does not have a record in the routing table about destination network for some package, it will discard the package.

ASA, like most cisco devices, provides several ways to remote control. The telnet is easiest and no most dangerous way. More secure command line access is provided by the SSH protocol. However, when you use SSH, you should specify which hosts you can use to manage,

and you must also specify the RSA keys that are required to encrypt user data.

Cisco Routers Configuration.

First of all, the physical connection should be configured and the parameters of the interfaces should be set up. This phase includes the configuration of the physical interfaces - port speed, duplex, flow control, EtherChannel creation, 802.1Q sub-interfaces for using VLANs in the network.

Then you should configure the router's IP-address, the local and global interfaces.

The next step is to configure routing functions:

- routing settings, including static and dynamic routing;
- setting of automation issuing for DHCP addresses, including creation dynamic pools, static bindings;
- NAT Address Conversion Settings - dynamic and static rules for port wrap;
- secure access configuration, including a access list of physical interfaces and a list for virtual ports;
- users, authorization, privileged mode;
- setting up access control list (ACL);

Access Control List (ACL) is a consistent list of permission or prohibition rules that apply to higher-level addresses or protocols. ACLs allow you to effectively monitor incoming and outgoing network traffic. ACLs can also be configured for all network routing protocols.

ACL is an array of IOS commands that defines, if router sends package or resets them, based on the information in the package header. ACLs do such tasks of limit network traffic to improve network performance. Access control lists provide the basic security level of access to the network. ACLs can open access to a part of the network of one node and close it to other nodes.

ACLs filter traffic based on the type of traffic.

Access control lists sort nodes to determine if network services access is required. Using ACLs, you can allow or deny access to certain types of files, such as FTP or HTTP.

4. Testing software

Hping3

Hping3 is a free package generator and analyzer for TCP/IP protocol. Hping is one of the mandatory tools for auditing security and testing of firewalls and networks. Like most tools used in computer security, hping3 is useful for security experts, and is used to:

- Traceroute/ping/probe hosts;
- test firewall rules;
- testing of IDS (intrusion detection systems);
- Network research;
- Stress tests of the network;
- study TCP/IP (Hping was used in AFAIK network courses);
- writing real programs related to TCP/IP testing and security;
- automate tests for traffic filtering;
- create a working model of exploits;
- Network and security intelligence studies when emulating the complex TCP/IP behavior;

Zenmap

Zenmap is the official GUI for Nmap Security Scanner. Zenmap is an open source utility for network research and security testing. It was designed to quickly scan large networks, although it works well for single purposes. Zenmap uses IP-packages in its original ways to determine which hosts are available on the network, which services (program name and version) they offer, which operation systems (and OS versions) they use, which types of package filters/ firewalls are used and other features. Zenmap is commonly used for security testing, but many network and system administrators find it useful for common tasks such as network structure control, service scheduling management, and host/service time tracking.

Nmap Output is a list of scanned targets with additional information for each one. The main information is the "important ports table". This table contains the port number, protocol, service name, and state. The state may be “open”, “filter”, “close”, or “unfilter”. “Open” state means that the application on the target machine is ready for connection/ accepting of package to this port. “Filter” state means that the firewall or some network filter in the network blocks the port, and Nmap can`t say if the port open or closed. “Close” ports are not associated with any application, so they can be opened at any time.

Wireshark

Wireshark is a network traffic analyzer. Its task is to intercept network traffic and detail display it. Wireshark can intercept the traffic of various network devices, displaying its name (including wireless devices). Supporting of any device depends on many factors, such as the operation system, and has many protocol decoders (TELNET, FTP, POP, RLOGIN, ICQ, SMB, MySQL, HTTP, NNTP, X11, NAPSTER, IRC, RIP, BGP, SOCKS5, IMAP4, VNC, LDAP, NFS, SNMP, MSN, YMSG and others). Wireshark allows you to save and open previously

saved network traffic. System administrators use it to solve network problems, developers' use it to debug network applications, and ordinary users use it to study the network protocols.

5. Practical implementation of the cyber security polygon on the GNS3 applicable software

For the construction of a cybersecurity polygon based on the GNS3 applicable software, a distributed simplified network topology for small enterprises has been selected. This network topology uses one Cisco ASA 5520 firewall, which divides the company's network into a demilitarized zone, an internal and an external network. The zonal model is fairly flexible, interfaces are assigned to zones, and the checking policy is assigned to the traffic that is transmitted between zones. The network topology built with GNS3 is shown on Fig. 1.

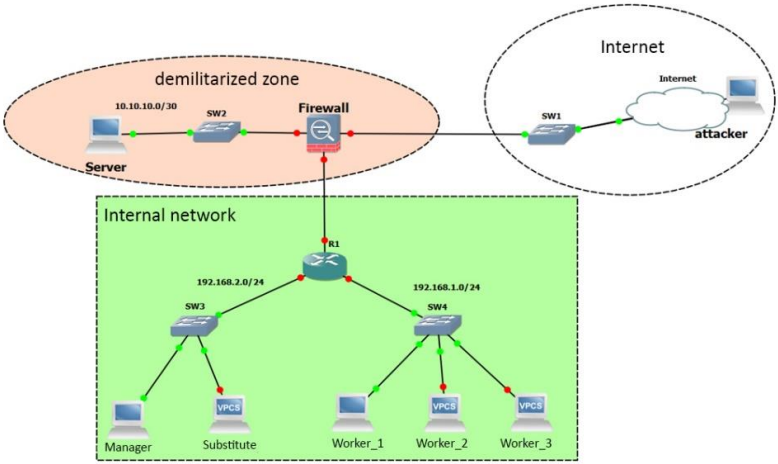


Fig. 1. Network topology in GNS3

The virtual network routing table is shown below in Table2. Our topology of the network is divided on the subnet, all network devices have their interfaces for connecting with IP address, subnet mask and default gateway.

Description of network devices

In this topology, for the construction of a cyber security polygon, the following network devices are used: the Cisco ASA 5520 firewall (hostname: Firewall), the Cisco 3745 router (hostname: R1), network switches (SW1, SW2, SW3, SW4). The loopback interface (Internet) is

used to access to the Internet. Also there are two virtual machines with the operation system Windows XP SP3 (Manager PC, Worker_1-PC) to simulate real computers in the network. Virtual PC Simulator (VPCS) is used to simulate other PCs in the network (Worker_2-PC, Worker_3-PC and Substitute-PC). VPCS uses insignificant PC resources. VCPS creates virtual interfaces for its work and uses only 2 UDP ports - one for transmitting information, another for reading. VCPS has quite limited functionality, but it can save PC resources.

Table 2. Virtual network routing table

Device	Interface	IP-address	Subnet mask	Default gateway
Firewall	Gig0	192.168.125.130	255.255.255.0/24	192.168.125.1
	Gig1	20.20.20.1	255.255.255.252/30	-
	Gig2	10.10.10.1	255.255.255.252/30	10.10.10.2
R1	Fa0/0	20.20.20.2	255.255.255.252/30	20.20.20.1
	Fa1/0	192.168.2.1	255.255.255.0/24	-
	Fa2/0	192.168.1.1	255.255.255.0/24	-
Internet		192.168.158.1	255.255.255.0/24	-
Manager-PC	Net. adapter	192.168.2.10	255.255.255.0/24	192.168.2.1
Substitute-PC	Net. adapter	192.168.2.11	255.255.255.0/24	192.168.2.1
Worker_1-PC	Net. adapter	192.168.1.10	255.255.255.0/24	192.168.1.1
Worker_2-PC	Net. adapter	192.168.1.11	255.255.255.0/24	192.168.1.1
Worker_3-PC	Net. adapter	192.168.1.12	255.255.255.0/24	192.168.1.1
Server	Net. adapter	10.10.10.2	255.255.255.252/30	10.10.10.1

Description of network topology and network device tasks.

The managers' network and workers network is located on the firewall interface Gig1. The router R1 routes traffic between them. Cisco ASA configures NAT to prevent and limit external requests to internal hosts. Firewall is configured to connect to the ASDM via HTTPS, for more management and monitoring. Access to Cisco ASDM is done directly through a Web browser from any Java-based network computer, thus allow security administrators quickly and securely to access Cisco ASA security devices and have a similar functionality as the console connection. For safe remote connection to the R1 router, it is configured with SSH version 2, and all other non-SSH connections are forbidden. In the demilitarized zone there is a server that acts as a Web

server and an FTP server. The web server is accessible for Managers and Workers, and from the Internet. Users have access to the Internet only through the 80th port and FTP. Other ports are closed. From the Internet the access is open only to the Web site and only to the 80th port.

Cisco ASA 5520 specifications.

The Cisco ASA 5500 is the latest multifunction device that uses the most advanced information security technologies, by combining proven products from Cisco: firewall, network intrusion prevention, network anti-virus and VPN services. The Cisco ASA 5520 can enhance the security of applications by blocking network worms, by using AIP SSM - a high-performance intrusion prevention system, by using CSC SSM - a full-featured anti-malware service.

The Cisco ASA 5520 has the following specifications:

- Bandwidth: 225 Mbps
- Simultaneous VPN session: 750
- Simultaneous SSL VPN Scripts: 750
- Interfaces: Four 10/100/1000 Ethernet ports, control port, two USB ports

Cisco 3745 specifications.

The Cisco 3745, is a high-performance router with modular architecture, allows you to implement integrated data, voice, and fax transmissions over TCP/IP networks. It has the following technical characteristics:

- Bins for network modules: 4 pcs.
- Bins for WAN Interface Cards (WIC): 3 pcs.
- Embedded Local Area Network (LAN) ports: 2 10/100 Ethernet ports
- It is possible to use a backup power supply
- Performance - 225.000 packs/second.

Possible modules for Cisco 3700 series routers: LAN modules, WANs, combined network modules, voice and service modules.

6. Network scan and testing

The Zenmap program is the official GUI (Graphical user interface) for Nmap Security Scanner, network exploration utility and port scanner. Ranges of IP-addresses are selected for save time of scanning process. We was got the information about the network from global network, as a result of the NAT setting, only the configured Cisco ASA interface is displayed, as shown in Figure2.

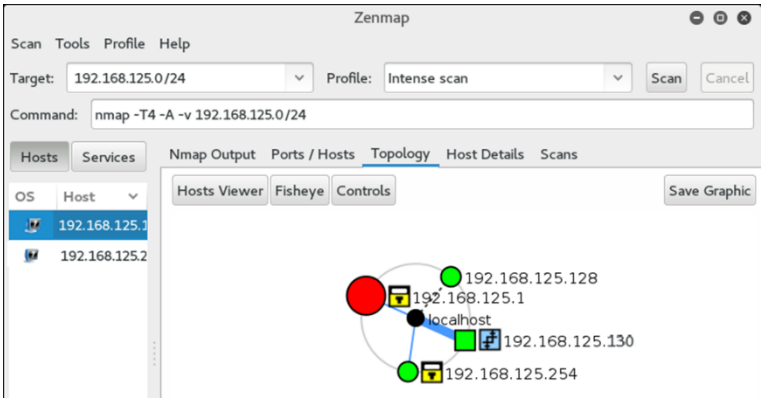


Fig. 2 Local network scan from the external network

If an attacker has access to the network from within, then he will be able to scan the topology of the network and find possible vulnerabilities. The utility allows you to determine which hosts are available in the network (scanned network topology is shown in Figure 3), the version of the operation system, running services, the names of running applications and ports number and ports state, for example, if you do not disable the standard connection to the router through Telnet protocol, then Zenmap will detect it, the result is shown in Figure 4.

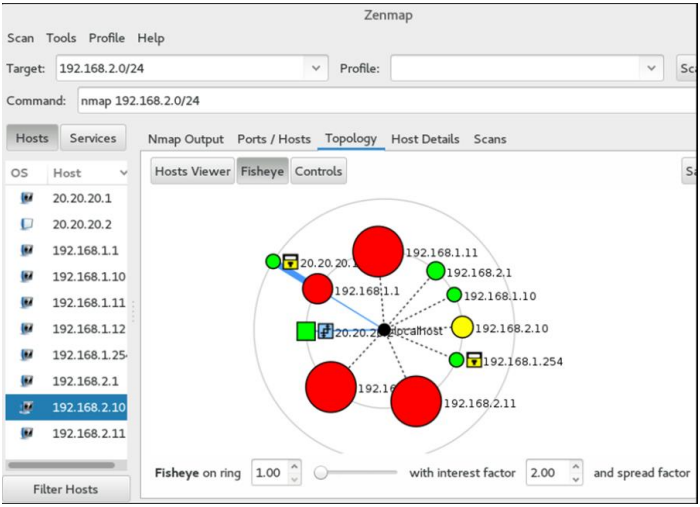


Fig. 3 Local network scan from the inside

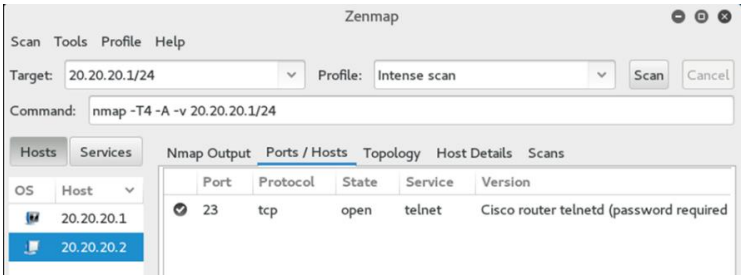


Fig. 4 Scan of Cisco 3745 Router with enabled Telnet

In general, after a detailed scan of the network device, we will have the next result: the Cisco 3745 utility with enabled Telnet, the operation system, open port, network address, as shown in Figure 5.

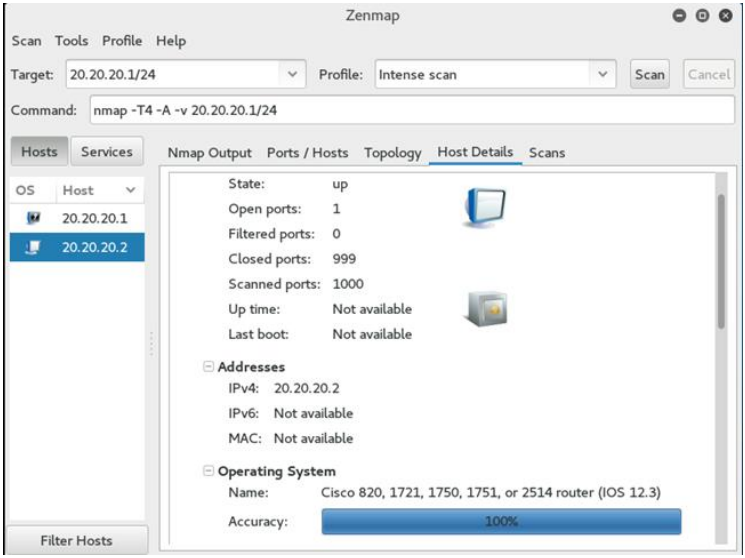


Fig. 5 Zenmap detailed information about scanned host

Cisco ASA has a scanning protection function, so if the messages contain the same source address, this message may be about gathering basic information or about trying to scan ports. Such IP package is rejected by the ACL, as shown in Figure 6.

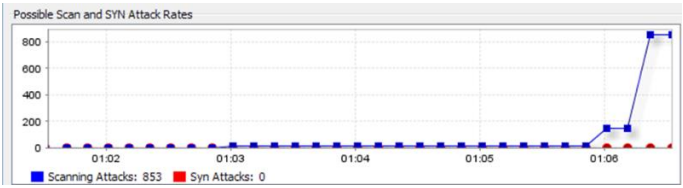


Fig. 6 Cisco ASA rejected package statistics

Network Stress-Test

In computer terminology, denial-of-service attack (DoS) or distributed denial-of-service attack (DDoS) is an attempt to make the machine's resources inaccessible to users. However goals of the DoS can be different, its main purpose is to suspend the services of the Internet-connected host for some period of time.

One of the common methods of attack is the saturation of the target machine with external connection requests, so it can not respond to legitimate traffic or corresponds so slowly that it is seams unavailable.

Syn-Flood Attack is an attack where an initiator puts a fake Source IP-address in the SYN package or ignores replies from the server - Syn + Ack. When you open thousands of such half-hearted sessions, the resources of the server are consumed, server is forced to memorize the parameters of such session and as a result may refuse.

For DoS attack, we have used the Hping3 tool (Kali Linux tool), by using the random IP address for DoS source. This program has no graphical interface. The following commands are selected for the Syn-Flood Attack, it is shown in Fig. 7:

-hping3 - the name of the application.
-c 10000 - the number of packages to send.
-d 120 - the size of each package that will be sent to the target machine.
-S - send only SYN packets.
-w 64 - the size of the TCP window.
-p 80 - destination port, you can use any port.
-flood - send package as fast as possible without taking care of displaying incoming package (Syn-Flood Attack).
-rand-source - use of random source IP addresses. You can also use -a or -spoof to hide hostname
-192.168.125.130 - IP address of the target machine. You can also use the site's address.

```
root@kali:~# hping3 -c 10000 -d 120 -S -w 64 -p 80 --flood --rand-source 192.168
.125.130
HPING 192.168.125.130 (eth0 192.168.125.130): S set, 40 headers + 120 data bytes
hping in flood mode, no replies will be shown
```

Fig. 7 DoS attack with Hping3

Cisco ASA automatically transmits packages that arrives on the server by port 80, and in case of Dos attack the server is under additional load, which means a denial of access to ordinary users.

We have pinged the network without loading and the network under load. We have also used the Wireshark program to analyze Ethernet network packages or other networks (sniffer). We can analyze a DoS attack by using the Wireshark. After filtering traffic between an attacker and Cisco ASA by ICMP criterion, it can be seen that due to a DoS attack, the Cisco ASA becomes under the load, so processing requests queues are created and responses come with a certain delay or response absolutely absent, as shown in Figure 8.

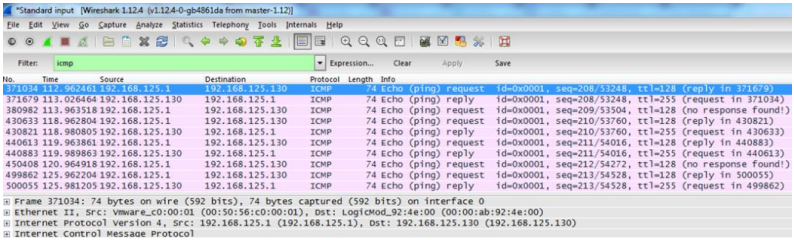


Fig. 8 Filtered ICMP traffic during a DoS attack

To solve this problem, ASA uses TCP SYN Cookies: ASA protects the server and does not broadcast all connections to it. Instead of remembering all these half-sessions, the ASA responds to each of them, but the actual connection to the server only occurs after receiving Ack's 3rd response. Embryonic-conn-max 5 means that the maximum resolution is up to 5 half-connections. You need to set the following settings:

access-list outside_mpc line 1 extended access tcp for any object dmz-server real

class-map no-syn-flood-class

match access-list outside_mpc

policy-map NO-SYN-FLOOD

class no-syn-flood-class

set connection conn-max 0 embryonic-conn-max 5 per-client-max 0 per-client-embryonic-conn-max 0 random-sequence-number enable

service-policy NO-SYN-FLOOD interface outside

Without additional settings for Syn-Flood attack, we have 1625 active connections to the server, so as a result we have a denial of service, as shown in Figure 9.

With such protection settings, the ASA will create a separate queue for half-sessions, it will not be able to affect the server, and as a result we get only 5 half-connections, as shown in Fig. 10. It indicates the protection efficiency from shown attack type with the correct setting of the firewall.

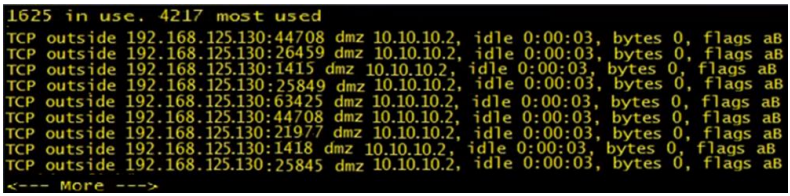


Fig. 9 SYN-flood attack active sessions without protection settings

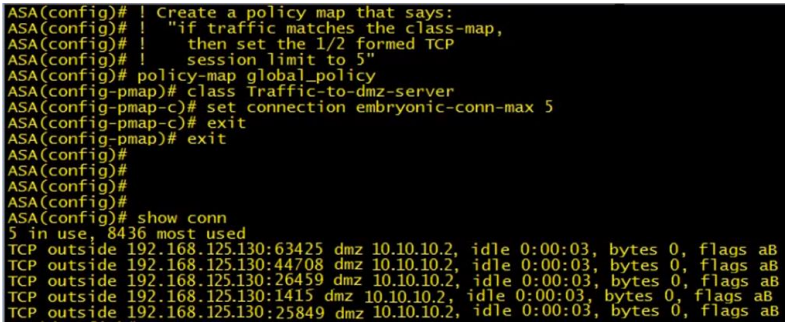


Fig. 10 SYN-flood attacks active sessions with protection settings

7. Conclusions

The ways of constructing a virtual cybersecurity polygon on different platforms have been described, and received conclusions for each of the emulators based on the advantages and disadvantages each of them.

The topology of the cybersecurity polygon have been developed and implemented on the basis of the GNS3 application software. The topology of a computer network consists of a Cisco ASA 5520 firewall that divides the company's network into a demilitarized zone, an internal and an external network. The router Cisco 3745 has been configured routing traffic between Workers and Management networks and remote connection through secure protocol SSH version 2. In a demilitarized

zone, a Web server have been configured to access both from the external network (Cisco ASA automatically transmits packages to the server by 80 port) and from the internal network. Also the FTP server is been configured for only internal network access. Cisco ASA has configured static NAT, for Internet access to the local server, and for external forbidden access to the internal network and also has the Cisco ASDM graphical interface. The network settings have been corrected.

The cybersecurity polygon have been tested by network scanning and network devices ports scanning with Zenmap utility. As a result of the ASA and NAT settings, the external scanning gave only information about IP address of the ASA interface, and the internal scanning presented all information about an internal network including an internal Cisco ASA interface.

The Hping3 utility has been used for a network stress-test. The Syn-Flood Attack has been implemented and shown counteraction ways to this attack.

References

1. Korniyenko, B. (2012) Model of Open Systems Interconnection terms of information security. Science intensive technology, № 3 (15), pp. 83 – 89., doi.org/10.18372/2310-5461.15.5120 (ukr).
2. Korniyenko, B., Yudin, O., Novizki,j E. (2013) Open systems interconnection model investigation from the viewpoint of information security. The Advanced Science Journal, issue 8, pp. 53 – 56.
3. Korniyenko, B., Yudin, O. (2013) Implementation of information security a model of open systems interconnection. Abstracts of the VI International Scientific Conference "Computer systems and network technologies» (CSNT-2013), p. 73.
4. Korniyenko, B. (2016) Information security and computer network technologies: monograph. ISBN 978-3-330-02028-3, LAMBERT Academic Publishing, Saarbrucken, Deutschland, 102 p.
5. Korniyenko, B., Galata, L., Kozuberda, O. (2016) Modeling of security and risk assessment in information and communication system. Sciences of Europe, V. 2., No 2 (2), pp. 61 -63.
6. Korniyenko, B. (2016) The classification of information technologies and control systems. International scientific journal, № 2, pp. 78 - 81.
7. Korniyenko, B., Yudin, O. Galata, L. (2016) Risk estimation of information system. Wschodnioeuropejskie Czasopismo Naukowe, № 5, pp. 35 - 40.
8. Korniyenko, B., Galata, L., Udowenko, B. (2016) Simulation of information security of computer networks. Intellectual decision making systems and computing intelligence problems (ISDMCI'2016): Collection of scientific papers of the international scientific conference, Kherson, Ukraine, pp. 77 - 79.

9. Korniyenko, B. (2017) Cyber security - operating systems and protocols. ISBN 978-3-330-08397-4, LAMBERT Academic Publishing, Saarbrucken, Deutschland, 122 p.
10. Korniyenko, B., Galata, L. (2017) Design and research of mathematical model for information security system in computer network. Science intensive technology, № 2 (34), pp. 114 - 118.

ПРАВОВІ ПИТАННЯ ЗАБЕЗПЕЧЕННЯ ІНФОРМАЦІЙНОЇ І КІБЕРНЕТИЧНОЇ БЕЗПЕКИ ПЛАТЕЖІВ

Губська Д.О.
Академія Адвокатури України, Київ, Україна

У статті розглядаються поточний стан забезпечення інформаційної та кібернетичної безпеки платежів на основі аналізу правових засобів та процедур, що застосовуються постачальниками та користувачами платіжних послуг, проаналізовано окремі аспекти організації досудового розслідування злочинів, що посягають на охоронювані законом суспільні відносини у сфері переказу коштів, функціонування платіжних систем та електронних засобів платежу, фіксації та збору електронних доказів, проведення негласних (слідчих) розшукових дій, акцентується увага на потребі запровадження технологій біометричної ідентифікації, як з метою мінімізації ризиків компрометації конфіденційних платіжних даних користувачів, так і з метою уможливлення встановлення особи злочинців органами досудового розслідування, притягнення їх до кримінальної відповідальності та відновлення порушених прав потерпілих.

Швидкий розвиток ринку платіжних систем в Україні супроводжується упровадженням новітніх платіжних послуг та операцій, що останніми роками вже є достатньо поширеними на платіжному ринку ЄС у зв'язку з імплементацією в національні законодавства країн Євросоюзу прогресивних положень Директиви 2015/2366/ЄС «Про платіжні послуги на внутрішньому ринку». [1]

Статистика безготівкових розрахунків Національного банку України підтверджує, що впродовж 2016-2017 року і в Україні зростає частка розрахунків, які здійснюються через мережу Інтернет з використанням платіжних карток та інших електронних засобів платежу. За 2016 рік збільшилася загальна кількість:

• активних платіжних карток – на 1,55 млн шт. (5%) і становила 32,4 млн шт.;

• безконтактних платіжних карток – на 0,54 млн шт. (37,4%) і становила 1,99 млн шт.;

• платіжних карток з функцією електронних грошей – у 6 разів і становила 54 тис. шт. Позитивна динаміка збільшення

безготівкових розрахунків із використанням платіжних карток як за сумою, так і за кількістю свідчить про те, що такий платіжний інструмент, як платіжна картка, набуває все більшого поширення серед громадян саме як інструмент для безготівкових розрахунків, а неотримання готівкових коштів. Упродовж 2016 року спостерігалося збільшення кількості активних платіжних карток та обсягів безготівкових платежів. На початок 2017 року в Україні працювало 45 платіжних систем та 39 систем переказу коштів, ефективна діяльність яких безпосередньо впливає на стабільність фінансової системи країни.

Інтернет-платежі досягли позначки 61% від загальної суми безготівкових операцій з використанням платіжних карток і становили 346 млрд грн. У 2016 році спостерігалося також розширення платіжної інфраструктури. За 2016 рік кількість підприємств торгівлі та сфери послуг, що надають можливість своїм клієнтам здійснювати безготівкові розрахунки, зросла на понад 11% (на 14 674 од.) і становила 145 938 од. Інфраструктура використання платіжних інструментів також зростала та станом на 01.01.2017 налічувала майже 33,8 тис. За підсумками 2016 року ми відзначаємо зростання частки безготівкових платежів до 35,5% (на початок 2016 року – 31%) та зниження рівня готівки в економіці до 13,8% (на початок 2016 року – 14,3%). [2,с. 99-101] [3].

Збільшується зацікавленість користувачів у розрахунках електронними гаманцями та електронними грошима. Активно розвиваються платіжні системи, які забезпечують проведення платежів та розрахунків в мережі Інтернет (такі, як, ApplePay, GoogleWallet, AnroidPay,PayPal тощо) з використанням платіжних карток та електронних грошей. Перевагами таких систем є, зокрема простота, зручність, швидкість розрахунків та можливість здійснення миттєвих транскордонних переказів. Українські громадяни зацікавлені у користуванні електронними інвойсами, електронними гаманцями, мобільним банкінгом та послугами українських фінансово-технологічних (фінтех) компаній та міжнародних систем інтернет-розрахунків.

Закон України «Про платіжні системи та переказ коштів в Україні» є основним спеціальним нормативно-правовим актом в системі законодавства України, що регламентує правовідносини у означеній сфері та термінологічний апарат. *Платіжний інструмент* – це засіб певної форми на паперовому, електронному чи іншому носії інформації, який використовується для ініціювання переказів. До платіжних інструментів належать документи на

переказ та електронні платіжні засоби. У свою чергу, *електронний платіжний засіб* - це платіжний інструмент, який надає його держателю можливість за допомогою платіжного пристрою отримати інформацію про належні держателю кошти та ініціювати їх переказ. Закон не обмежує функціонування платіжного інструменту або електронного платіжного засобу певною однією формою, передбачаючи, що ним є і платіжна *картка* (пластикова або віртуальна), і *мобільний платіжний інструмент* - електронний платіжний засіб, реалізований в апаратно-програмному середовищі мобільного телефону або іншого бездротового пристрою користувача (персонального комп'ютера, планшету тощо). Так, у статті 15 Закону зазначається, що *електронні гроші* - одиниці вартості, які зберігаються на електронному пристрої, приймаються як засіб платежу іншими особами, ніж особа, яка їх випускає, і є грошовим зобов'язанням цієї особи, що виконується в готівковій або безготівковій формі. Стаття 15 Закону також визначає особливості емісії електронних грошей та особливості правового статусу суб'єктів правовідносин, що виникають у зв'язку з обігом та використанням електронних грошей (користувачі електронних грошей, суб'єкти господарювання, комерційні агенти). .[4, 137] Детальна ж регламентація цих правовідносин здійснена на рівні підзаконних нормативних актів Постанов Правління НБУ.

У 2016 році Національний банк України затвердив концепцію з розвитку акцепторного платіжного доручення та працює над законодавчим урегулюванням цього питання, а саме впровадженням єдиних стандартів квитанції та її формування в електронному вигляді всіма постачальниками послуг (т.зв. e-invoicing). Його втілення сприятиме підвищенню контрольованості і прозорості грошових потоків, скороченню тіньового сектору економіки, додатковому залученню коштів у банківську систему та відповідно збільшенню інвестицій в економіку держави, поповненню бюджету за рахунок більш повного оподаткування, зменшенню обсягів розрахунків готівкою та зростанню ВВП. Окрім цього, регулятор продовжує працювати над розвитком системи дистанційної ідентифікації громадян та юридичних осіб BankID, запровадженням міжнародних стандартів платежів IBAN, розвитком безготівкового обслуговування в торговельних мережах. Національний банк всіляко підтримує розвиток та запровадження інноваційних технологій та альтернативних способів безготівкових платежів (таких як впровадження мобільних розрахунків та оплат за допомогою QR-кодів тощо).[2,95]

Водночас у чинній станом на 12.11.17 редакції Закону України “Про платіжні системи та переказ коштів в Україні” не до кінця визначені особливості роботи провайдерів таких послуг та правові умови для використання користувачами, у тому числі фізичними особами, електронних інвойсів, електронних гаманців та грошей, для здійснення безготівкових переказів та розрахунків за товари (роботи, послуги) в режимі он-лайн на території України та для транскордонних переказів на рахунки громадян, відкриті українськими банками та платіжними установами.

Положення Директиви 2015/2366/ЄС «Про платіжні послуги на внутрішньому ринку», поряд з класичними формами переказу коштів, унормовують також такі види платіжних операцій (або операцій, що пов’язані із платежем):

- виконання платіжних операцій у випадку, коли згода платника на виконання платіжної операції виражена за допомогою будь-яких пристроїв електронної комунікації, цифрових або інформаційних пристроїв, платіжних пристроїв, ПТКС, у тому числі за допомогою терміналів з оплати готівкою (термінали «cash-in»), і у випадку, коли платіж здійснюється оператором системи або інформаційної мережі або мережі електронної комунікації, які виступають винятково в якості посередника між користувачем платіжних послуг і постачальником товарів і послуг;
- надання консолідованої інформації про стан рахунку (рахунків), у випадку, коли згода платника на отримання та надання інформації виражена за допомогою будь-яких пристроїв електронної комунікації, цифрових або інформаційних пристроїв, платіжних пристроїв, ПТКС, у тому числі за допомогою терміналів з оплати готівкою (термінали «cash-in»), і у випадку, коли надання інформації здійснюється оператором системи або інформаційної мережі або мережі електронних комунікації, які виступають винятково в якості посередника між користувачем платіжних послуг і постачальником товарів і послуг;
- послуга ініціювання платежу.

Директивою 2015/2366/ЄС визначено також надважливі з точки зору забезпечення інформаційної та кібернетичної безпеки платежів поняття «аутентифікація», «персоналізовані засоби безпеки», «конфіденційні платіжні дані»:

аутентифікація - процедура, за допомогою якої постачальник платіжних послуг може перевірити використання визначеного платіжного інструмента, включаючи його персоналізовані елементи безпеки;

конфіденційні платіжні дані - дані, включаючи персоналізовані засоби безпеки , які можуть бути використані з метою протиправних дій. Для цілей діяльності постачальників послуг ініціювання платежу і постачальників послуг надання інформації про стан рахунку, конфіденційними платіжними даними не може вважатися ім'я власника і номер рахунку.

персоналізовані засоби безпеки - персоналізовані елементи, надані постачальником користувачеві платіжних послуг в цілях аутентифікації;

строга аутентифікація— аутентифікація, заснована на використанні двох або більшої кількості елементів, віднесених до категорії відомих (відомих тільки користувачеві), наявних у розпорядженні (тільки у розпорядженні користувача) **і невід'ємних (відносно користувача),** які є незалежними, порушення одного з яких не впливає на надійність інших, і розроблених з метою захисту конфіденційності даних аутентифікації.

Деякі норми Директиви вже знайшли втілення у Проекті Закону про внесення змін до деяких законодавчих актів України щодо регулювання переказу коштів, 4 листопада 2016 року зареєстрованого під № 5361. [5]. 9 листопада 2017 року проект Закону перереєстровано під № 7270 [6]. Запропоновані законопроектом зміни мають за мету удосконалити регулювання платіжних систем та контроль за операційними ризиками їх учасників, упорядкувати діяльність платіжних установ на ринку послуг з переказу коштів, надати право платіжним установам відкривати клієнтам рахунки й електронні гаманці, випускати платіжні карти й електронні гроші, а також їх обслуговувати, демонополізувати оплату комунальних та подібних послуг за рахунок використання електронних інвойсів, удосконалити порядок здійснення оверсайту платіжних систем, привести національне кримінальне законодавство до вимог європейського законодавства.

У той же час, науковий та практичний інтерес, з точки зору важливості забезпечення інформаційної та кібернетичної безпеки платежів, становлять, зокрема, питання, пов'язані з автентифікацією, тобто процедурою встановлення належності користувачеві інформації в системі пред'явленого ним ідентифікатора. З позицій інформаційної безпеки, автентифікація є частиною процедури надання доступу для роботи в інформаційній системі, наступною після ідентифікації і передує авторизації.

Чи є дані облікових записів (логіни, паролі користувачів) сервісів, що не є суто платіжними за призначенням, але при підключенні платіжних інструментів, по суті стають ними (Google Wallet, Android Pay, Apple Pay), конфіденційними платіжними даними? Безумовно, так. Ті з них, які надані постачальником користувачеві платіжних послуг в цілях автентифікації є персоналізованими засобами безпеки. У якості двох або більшої кількості елементів, віднесених до категорії відомих тільки користувачеві, наявних тільки у розпорядженні користувача, у різних комбінаціях виступають, як правило, номер мобільного телефону, електронна адреса, пароль, встановлений самим користувачем, смс-пароль, електронний підпис, електронно-цифровий підпис тощо.

Традиційну автентифікацію за допомогою пароля називають ще одно факторною або слабкою. Оскільки за наявності певних ресурсів перехоплення або підбір пароля є справою часу. Не останню роль в цьому грає людський чинник— чим стійкішим до взлому методом підбору є пароль, тим важче його запам'ятати людині і тим вище ймовірність що він буде додатково записаний, що підвищує ймовірність його перехоплення або викрадення. І навпаки— легкі для запам'ятовування паролі в плані стійкості до злому є дуже не вдалими. Утім, нерідко навіть при використанні у якості двох факторів одноразові смс-паролі у комбінації з номером мобільного телефону або електронної адреси, за відсутності умислу/необережності самого користувача, несанкціонований доступ до цих даних можуть отримати сторонні особи (наприклад, застосовуючи шкідливі програми, що дають зловмисникам змогу перехоплювати чужі повідомлення та дзвінки). Поливанюк В.Д. виділяє 4 групи способів перехоплення такої інформації:

1. Перехоплення інформації: безпосереднє перехоплення, і електромагнітне перехоплення.

2. Несанкціонований доступ до інформації: "комп'ютерний абордаж", "містифікація", "маскування", "непостійний вибір", "за хвіст", "аварійний", "за дурнем", "пролом", "люк", "склад без стін", "системні роззяви".

3. Маніпуляції з інформацією: "троянський кінь в ланцюгах", "троянська матрьошка", підміна даних, підміна коду, комп'ютерні віруси, "салямі", "логічна бомба", "асинхронна атака", моделювання, "повітряний змій", "пастка на живця".

4. Отримання і використання інформації із злочинною метою: крадіжка програмного забезпечення, крадіжка устаткування за

допомогою комп'ютерних операцій, крадіжка грошей за допомогою отримання секретних кодів, крадіжка грошей шляхом комп'ютерного пересилання, крадіжка грошей шляхом маніпуляції з комп'ютером, внесення змін в програму. [7]

При порушення кримінальної справи, у більшості випадків таке втручання при проведенні комп'ютерно-технічної експертизи відповідно до п. 1.2.2 Інструкції про порядок призначення та проведення судових експертиз та експертних досліджень, фіксації достатніх та належних електронних доказів в порядку передбаченому ст. ст. 84, 85 КПКУ [11,88], можна довести та встановити IP-адреси, IMEI-коди пристроїв, з яких вчинялися протиправні дії. Однак, навіть встановлення пристроїв, з яких вчиняли протиправні суспільно небезпечні кримінально карані дії зловмисники, поки що не завжди дає змогу встановити особу цих злочинців та притягнути їх до кримінальної відповідальністю за ст.ст. 200, 190, 185, 361, 361-1, 231 ККУ (залежно від фактичного складу злочину) [10, 131].

Злочин, кримінальна відповідальність за який передбачена ст. 361 ККУ, вважається закінченим з моменту настання одного або декількох із зазначених у статті наслідків: 1) виток інформації; 2) втрата інформації; 3) підробка інформації; 4) блокування інформації; 5) спотворення процесу обробки інформації; 6) порушення встановленого порядку маршрутизації інформації. Факт перегляду інформації за результатом несанкціонованого доступу можна кваліфікувати за ст. 361 у разі настання наслідку ─ витоку інформації з обмеженим доступом. Одержання несанкціонованого доступу правопорушником, у більшості випадків, блокує інформацію (або інформаційну послугу) призначену для законного користувача (наприклад, несанкціонований доступ з використанням чужих паролів). [8,20]. Якщо ж несанкціоноване втручання призвело до витоку інформації про конфіденційні платіжні дані або персоналізовані засоби безпеки користувача, та подальшого переказу на інші рахунки, тобто фактично викрадення грошових коштів, зазначене слід кваліфікувати за сукупністю зі ст. 200 ККУ, а інколи й зі ст. 185, 190 ККУ. [10, 131].

За Wikipedia, елементами, або факторами строгої автентифікації є одночасне використання декількох елементів:

- властивості, якою володіє користувач;
- знаннями — інформацією, яку знає користувач;
- володіння — річ, якою володіє користувач.

Який же елемент безпеки є невіддільним, невід'ємним відносно користувача є таким, який неможливо підробити/перехопити і є його незмінною властивістю? Біометрія (biometrics) -технологія ідентифікації особи, яка використовує фізіологічні параметри суб'єкта (код ДНК, відбитки пальців, райдужну оболонку ока, зображення обличчя, тембр голосу). [9]

Отже, перехід постачальників платіжних послуг, систем інтернет-розрахунків до строгої автентифікації користувачів з використанням біометрії в якості елементу ідентифікації та персональних засобів безпеки, з одного боку суттєво б підвищило рівень захищеності мобільних платежів, Інтернет платежів, платежів в середовищі card-not-present та мінімізувало кількість злочинів з використанням електронних платіжних засобів. З огляду на це, доцільним вбачається економічне стимулювання та законодавче закріплення обов'язковості забезпечення платіжними сервісами біометричної ідентифікації користувачів.

У разі ж вчинення злочину та відкриття кримінального провадження, якщо відомості про злочин та особу, яка його вчинила, неможливо отримати в інший спосіб, слідчим доводиться вдаватися до тимчасового обмеження конституційних прав підозрюваних осіб та проводити негласні слідчі розшукові дії з врахуванням ст. 246, 252, 253, 254, 265 КПК України, для того, щоб зібрати та оцінити належність та достатність електронних доказів Аналіз слідчої практики за 2012 - 2017 рр., по категоріям справ, що пов'язані із вчиненням злочинів з використанням електронних платіжних засобів та/або посяганням на інформаційну/кібернетичну безпеку, вказує на те, що при проведенні негласних слідчих розшукових дій, передбачених ст.263, 264 КПК України, п. 1.11.5. та п. 1.11.6 Інструкції про організацію проведення негласних слідчих (розшукових) дій та використання їх результатів у кримінальному провадженні, а саме:

1.11.5.1. Зняття інформації з транспортних телекомунікаційних мереж, що поділяється на:

- контроль за телефонними розмовами, що полягає в негласному проведенні із застосуванням відповідних технічних засобів, у тому числі встановлених на транспортних телекомунікаційних мережах, спостереження, відбору та фіксації змісту телефонних розмов, іншої інформації та сигналів (SMS, MMS, факсимільний зв'язок, модемний зв'язок тощо), які передаються телефонним каналом зв'язку, що контролюється;

- зняття інформації з каналів зв'язку, що полягає в негласному одержанні, перетворенні і фіксації із застосуванням технічних засобів, у тому числі встановлених на транспортних телекомунікаційних мережах, у відповідній формі різних видів сигналів, які передаються каналами зв'язку мережі Інтернет, інших мереж передачі даних, що контролюються.

1.11.6. Зняття інформації з електронних інформаційних систем без відома її власника, володільця або утримувача полягає в одержанні інформації, у тому числі із застосуванням технічного обладнання, яка міститься в електронно-обчислювальних машинах (комп'ютер), автоматичних системах, комп'ютерній мережі;[11,88;12]

- можуть бути порушені конституційні права та законні інтереси не винних осіб, що визнані підозрюваними у конкретному кримінальному провадженні, при цьому втановлення особи злочинця та притягнення його до кримінальної відповідальності може так і не відбутися.

- у КПКУ та Інструкції недостатньо пропрацьований механізм оформлення та поетапності проведення Н(С)РД.

Щодо не зовсім виважених підходів нормотворців КПК та відомчих документів у своїх наукових працях писали М. Шумило, О. Книженко, С. Кудінов. М. Стащак, В. Уваров, С. Савицька, Д. Никифорчук також досліджували окремі проблемні питання, пов’язані з проведенням негласних слідчих (розшукових) дій.

Застосування ж технологій біометричної ідентифікації в цілях досудового розслідування дозволило б не лише знаходити пристрої, з яких здійснювалося несанкціоноване втручання, а й при проведенні комп’ютерно-технічної, дактилоскопічної експертизи встановлювати конкретних винних осіб та причинно-наслідковий зв'язок між їхніми протиправними діями у кіберпросторі, що вчинялися з цих пристроїв, зокрема і з використанням електронних засобів платежу, та притягувати їх до кримінальної відповідальності

Список літератури

1. Директива 2015/2366/ЄС «Про платіжні послуги на внутрішньому ринку»/ Веб-сайт Єврокомісії. Режим доступу: https://ec.europa.eu/info/law/payment-services-psd-2-directive-eu-2015-2366_en

2. Національний банк України. Річний звіт за 2016 рік. – Режим доступу: https://bank.gov.ua/doccatalog/document?id=49064031
3. Національний банк України. Платіжні системи та розрахунки. Загальні показники розвитку ринку платіжних карток в Україні. – Режим доступу: https://bank.gov.ua/control/uk/publish/category?cat_id=79219
4. Закон України «Про платіжні системи та переказ коштів в Україні». - Відомості Верховної Ради України (ВВР), 2001, N 29, ст.137
5. Проект Закону про внесення змін до деяких законодавчих актів України щодо регулювання переказу коштів № 5361 від 04.11.2016. - Режим доступу: http://w1.c1.rada.gov.ua/pls/zweb2/webproc4_1?pf3511=60425
6. Проект Закону про внесення змін до деяких законодавчих актів України щодо регулювання переказу коштів 7270 від 09.11.2017. - Режим доступу: http://w1.c1.rada.gov.ua/pls/zweb2/webproc4_1?pf3511=62847
7. Криміналістична характеристика злочинів, вчинених у банківській системі з використанням сучасних інформаційних технологій. Поливанюк В. Computer crime research center/ - http://www.crime-research.ru/library/Polivan3.htm
8. Організація розслідування фактів несанкціонованого переказу коштів з рахунків клієнтів банку, які обслуговуються за допомогою систем дистанційного обслуговування: Методичні рекомендації / В.В. Корнієнко, В.І. Стреляний. – Х., 2015. - 71 с.
9. Технології аутентифікації особи за біометричними характеристиками. В.О. Романов, І.Б. Галнелюка, П.. Клочан. - Компютерні засоби, мережі та системи. УДК 381.3.
10. Кримінальний Кодекс України. - Відомості Верховної Ради України (ВВР), 2001, № 25-26, ст.131.
11. Кримінальний Процесуальний Кодекс України. - Відомості Верховної Ради України (ВВР), 2013, № 9-10, № 11-12, № 13, ст.88.

12. Інструкція про організацію проведення негласних слідчих (розшукових) дій та використання їх результатів у кримінальному провадженні, затверджена 16.11.2012 Наказом № 114/1042/516/1199/936/1687/5. – Режим доступу: http://zakon2.rada.gov.ua/laws/show/v0114900-12

STRUCTURAL MODEL OF ROBOT-MANIPULATOR FOR CAPTURE OF NO-COOPERATION CLIENT SPACECRAFT

Humennyi D.[1,2], Khlaponin Yu[2], Parkhomey I.[1,2], Rudnitska O.[2]
[1] – National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine
[2] - Kyiv National University of Construction and Architecture, Kyiv, Ukraine

In this work is represented conceptual model of robot-manipulator for capture and holding no-cooperation client spacecraft, which has Payload Adapter interface PAS 1666 S, PAS 1194 C, PAS 1666 MVS, PAS 1184 VS, when there are dynamic errors of linear and angular position of client spacecraft in the interval +/-5 per minute and +/-0.1 meters per minute respectively.

1. Introduction

Today on the geostationary Earth orbit (GEO) there are about 1500 satellites, 750 of which require reorientation, motion or another service operation [5]. Mostly all spacecrafts (SC) with weights more than 500 kg were placed into GEO by medium-lift and heavy-lift launch vehicle and were equipped by Payload Adapter, which is compatible with PAS PAS1666 S, 1194 C, PAS 1666 MVS, PAS 1184 VS. In most cases such design of SC and its control system did not provide docking and orbital motion in GEO after initial adjustment. Therefore such vehicles are not equipped by automatic docking means (IDBS) and need special methods, instrumentality and scripts for this operations.

The analysis of next reports: Department of Mechanical Engineering Massachusetts Institute of Technology [7], International Astronautical Congress [8], University of Nebraska - Lincoln U.S. Air Force Research U.S. Department of Defense [9] and publications [4], [5], [6] showed that at the moment there are no docking means for service spacecraft (SSC) with no-cooperation spacecraft (NCSC) in automated or automatic modes. Reasons for this are complexity of such operations, their cost, lack of hardware-technical solutions, single standards of interfaces and functioning protocols of equipment in situations, which arise during process of approach, berthing, docking, undocking, unberthing and projection of SC. [6]. Docking with NCSC in manual mode is considered in works [4], [5]. However, duration of this operation is longer than five hours and needs both presence of cosmonaut (astronaut) and complex equipment on SSC.

Thereby, it is obvious that with further evolution of Astronautics execution of docking operations with NCSC will appear more often and involvement of human as an object, which controls docking process, will not have economical, social or scientific sense.

2. Problem formulation

Process of docking NCSC with SSC is performed by using a script, which is presented in table 1. This script includes nine stages. Five of them are essential for orbital docking of SC [4] and must be fulfilled automatically.

Stages 1, 2 have turn-key solution [12], [13], and stages 3, 4 were carried out within Space Orbiter programs and described in publications [12], [13], [14], [15]. Stages 6-8, which provide process of "soft" docking, are performed only in manual mode. Process of their automation needs development of new control means and systems.

Table 1. Sequence of operations, which are held during docking with NCSC

№	Stage name	Stage description
1	NCSC position determination	Determination of the positional relationship of NCSC and SSC with GPS navigation and radio-location tools.
2	Approach of NCSC and SSC to a distance 200 m.	Approaching SSC to a distance of the first suspension for the preparation of docking equipment
3	Approach of NCSC and SSC to a distance 20 m.	Approaching SC to a distance of a second suspension for testing docking and fixing equipment of apparatus' continuation.
4	Search for the docking plane on NCSC	SSC flight-around NCSC for searching of the docking plane. Determination of the position of docking port. Comparison of apparatuses dynamic behaviors.
5	Berthing SSC to NCSC on a 2 meters distance	Approaching SSC to NCSC on the service distance with further approach. Positional relationship and orientation of apparatuses are coordinated.
6	Determinati on the position of docking point	Optical, analytic and radio-locating determination of docking port on NCSC's docking plane. Estimate of entering cone.
7	Capture and	Jogless containment of NCSC with

	retention of NCSC by SSC tools	mechanical tools of SSC and further fixation of linear and angular displacement of NCSC.
8	Matching of electric and mechanic parameters of apparatuses	Matching of rotation angles in relative planes electric level 0 V and digital interfaces.
9	Unberthing of SSC from NCSC	Unberthing of SSC on a distance of 20 m with further return to the previously specified orbital position.

3. Conceptual design of robot-manipulator

Suitable Docking port (DP) in modern SC is nozzle of apogeum motor (fig. 1) and SpaceCraft adapter ring (S/C) Payload Adapter (fig. 2). Specification on them can be found accordingly in service instructions [1], [2], [3]. This nodes are characterized by high rigidity and coaxiality axis of apparatuses' mass, what allows to motion NCSC with mechanic tools which are located in SSC. [1], [2], [3].

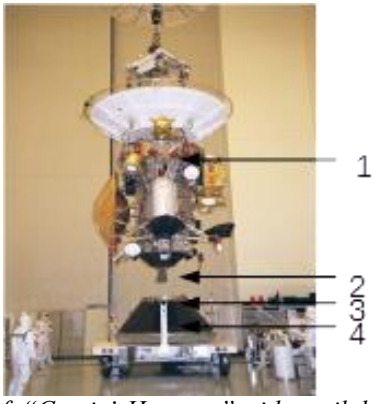


Fig. 1. Spacecraft "Cassini-Huygens" with available S/C adapter ring and nozzle of apogeum motor in process of connecting to PAS: 1 - system block of SC; 2 - nozzle of apogeum motor; 3 - S/C adapter ring; 4 – PAS (photo is used with permission of European Space Agency, Communication Department)

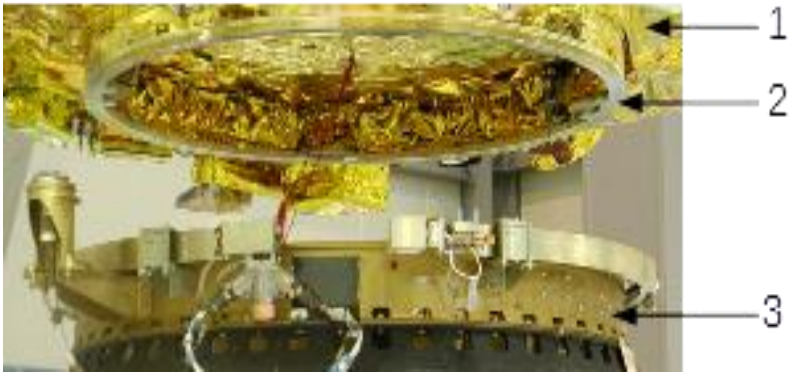


Fig. 2. Connection of SC with Payload adapter PAS 1194 C with means of S/C adapter ring: 1 - system module; SC; 2 — contact plane of SC with PAS by S/C adapter ring means; 3 - Payload adapter (photo is used with permission of European Space Agency, Communication Department)

Capture and holding of NCSC by the listed above ports needs compensation of error relatively to the positional relationship, which is caused by means of radar sensor system. This system is defined by low resolution on the distance of 20 meters [11], and has stationary character and is set by six degrees of freedom (DOF).

Typical design of S/C Adapter, which remains on SC after its unberthing from the NCSC's launch vehicle (fig. 3) has mounting hardware (keys) to Payload Adapter. Keys provides “unberthing” of SC from launch vehicle through pyrotechnic fasteners. After unberthing from launch vehicle S/C Adapter stays on SC and in not in use again for period of active existence of satellite in GEO. Mechanical characteristics of pyrotechnic fasteners' mounting hardware allow holding on NCSC to the S/C Adapter 's mounting hardware in case of uncoordinated docking with further coordination of dynamical and mechanical parameters of butt-jointed apparatuses.

In such a way, for providing automatic capture of SC for S/C Adapter's elements it is necessary to set specification of capture adapter with range of diameters 800-1400 mm, and at the same time save the coaxiality of console, final effector, NCSC and SSC. It is important to note necessity of jogless capture in the conditions of dynamic positioning error of apparatuses. In works [10], [16] the possibility of such a capture is considered however coaxiality of SC is not provided.

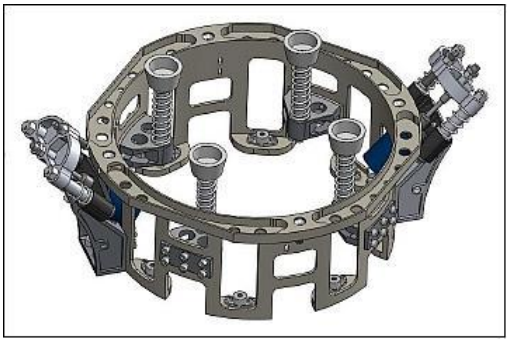


Fig. 3.Typical design of S/C Adapter

Structure of effector has to provide the opportunity of capturing NCSC with consideration of all such factors: the presence of linear and angular error of locating; limited time of being on the distance of berthing of NCSC and SSC; huge amount of S/C Adapter's standards, that differs by sizes and design; the lack of standard markers on NCSC adapter; the lack of friction force between NCSC and environment.

Then, with taking to the account the above, capture of NCSC by RM has to meet the requirements, specified in table 2 and described in [4].

Table 2. Conditions to th opportunities of RM for capture of NCSC

№	Description	Value
1	Distance between the base	1.5-3 m
2	Mutual orientation	Quasispherical
3	Relative angular speed of	to 10˚ per 1 min
4	Relative linear speed of	to 0.1m per 1
5	Zone of positional	to 0.1m
6	Permissible dimensions of	0.6-3m
7	Permissible mass characteristics of SC	to 5000 kg

Conditions for SSC, which are listed in table 2 (it. 1 and 2) are caused by specifications of NCSC, in particular by generalized length of satellite's solar batteries, location of antennas, radius of SC's entering cone, linear and angular errors of positional relationship of SSC and NCSC. In such conditions, delivery of RM's effector to S/C Adapter is

available by using telescopic console, which is equipped by two corner hinges: double-axis hinge 1 and triple-axis hinge 2, which are located in places of console fastening to NCSC (p. A) and final effector (p. B) in accordance, which kinematic structure is shown on fig 4.

Double-axis hinge 1 and telescopic link of robotic console 3 provides work of final effector in polar coordinate system. Triple-axis hinge 2 provides cardan joint of final effector. That allows to compensate errors of its locating. Such kinematic structure of robotic console prevents the emergence of link's singularity, however does not provide problem of final effector locating in S/C Adapter area for committing the jogless NCSC capture. Solution of this problem is possible only by equipping the effector by sensor means, which are suitable for cooperation and relative position of final SSC effector and NCSC S/C Adapter.

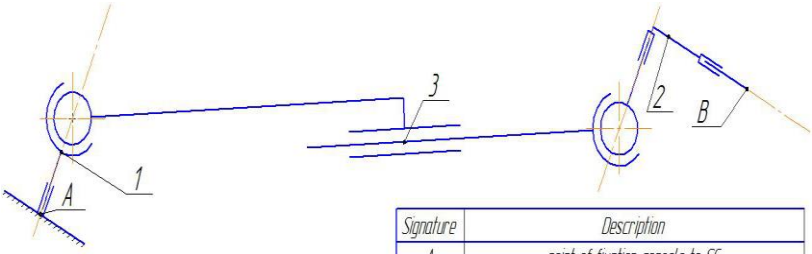
Considering available linear and angular errors of apparatuses positional relationship, guaranteed effector positioning for its further girth and capture of S/C Adapter is possible in segment of work zone (fig. 5 it. 1). Opportunity of capturing NCSC with deviation close to zero although possible in areas, which are shown on fig. 5 (it. 2 and 10). Another zones, which are shown on fig. 5 are assigned for equipping SC or used as a part of technical operations of apparatus work.

Zone of insensibility (table. 2, it. 5) and odd of S/C Adapter's radii (table. 2, p. 6) impose restrictions on the choice of the effectore's design and set of sensor system means. “Zone of insensibility” implies impossibility of sensor system means, which are located within the SSC control system (CS), give accurate coordinates of distance and orientation of NCSC. Error with per minute can cause a collision of effector with S/C Adapter and give it relative acceleration.

It is possible to compensate error on account of effector's physical influence on S/C Adapter after its girth. Adaptation to the diameter of PAS is reached by usage of effector with collet principle of a squeeze. Kinematic structure of such effector is shown on fig. 6. Contact area of effector has unchangeable form because of unification of key design in PAS S/C Adapter. Form of the adapter's key is shown on fig. 7.

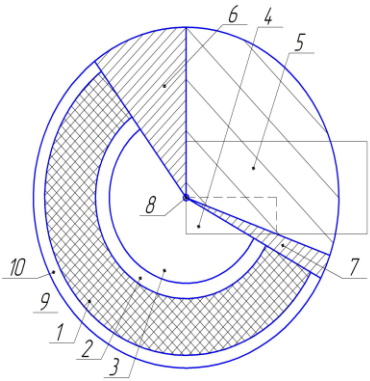
Jogless capture of PAS S/C Adapter is achieved by compensation of force, which effector has applied, and opposite jet force, which adaptor has applied to effector. This effect is reached due to limitations that are imposed by effector's links, after effector girth adaptor until the approach of capture. Boundary values of the squeeze of S/C Adapter are determinate by the selected algorithm of capture, force sensors, which are integrated in the composition of effector's actuators and sensors of

optical interruption of the ray (fig. 6, s2), which goes off when the ray crosses elements of S/C Adapter. Prevention of not full capture (fig 6, S/C1) is provided by coordinated action of SSC's vision means.



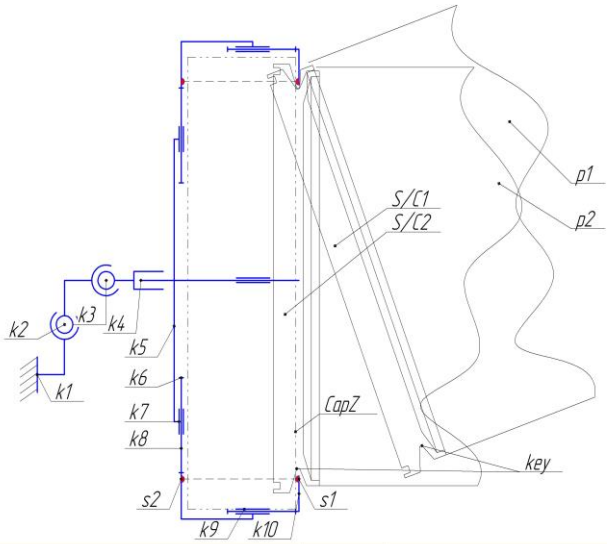
Signature	Description
A	point of fixation console to SC
B	point of fixation effector to console
1	double-axis hinge
2	triple-axis hinge
3	linear conection of telescopic pare of links

Figure 4. Kinematic structure of telescopic console



Signature	Description
1	Work zone of RM
2,10	Restriction of access angles to the work zone
3, 9	Outside the work zone
4	Outside the work zone Container with RM as a part of SC
5	Outside the work zone Design of SC
6, 7	Outside the work zone Equipmet of SC
8	Base point of RM

Figure 5. Operating area of RM's effector in the pale of longitudinal section



Signature	Description	Signature	Description
k1	Base point of effector	s1	Source of laser ray
k2, k3, k4	corner hinges of effector's orientation	s2	Optical diode
k5	Base link of kinematical link pair of effector	S/C1, S/C 2	Position of S/C Adapter in the groove of effector and behind the groove
k6	Damper system	p1, p2	Body of SC in case of S/C1 and S/C2
k7,k9	Translational connection of links	key	Element of key for S/C1 and S/C2
k10	L-shaped link	CapZ	Field of capture/girth of the object

Figure 6. Kinematic structure and main node of such collet effector

Residual fixation during capture is due to reduction of the free movement zone of adapter within the board lines of effector's zone (fig 6, CapZ), what comes as the result of translation motion in the matching hinge (fig. 6, k7, k9).

Process of S/C Adapter's fixation is accompanied by mutually rotation of effector, that prevents angular momentum between effector and adapter. Such rotation is provided by hinges, which are shown on fig. 6, (k2, k3, k4).

4. Docking program of SC

In docking program of SCs, that provides joint movement of apparatuses, prerequisite is prior alignment of their axes. At the same time, design of apparatuses provides axial of mass center point (MCP), that provides positional relationship of shared MCP on a thrust vector of the main SC's engine. For the convenience, process of berthing is divided into two phases:

berthing with capture;

orientation of SC by one shared axis.

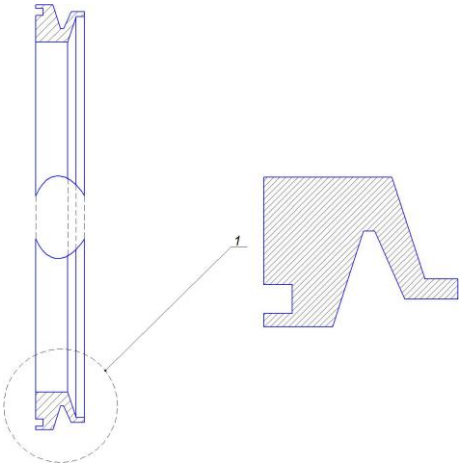


Figure 7. Design of key in composition of adapter PAS C/S Adapter: 1 — groove of a key

During the first phase RM CS initiate execution of effector preparation operations for capture. Occurs adjustment of effector's and S/C Adapter axes.

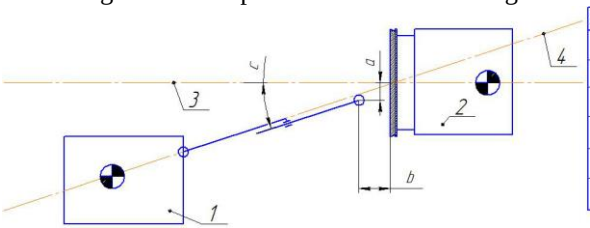
Berthing SSC to NCSC, their positional relationship is set by service spacecraft CS, however CS sets positional error, which can call invariant of SC position. In this case there is a need for error compensation through mains of console and effectore, which are part of the robot-manipulator.

Through pair of hinges, which are placed on the base and on the end of the RM's console and thanks to the telescopic pair of links, final effector can be positioned in the plane of S/C Adapter. Control of effector positioning is possible due vision means. Process of compensation of linear and angular errors within the plain is shown on fig. 8.

It is significant to note that on the stage of berthing robot's effector does not interact with NCSC, which leads to preservation of the Newton's first law.

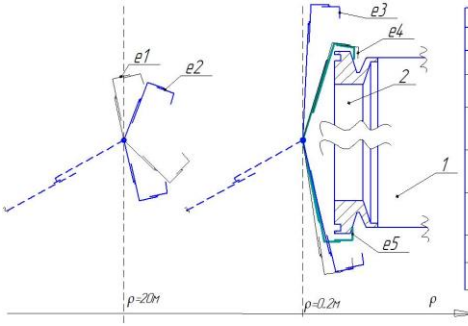
Capture procedure provides cooperation of SSC and NCSC mass center points (MCP) without their displacement through forces, which are applied by effector and are equal by value and opposite by direction. During the berthing RM can act on NCSC with essential force, which in

the case of inaccurate capture, will give NCSC uncontrolled movement. For the capture of NCSC without increase acceleration to it, effector of RM has to cover structural components of S/C Adapter without contacting them. This procedure is shown on fig. 9.



Signature	Description
a,b	linear positional errors
c	angular posotional errors
1	servise spacecraft
2	no-cooperation spacecraft
3	axis of the optimal berthing
4	actual axis of bearthing

Figure 8. Rendering of positional relationship error of SSC and NCSC



Signature	ρ	Description
ρ	-	Distanse between NCSC and SSC
e1	20	Stage 1. Alignment of axes of NCSC and SSC
e2	20	Stage 2. Effector orientation parelel to the plane of S/C Adapter
e3	0,2	Stage 3. Grasping of S/C Adapter by effector with taking into account errors of positional errors
e4	0	Stage 4. Insertion of the capture mechanism in the groove of the S/C Adapter
e5	0	Stage 5. Capture the S/C Adapter by effector through contraction of the effector's links
1	-	S/C Adapter
2	-	No-cooperation spacecraft (NCSC)

Figure 9. Capture procedure S/C Adapter of NCSC by effector of RM

First step is held on the “decision-making” distance (20 meters between NCSC and SSC). On this step initialization operations of sensor system, executive devices and RM CS are performed. On the distance ρ=20 meters effector of RM insert into the manipulation work zone. According to data from RM CS, vision, sensors of links orientation and position concordance of effector's position by three coordinates is carried out. At the same time, angular orientation of effector is remains random. (fig. 9, e1).

On the second step effector places in parallel with plane of S/C Adapter. Second step is carried on the distance of “berthing” (about two meters between NCSC and SSC). On this step takes place calibration of vision according to sensors of orientation and position of RM's links. The S/C Adapter is classified. Algorithm of berthing and capture is determinate (fig. 9, e2).

On the third, forth and fifth steps (fig. 9, e3 e5) operations of girth and capture of structural part of S/C Adapter are carried by effector.

Taking into consideration huge number of S/C Adapter versions, effector can change zone of girth and capture in dependence of determinate algorithm, which was chosen of the second step.

During the second phase CS of RM initiate coordination of SSC and NCSC axes. For this angle (fig. 8) leads to zero. After coordination of axes hinge joint of SSC, RM, NCSC design is fixed, what confine relative displacement. This gives new position of MCP of all the design. Position of MCP coincide with axis of apparatuses (which goes through their apogeum motors). This orientation of the apparatuses allows to use apogeum motor of SSC as the motor of whole construction.

5. Conclusion

Researches, that were held in this work, showed, that procedure of no-cooperated capture and docking operations is economically and scientifically actual engineering problem, and development of space industry of the human menage needs technologies for satellites extirpation, their recycling, execution of mechanical operations etc.

The proposed principle and design conception of service of the spacecraft robot-manipulator allows to carry out capture procedure of spacecraft with wights up to 5000 kilos in conditions of incomplete definition of their spatial position and suggested adapter design, which works on the principle of collet. This conception allows to make girth, capture and holding of SC which is equipped by PAS 1666 S, PAS 1194 C, PAS 1666 MVS, PAS 1184 VS with positional error 10° per minute and

The proposed solutions needs further development and solving of such actual problems:

a) development of kinematic structure of robotic console and effector, which have to consider features of fastening console to basic structure of SSC, and principle approaches to provide linear and angular motion of the links;
b) calculation of angular and linear position of the links and effector, determination of speeds and accelerations of links, calculating moments and determination of work zone and singularity zone;
c) selection of the engines and actuators types, which allows effectively provide linear and angular motion of the colsole's and effectore's links in terms of space. In particular, chosen engines must be characterized by low wight, huge moment, low consumption, working voltage range less that 20 Volt, minimum requirements to climatic conditions and possibility of

accurate refining of angles and rotation speeds. Furthermore, actuators have to provide translational motion of the links in space conditions as well as angular motion of the links in two or three planes;

d) choice of the materials for console's and effector's links has to provide minimal weight with high rigidity. Important is appropriateness of chosen materials in application conditions and with minimal level of capacity, permeance and radio-wave absorption;
e) development of action script of the robot-manipulator system, which contains control algorithm of RM and process of its co-operation with SSC CS ;
f) development of technical equipment and script of output from transfer position;
g) development of mathematical and computer methods of simulation RM actions and control.
h) development of methods and means of vision for RM;
i) development of interaction protocols between RM CS and SSC CS;
j) development of hardware for RM CS, sensor system and power supply means;
k) development of ground part of diagnostic, arrangement and management equipment as well as equipping of auxiliary service;
l) development of marking standards of S/C Adapter and improvement systems for burial of SC;
m) determination of spacecraft's orientation engine location for cases of variable position of MCP.

Acknowledgements

This publication would not be prepared without consultation of colleagues from SOE “Південне”, PrJSC “НВК КУРС” and Department of Technical Cybernetics NTUU “Igor Sikorsky KPI”.

References

[1] Eurockot. Rockot User's Guide, EHB0003, Issue 5, Revision 0. August 2011
[2] Ariane 5 User’s Manual Issue 5 Revision 1. July 2011
[3] Polar Satellite Launch Vehicle User’s Manual Issue-6 Rev 0 No: PSLV-VSSC-PM-65-87/6. March 2015
[4] Гуменний Д. Технічний вигляд ефекторної складової роботизованої системи механічного захвату для

виконання задач орбітального сервісу: міжанродна конференція [“16th Ukrainian Conference on Space Research”], Odessa, 08.2016
[5] Гуменний Д. Розробка концепції та системи управління багатоцільовим роботом-маніпулятором для виконання орбітального сервісу: міжнародна конференція [“15th Ukrainian Conference on Space Research”], Odessa, 08.2015
[6] DARPA-SN-14-51 1 Request for Information (RFI) DARPA-SN-14-51 Robotic On-Orbit Servicing Capability with Commercial Transition Defense Advanced Research Projects Agency (DARPA) Tactical Technology Office (TTO)
[7] West, Harry, et al. "Experimental simulation of manipulator base compliance." Experimental Robotics I. Springer Berlin Heidelberg, 1990.
[8] Skomorohova, Ruslan S., Andreas M. Heinb, and Chris Welchc. "In-orbit Spacecraft Manufacturing: Near-Term Business Cases." (2016).
[9] Flores-Abad, Angel, et al. "A review of space robotics technologies for on-orbit servicing." Progress in Aerospace Sciences 68 (2014): 1-26.
[10] Mohtar, Tharek, et al. "Docking mechanism concepts for the strong mission" (2015).
[11] Севостьянов, Ю. В., Каратєєв С. М. "Пропозиції щодо розробки бортового імпульсно–доплерівського радіолокаційного комплексу з системою фазованих антенних решіткок для військових літальних апаратів." Озброєння та військова техніка 4 (2011): 28.
[12] Langley, Robert D. "Apollo experience report: The docking system." (1972).
[13] Bluth, B. J., and Martha Helppie. "Soviet space stations as analogs." (1986).
[14] Хорольский, П.Г., Дубовик Л.Г. "Методика прогнозирования тактико-технических характеристик космического тральщика." Восточно-Европейский журнал передовых технологий 4.3 (64) (2013).
[15] Polites, Michael E. "Technology of automated rendezvous and capture in space." Journal of Spacecraft and Rockets 36.2 (1999): 280-291. DOI: 10.2514/2.3443

[16] Ткач, М. М. "Керування рівновагою антропоморфного крокуючого апарата за інформацією про екстремуми на поверхні руху/Ткач ММ, Гуменний ДО Стратегии качества в промышленности и образовании." Proc. of Annual Conf. 2012.

ВИЗНАЧЕННЯ ПУБЛІКАЦІЙНОЇ АКТИВНОСТІ В НАУКОВИХ НАПРЯМКАХ, ЯКІ СПРИЯЮТЬ ОБОРОНОЗДАТНОСТІ КРАЇНИ (ЗАХИСТ ІНФОРМАЦІЇ, КОМП'ЮТЕРНА БЕЗПЕКА)

Добровська С.В.
Інститут проблем реєстрації інформації НАН України, м. Київ

Проведено розширений пошук у реферативній БД "Україніка наукова" масиву записів з напрямку «Захист інформації». Досліджено зростання публікаційної активності цієї тематики. Побудовано діаграму динаміки публікаційної активності в базі даних "Україніка наукова" з тематики «Захист інформації». Зроблено аналіз журналів, в яких найбільше публікується статей за напрямом «Захист інформації».

Сьогодні інформаційна боротьба стає однією з основних форм вирішення суперечностей між державами. Стрімко зростає рівень та значно розширюється спектр інформаційних загроз. Маючи системний і цілеспрямований характер, зовнішній негативний інформаційний вплив призводить до появи загроз національній безпеці України в інформаційній сфері, які завдають державі відчутних збитків. Важливість науково-практичних досліджень з напрямку захисту інформації ґрунтується на тому, що безпека інформаційного простору є визначним чинником військово-політичної та економічної стабільності держави, а також захищеності національних інтересів у різних сферах життєдіяльності.

Відповідно до законодавства України поняття "інформаційна безпека" має таке визначення: "стан захищеності життєво важливих інтересів людини, суспільства і держави, при якому запобігається нанесення шкоди через: неповноту, невчасність та невірогідність інформації, що використовується; негативний інформаційний вплив; негативні наслідки застосування інформаційних технологій; несанкціоноване поширення, використання, порушення цілісності, конфіденційності та доступності інформації." [1]

Реферативна база даних (БД) "Україніка наукова", яка створюється Інститутом проблем реєстрації інформації та Національною бібліотекою України імені В. І. Вернадського, на сьогодні є лідером в Україні за обсягом власних інформаційних ресурсів, що дає змогу проведення аналітичних досліджень

поточного стану в науці, техніці та технологіях. Вона кумулює наукову інформацію з усіх галузей знань, забезпечує вільний доступ та пошук інформації, а також архівне збереження фонду БД на компакт-дисках.

Велика ретроспектива та об'єм даних надають можливість застосування методу підрахунку кількості публікацій. Цей метод базується на індикаторі «кількість наукового продукту» (тобто статей, книг, монографій). Його застосовують як для оцінки результатів наукової діяльності окремого вченого, установи і держави в цілому, так і для аналізу потоків науково-технічної інформації з метою визначення тенденцій розвитку. Метод містить етапи: виокремлення локальних груп конкретних потоків інформації; визначення на різних часових відрізках інтенсивності та напрямку потоку, побудова мереж фактичного взаємовпливу за допомогою якісного аналізу; отримання висновків щодо реальних тенденцій процесу розвитку певної галузі науки. Цей традиційний метод застосовується вже довгий час. Активні прихильники застосування публікаційної активності як критерію оцінки науково-дослідної діяльності не відкидають його недоліків, але вважають, що вони не перекривають його переваги [2]. Показник «кількість публікацій» не дає повного представлення про значущість наукової продукції, але той факт, що публікація вводить в науковий обіг певну інформацію, не піддається сумніву. Кількість публікацій є важливим показником для визнання вченого, його внеску в науку. Велика кількість монографій та статей в певній науковій галузі свідчить про її динамічний розвиток, а швидкість зростання — про актуальність.

Поточний стан записів у БД на листопад 2017 року складає 650 109 записів. Досліджено потік наукових публікацій у реферативній базі даних "Україніка наукова" з інформаційної безпеки (захисту інформації). Науковий напрямок «Захист інформації» входить до тематичного розділу «Електроніка. Обчислювальна техніка» серії 2 «Техніка. Промисловість. Сільське господарство» українського реферативного журналу «Джерело» та реферативної бази даних «Україніка наукова».

За напрямом «Захист інформації» проведено розширений пошук у реферативній БД "Україніка наукова" за кількістю прореферованих записів, зокрема, статей в періодичних виданнях в 2005 — 2016 рр. та за певним індексом рубрикатора НБУВ. [3]

За рубрикатором НБУВ в БД “Україніка наукова” “Інформаційна та обчислювальна техніка” має індекс 397. Дані з захисту інформації шукаємо за індексом рубрикатора 3970.40. [3]

Загальна кількість записів в базі даних «Україніка наукова» з тематики “Захист інформації” 1531. Це невисокий показник наповнення бази даних з вузької тематики (0,24 %). Книжкові видання налічують 941, журнали та продовжувані видання – 531, автореферати – 51, наукова періодика України – 7.

Дані щодо кількості публікацій із тематики “Захист інформації” в реферативній БД «Україніка наукова» [3] за роками 2005 - 2016 зведено в таблицю 1.

Таблиця 1 Кількість записів в РБД «Україніка наукова» з захисту інформації

Рік	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
Кількість статей	82	66	140	134	168	197	224	162	140	69	26	22
Кумулятивна КП	183	249	389	523	691	888	1112	1274	1414	1483	1509	1531

Побудовано діаграму динаміки публікаційної активності з напрямку “Захист інформації” (див. рис. 1).

Рис. 1. Динаміка публікаційної активності з тематики «Захист інформації»

Наочно видно нерівномірне наповнення БД кожного року. Пік приходиться на 2011 р. Так, у 2011 р. кількість статей у базі даних збільшилась у 4,3 рази у порівнянні з 2003 р. Але у 2012 – 2015 рр. відбулось поступове зменшення статей з тематики "Захист інформації". За 2016 р. база даних «Україніка наукова» з цього напрямку ще поповнюється.

Також побудовано діаграму кумулятивної кількості публікацій (КП) з тематики "Захист інформації" (див. рис. 2).

Рис 2. Кумулятивна кількість публікацій за роками з тематики «Захист інформації»

Зроблено аналіз журналів, в яких найбільше публікується статей за напрямом "Захист інформації". Найбільша кількість записів за 2005 – 2016 рр. опублікована в провідних журналах (див. табл. 2).

Таблиця 2 – Провідні фахові видання з тематики "Захист інформації"

Журнали	Вид	Рік заснування
Вісник Національного університету "Львівська політехніка"	збірн	1964
Восточно-Европейский журнал	журн	2002

передовых технологий		
Захист інформації	журн	1999
Искусственный интеллект	журн	1995
Кибернетика и системный анализ	журн	1965
Правове, нормативне та метрологічне забезпечення системи захисту інформації в Україні	збірн	2000
Проблемы управления и информатики	журн	1956
Радіоелектроніка. Інформатика. Управління	журн	1999
Радіоелектронні і комп'ютерні системи	журн	2003
Реєстрація, зберігання і обробка даних	журн	1999
Системи обробки інформації	збірн	1996
Системи управління, навігації та зв'язку	збірн	2007
Управляющие системы и машины	журн	1972

Найбільша кількість записів у БД серед провідних журналів з тематики "Захист інформації" припадає на 3 журнали: "Правове, нормативне та метрологічне забезпечення системи захисту інформації в Україні", "Захист інформації" та "Системи обробки інформації". Перші 2 видаються в Києві, 3-й в Харкові. [4]

Висновок. Використання реферативної БД «Україніка наукова» дозволяє проведення наукометричних досліджень, які надають можливість спрогнозувати тенденції розвитку наукового знання з тематики «Захист інформації» в Україні. В результаті дослідження публікаційної активності вчених і спеціалістів України з тематики "Захист інформації" з'ясовано, що протягом 2003 – 2011 рр. спостерігається зростання кількості публікацій з цього напрямку, яке відзначається стрімким розвитком за 2010 — 2011 рр., але потім – спад. Дослідження потоку наукових публікацій показали, що «Захист інформації» є перспективним, а також науковим напрямком, що розвивається в Україні.

Література

1. Закон України «Про Основні засади розвитку інформаційного суспільства в Україні на 2007-2015 роки»

2. Ардашкин И. Б., Сидоренко Т. В. Публикационная активность и ее роль в оценке профессиональной деятельности научно-педагогических работников вузов (российский опыт) – Образование и наука - 2016Б1(30)-DOI: 10.17853/ 1994-5639-2016-1-145-158.
3. Реферативна база даних "Українiка наукова" [електронний ресурс]. — Режим доступу: — http://www.nbuv.gov.ua.
4. Добровська С. В. Аналіз публікаційної активності з напрямку «Захист інформації» за реферативною базою даних “Українiка наукова” // Информационные технологии и безопасность: основы обеспечения информационной безопасности: Материалы международной научной конференции ИТБ-2014. – К.: ИПРИ НАН Украины, 2014. — Вып. 14. — С. 153-159.

АНАЛІЗ ТЕХНОЛОГІЙ ІНФОРМАЦІЙНО-ПСИХОЛОГІЧНИХ ВІЙН ТА ІНФОРМАЦІЙНО-ПСИХОЛОГІЧНОЇ ЗБРОЇ РОСІЙСЬКОЇ ФЕДЕРАЦІЇ

Довгополий А., Білобородов О.
Центральний науково-дослідний інститут озброєння та військової техніки Збройних Сил України, м. Київ, Україна

Класичні війни, проксі війни, гібридні, мережеві, мережецентричні та приватизовані війни є реаліями сьогодення. У наш час економічні, фінансові та інформаційні інструменти мають рушійну силу сумірну, або навіть більшу від застосування традиційної сили.

Вчені та політики РФ протягом останніх 30 років приклали значні зусилля для розробки засад і теорії інформаційно-психологічної війни (ІПсВ) та створення інформаційно-психологічної зброї (ІПсЗ) [1; 2]. Один з розробників Доктрини інформаційної безпеки РФ визначив, що ІПсВ може визначатись як масштабне застосування засобів і методів інформаційно-психологічного впливу по відношенню до населення країни, окремих соціальних груп чи особистостей, та захист від аналогічних дій в свою адресу, реалізованим державою чи іншим актором міжнародної політики для забезпечення реалізації своїх інтересів [3].

Метою доповіді є представлення результатів аналізу технологій інформаційно-психологічного впливу на психіку людини та суспільну свідомість, аналіз інформаційно-психологічної зброї, яка може бути застосована РФ проти України.

1. Погляди на методичні та організаційні основи ведення ІПсВ.

Основні цілі ІПсВ зазвичай є [3; 4]:

- забезпечення прийняття рішень і спонукання до виконання дій влади країни-жертви, які б задовольняли потреби країни-агресора;
- підрив легітимності політичної влади та міжнародного авторитету країни-жертви;
- дестабілізація ситуації у країні-жертві, провокування політичних протестів, соціальних конфліктів, підрив морально-психологічного стану населення країни-жертви;
- підрив обороноздатності країни-жертви та боєздатності її збройних сил;
- підтримка дій внутрішніх сил, направлених на знищення чи завдання шкоди своїй державі, у тому числі шляхом корумпування влади й політичної еліти;

- заміна соціально-культурної ідентичності всього населення або його частини у країні-жертві, зміна національних цінностей та засад державотворення.

В якості основних об'єктів інформаційно-психологічного впливу у країні-жертві є:

політична еліта;

населення країни у цілому;

окремі соціальні групи, наприклад, за національною, релігійною, професійною або іншими ознаками;

окремі високопосадовці, наприклад, очільник країни-жертви, міністри, військові високопосадовці, політичні опозиціонери та інші.

Автори робіт [3; 5] до числа основних організаційних форм ІПсВ відносять:

психологічні операції проти вищого керівництва, населення та військ країни-жертви;

публічну дипломатію, яка є основним способом реалізації так званої "м'якої сили" [5] і складається із зовнішньополітичного інформування, наукових та освітницьких обмінів, програм в області культури та інші форми гуманітарного співробітництва;

зовнішньополітичну пропаганду, націлену на розповсюдження поглядів, ідей і теорій для формування у громадян заданого керівництвом РФ світогляду, понять, які відображають інтереси суб'єкта пропаганди – РФ та стимулюють до відповідних практичних дій [6];

спеціальні таємні психологічні операції, які вносять політичну та соціальну дестабілізацію в країні-жертві.

2. Погляди на технології дестабілізації та методи руйнації легітимної влади країни-жертви.

Узагальнення світового досвіду, наприклад, "арабської весни", кольорових революцій та власні напрацювання вчених РФ дозволяють використовувати технології впливу на масову свідомість для дестабілізації суспільства та підриву легітимної влади країни-жертви.

Пропагандистське бачення світу, жорсткість масмедіа, перетворення їх в психологічну зброю потребують розуміння механізмів впливу на особистість, великі групи людей та цілі народи.

Зусиллями психологів-когнітивістів [7; 8] з'ясовано чому людська психіка сприймає пропаганду і чому люди завжди попадають у пастки медіаманіпуляторів. Природа людини така, що спроби явного і очевидного впливу викликають спротив такому

впливу, тому що людина бачить замах на власну ідентичність. Відповідно до теореми Томаса [8] “щоб спровокувати необхідну поведінку або настрій людей, необхідно створити реальність, яка буде уявлятись людям істинною”. Масштабну реальність для мас людей можуть створити масмедіа. Сфабриковану реальність люди повинні прийняти добровільно та бути впевненими, що це і є їх погляди на світоустрій. Якщо людина розуміє, що нею маніпулюють, то маніпуляція втрачає силу.

Медіаманіпулювання. При всіх відмінностях медіамашин Заходу, РФ та інших держав набір цілей маніпулювання часто однаковий. Відмінності полягають у тому, що західні ЗМІ основані на плюралізмі поглядів і думок. Власники засобів масової інформації створюють, обробляють, спритно оперують і цілком контролюють поширення інформації, яка визначає уявлення людей, їх установки, а в кінцевому рахунку, і поведінку. Навмисно фабрикуючи повідомлення, які спотворюють реальну соціальну дійсність, вони перетворюються в маніпуляторів свідомістю.

Виходячи з узагальнень російських дослідників, ЗМІ повинні дотримуватись деяких постулатів маніпуляцій, основні з яких наступні:

послання до суспільства не повинно розходитись з масовими цінностями і світоглядом цього суспільства;

маніпуляції починаються з підготовки інформаційного фону медіаспектаклю і оцінки реакції суспільства на маніпуляції;

якщо фон медіаспектаклю не викликає сумнівів, тоді він відповідає світогляду суспільства та його упередженим думкам. Упереджені думки – це те, що є в людині до того, як включається критичне мислення.

При медіаманіпулюванні використовуються фактори:

короткої пам’яті людини;

для людей немає об’єктивності, якщо вони втягнуті у конфлікт, який загрожує їх інтересам і цінностям;

людині для осмислення явища необхідно мати назву цього явища, а назва у прихованому вигляді пояснює і програмує реакцію на нього. Через слова і поняття створюється картина світу.

Для забезпечення медіаманіпулювання основними методами поширення інформації є фрагментація та негайність. Фрагментація (дроблення, локалізація) – домінуючий метод поширення інформації. Негайність – репортаж безпосередньо з місця події – на даний час є одним із головних принципів журналістики. Помилкове відчуття терміновості призводить до втрати здатності розмежовувати

інформацію за ступенем важливості. Розумовий процес сортування, який у звичайних умовах сприяє осмисленню інформації, не в змозі виконувати цю функцію. Концентрація уваги на подіях, що відбуваються в дану хвилину, руйнують необхідний зв'язок з минулим.

Роль телебачення у медіаманіпулюванні виходить з того, що слова в багато разів переконливіші, якщо вони підкріплені картинкою або синхронізовані з відеорядом. Дивлячись телевізор, люди несвідомо сприймають подію як таку, що відбувається у них на очах і вони є свідками події. Сфабриковані телесюжети, витончена суміш правди, напівправди та інсценування, все це для глядачів є "достовірною" реальністю [8].

Для подачі "запрограмованих" матеріалів створюються передачі, які зрозумілі за змістом та не потребують інтелектуальної напруги, у той же час будоражать емоції. Переважання емоцій над розумом – це особливість поведінки людини. Ці програми повинні бути зрозумілими 13-14 літнім підліткам [8].

Таким чином, ТБ з його швидкозмінними картинками не залишає шансу осмислити побачене й почуте.

Можливості інтернет-технологій в ІПсВ. Інтернет – це новий інструмент соціальної інженерії з невідомими раніше моделями прийняття рішень, який змінює пізнавальний базис сучасної людини, він визначає зміст інформації, яка доходить до людей по всьому світу.

Інтернет може використовуватись у внутрішній політичній боротьбі та зовнішніми деструктивними силами для збурення суспільства. Крім того, інтернет може використовуватись для антидержавної діяльності агентами іноземного впливу. Технології підриву легітимності влади в інтернет просторі постійно генеруються і вдосконалюються, що створює проблеми для системного протиборства подібній підривній діяльності. Ці технології деформують масову свідомість населення країни-жертви і викликають недовіру, презирство й ненависть до діючої влади.

Не дивлячись на те, що в інтернеті є системні центри генерації контенту, користувачі інтернету і соціальних мереж вважають, що контент створюється такими ж рядовими користувачами, як і вони, тому довіряють цьому контенту і не схильні до перевірки достовірності новин і фактів, наданих у мережах. Молодіжна аудиторія велику частку інформації й інтерпретацію подій знаходить в інтернеті. "Бунтарська істерія" молоді може ефективно

використовуватись для руйнування національної держави за короткий термін.

Політична активність соціальних груп поступово зміщується у соціальні мережі. Інструменти інтернету та соціальних мереж з допомогою мобільних комп'ютерів та засобів зв'язку дозволяють швидко мобілізуватись великим масам людей, отримувати інструкції і діяти синхронно.

Психометрія та Big Data. До початку кібернетизації суспільства у світовій практиці для визначення психологічних можливостей людини використовували результати психометричних досліджень, проведених шляхом анкетування. Після досліджень складався психометричний портрет (профіль) людини, на основі якого прогнозувалась поведінка людини в екстремальних умовах, у колективі, прогнозувалось кар'єрне зростання та інше. Психологи [9; 10] довели, що кожна риса характеру людини може вимірюватись за допомогою наступних п'яти вимірів:

- відкритість (наскільки людина готова до сприйняття нового);
- добропорядність (наскільки людина є перфекціоністом);
- екстраверсія (як людина відноситься до соціуму);
- дружелюбність та готовність людини до співробітництва;
- нейротизм (наскільки легко людина виходить з себе).

Психологічний профіль дозволяє вивчити всі пріоритети людини. Активна фаза впливу на людину полягає у підготовці адресного звернення з інформацією, яку у своїх думках очікує людина, тому можливо очікувати прогнозовану реакцію.

Розвиток інтернету, комп'ютерних баз даних та соціальних мереж дали доступ до інформації про людину без її згоди. Як показано у роботах [10-13] комплексний аналіз інформації про людину з соціальних мереж, інтернету, матеріалів з податкової служби, інформації зі смартфонів та інших джерел дозволяє проводити цільовий психологічний аналіз поведінки людини без анкетування. Наприклад, консалтингова компанія Cambridge Analytics розробила модель, яка дозволяє оцінити за певними параметрами особистість кожного повнолітнього громадянина США, а також здійснювати вплив на поведінку людей.

При використанні технологій Cambridge Analytics у виборчій компанії Д. Трампа (станом на кінець 2017 року – президент США) населення країни було розділено на 32 типи особистостей. Для кожного типу особистості були підготовлені персоналізовані послання і, виходячи з аналізу фахівців [12], ця технологія зіграла

одну з вирішальних ролей у результатах голосування. Аналогічні підходи використані при голосуванні за вихід Великої Британії з Євросоюзу (Brexit) [11] та у передвиборчій компанії Президента Франції.

На основі використання подібних технологій, з урахуванням досвіду застосування програмних засобів аналізу, в РФ у 2014 році була створена компанія Social Data Hub. Social Data Hub пропонує свої рішення рекламним агентствам, аналітикам, банкам, спеціалістам з маркетингу у соціальних мережах, політикам та іншим професіоналам. Інша компанія – Fubutech (того ж засновника) "допомагає чиновникам шукати дисидентів, співпрацює з крупними корпораціями" [14].

Таким чином, технології медіаманіпулювання на ТБ, вплив на поведінку людей через соціальні мережі, використання психологічного профілю людини у результаті психоаналізу за інформацією Big Data для управління поведінкою людини і є одним з видів інформаційно-психологічної зброї.

3. Нейролінгвістичне програмування, психодіагностика та психокорекція людини.

КГБ СРСР, а сьогодні спецслужби та військові РФ разом з вченими та політологами розробляли й розробляють методи зміни поведінки людини шляхом формування заданих програм у замаскованій вербальній (словесній) формі. НЛП – це вплив на думки людини на свідомому та підсвідомому рівнях, націлений на зміну поведінки шляхом формування наперед заданих програм поведінки у "замаскованій" словесній та мовній формах, включаючи техніки гіпнозу та технічні засоби інформаційного впливу.

Таким чином, НЛП може бути потужною інформаційно-психологічною зброєю, в основу якої покладена нейрологія, психофізіологія, лінгвістика, кібернетика і теорія комунікацій.

Психотехнології психозондування (психодіагностика) і психокорекція – принципово новий напрямок у науці, який дозволяє немедикаментозним шляхом діагностувати і корегувати психофізіологічний стан людини шляхом доступу до семантичної пам'яті (підсвідомості). Російський вчений академік І.В. Смірнов розробив способи і методи впливу на психіку (основну функцію головного мозку), які дозволяють управляти психофізіологічним станом людини [15-17]. Психозондування виявляє приховану у підсвідомості інформацію, яка є причиною поведінки та поступків людини. Психокорекція активізує психічні і фізичні можливості організму людини, які вона не використовує у своєму звичайному

стані. Психокорекція – це корегування внутрішньої картини світу людини, ініціація у людини стану і поведінки, адекватних заданій програмі. Психокорекція заснована на нейролінгвістиці – програмування за допомогою спеціально підготовлених мовних конструкцій – сугестивних фабул, тому ці технології мають назви "психолінгвістика" та "психосемантика".

Можливо допустити використання групової психокорекції через ТБ та радіо чи при масових заходах, для чого, як і при методиках Cambridge Analytics та Social Data Hub, формуються відповідні фабули для груп людей з однаковими психологічними профілями.

4. Консцієнтальна війна.

Розвиток методів впливу на свідомість людини та суспільну свідомість призвів до формування ще більш витончених теорій і технологій ІПсВ, які є основою так званої "консцієнтальної війни". Відповідно до роботи [18] "консцієнтальна війна – війна на руйнацію свідомості. В основі такої руйнації лежить знищення здібності людини до вільної ідентифікації, ким людина хоче стати і в межах якої культурно-історичної традиції". Тому світу і громадянам України, у тому числі, нав'язують "руський мір" і традиційні релігійні конфесії. Консцієнтальна війна, на відміну від відвертих військових дій, має ціллю не захоплення чужих територій, а націлена на захоплення свідомості людей, які проживають на цих територіях. Таким чином, головна ціль агресора у консцієнтальній війні – управління свідомістю людей.

Висновки:

Вчені, політтехнологи та військові РФ створили науковий фундамент та технології ведення інформаційно-психологічної війни. Частина цих напрацювань використовується проти України і є реальною загрозою національній безпеці, а інша частина несе потенційні загрози.

У цілому технологія впливу на поведінку являє собою інформаційно-психологічною зброю, яка, на відміну від звичайної зброї використовує:

- технології медіаманіпулювання у засобах масової інформації;
- вплив на людей через соціальні мережі;
- управління поведінкою людини з використанням психологічного профілю людини;
- нейролінгвістичне програмування на основі нейрології, психофізіології, лінгвістики, кібернетики і теорії комунікацій;
- психодіагностику та психокорекцію на основі психолінгвістики, психосемантики та сучасних інформаційних технологій;

- технології роботи із свідомістю людини, що спрямовані на поразку і знищення визначених форм і структур свідомості, а також деяких режимів її функціонування (консцієнтальна зброя).

Застосування інформаційно-психологічної зброї трансформує пам'ять індивіда, створюючи особистість з наперед заданими параметрами, які задовольняють агресора, роблять непрацездатною систему управління державою противника та його збройні сили. Зазначені обставини потребують удосконалення технологій і засобів протидії, що є напрямом подальших досліджень.

Перелік посилань:

1. Операции информационно-психологической войны: краткий энциклопедический словарь-справочник / В. Б. Вепринцев, А. В. Манойло, А. И. Петренко, Д. Б. Фролов; под ред. А. И. Петренко. – М.: Горячая линия – Телеком, 2005. – 495 с.
2. Волкогонов Д. А. Психологическая война: Подрывные действия империализма в области общественного сознания. Москва : Воениздат. 1983. 288 с.
3. Смирнов А. А. Информационно-психологическая война // Свободная мысль. 2013. №6 (1642). С. 81-96.
4. Почепцов Г. Г. Психологические войны. Москва : Рефл-бук; К. : Ваклер. 2002. 528 с.
5. Мухаметов Р. С. Роль публичной дипломатии в формировании международного имиджа России // Современная Россия и мир: альтернативы развития (Россия и Западная Европа: влияние образов стран на двусторонние отношения) : материалы международной научно-практической конференции. Барнаул: Изд-во Алт. ун-та. 2010. С. 111–116.
6. Зорина Е. Г. Технологии подрыва легитимности государственной власти, используемые в интернет-пространстве // Информационные войны. 2016. №1 (37). С. 78-80.
7. Карякин В. В. Стратегии непрямых действий, "мягкой силы" и технологии "управляемого хаоса" как инструменты переформатирования политических пространств // Информационные войны. 2014. № 3 (31). С. 29-38.
8. Аронсон Э., Пратканис Э. Эпоха пропаганды: Механизмы убеждения, повседневное использование и злоупотребление. СПб. : прайм – ЕВРОЗНАК. 2003. 384 с.
9. Глушко А. Н. Основы психометрии. Москва : ГВМУ МО РФ. 1994. 100 с.
10. Kosinski M. Publications. URL : http://www.michalkosinski.com/home/publications (дата звернення: 02.08.2017).

11. The Insider. Расследование Das Magazin: как Big Data и пара ученых обеспечили победу Трампу и Brexit. 06.12.2016. URL : http://theins.ru/politika/38490 (дата звернення: 08.08.2017).

12. Madanapalle A. Big data and psychographic profiling helped Donald Trump with the US presidential election. URL : http://www.firstpost.com/tech/news-analysis/big-data-and-psychographic-profiling-helped-donald-trump-win-the-us-presidential-election-3696847.html (дата звернення: 02.08.2017).

13. Did Michal Kosinski's methods facilitate Donald Trump's victory? Big Data & Business Intelligence. Interview with Michal Kosinski. 18 Jan. 2017. URL: http://www.cebit.de/en/news/article/michal-kosinski-38402.xhtml (дата звернення: 02.08.2017).

14. Писарев А. SocialDataHub: Как экс-рекламщик сделал бизнес на Big Data и продает услуги бывшим коллегам (и чиновникам). 10.02.2017. URL : http://incrussia.ru/fly/socialdatahub-kak-eks-reklamshchik-sdelal-biznes-na-big-data-i-prodaet-uslugi-byvshim-kollegam-i-chi (дата звернення: 15.08.2017).

15. Смирнов И. В. Безносюк Е. В. Методы неосознаваемой психотерапии, психокоррекции и психосемантического анализа с субсенсорным введением информации // Психотерапевт России. Москва : Медицина. 1994. № 3.

16. Безносюк Е.В., Смирнов И.В. Психокоррекция, психопрофилактика, психотерапия // Медицинская психология и психогигиена. 1990. С. 53.

17. Академик Игорь Смирнов: психотронное оружие, психозондирование – репортаж на РЕН-ТВ. 17.08.2013. URL : https://youtube.com/watch?v=3YiLqtnRi68 (дата звернення: 16.07.2017).

18. Громыко Н. В. Использование информационных технологий в качестве консциентального оружия // Альманах "Восток". 2005. № 7-8 (31-32).

РОЗПІЗНАННЯ АНОМАЛЬНИХ СТАНІВ В ІНФОРМАЦІЙНО-ТЕЛЕКОМУНІКАЦІЙНИХ СИСТЕМАХ ПРИ НЕЧІТКОМУ ОПИСІ ПОДІЙ

Зубок В. Ю., Захарченко О.І., Бєланов Ю.О.
Інститут спеціального зв'язку та захисту інформації НТУУ "Київський політехнічний інститут імені Ігоря Сікорського", Україна, м. Київ

Системи моніторингу мають широке використання в інформаційно-телекомунікаційних истемах (ІТС) для контролю працездатності шляхом відстеження критичних характеристик мережі в режимі реального часу або з певною періодичністю, та сигналізування про перехід числових характеристик через певні встановлені рівні. В той час, як існуючі системи виявлення кібератак здебільшого спираються на макроописи (сігнатури тощо), що представляють попередньо ідентифіковані відомі загрози безпеці, аналіз даних, отниманих від систем моніторингу працездатності дає можливість виявлення шкідливого зовнішнього впливу при відсутності чітких характеристик такого впливу та без попереднього опису притаманних такому впливові подій.

Постановка задачі. Протягом останнього часу зростає кількість та складність кібератак на промислові підприємства, органи влади та державні установи. Атаки виконуються шляхом впливу на веб-сайти, інші мережеві служби, телекомунікаційне обладнання, спрямовані на загрозу конфіденційності, цілісності, доступності даних, а також на послаблення спостерережності і керованості системами, де передається, оброблюється, зберігається інформація [1,2,3].

Зазвичай вважається, що виявлення атак є функцією окремих спеціалізованих та коштовних програмно-апаратних комплексів. В той же час у складі будь-якої ІТС присутні системи моніторингу (контролю працездатності). Програмне забезпечення моніторингу численних параметрів мережі а також стану і працездатності серверів використовує гнучкий механізм повідомлень, що дозволяє користувачам налаштовувати оповіщення по пошті практично для будь-якої події. Це дає можливість швидко зреагувати на проблеми з сервером. Таке програмне забезпечення пропонує широкі можливості звітності і візуалізації даних,

базуючись на зібраних даних. Інструменти моніторингу найчастіше чудово піддаються горизонтальному і вертикальному масштабуванню і можуть бути використані для виявлення факту зміни стану системи під впливом кібератаки.

Отже, моніторинг стану підсистем ІТС є рутинною процедурою. Накопичення та аналіз даних від таких підсистем дозволяє вести спостереження за груповими відхиленнями від звичайного стану в компонентів ІТС. Такі відхилення, в свою чергу, можуть бути наслідками прихованого шкідливого зовнішнього впливу чи використання ІТС для реалізації кібератак. Будь-яка атака, в залежності від механізму, залишає слід у вигляді зміни кількісних і якісних характеристик багатьох параметрів. Наприклад, якщо DDoS порівняно легко ідентифікується з боку жертви, то з боку ураженого комп'ютера-бота атаку виявити складно. Проте, наявність в підконтрольній мережі декількох комп'ютерів, що одночасно виконують однакові мережеві операції, виявити набагато легше.

Існуючі системи виявлення кібератак здебільшого спираються на макроописи (сігнатури тощо), що представляють попередньо ідентифіковані відомі загрози безпеці. Системи моніторингу ІТС відстежують критичні характеристики мережі в режимі реального часу або з певною періодичністю, та сигналізують про перехід числових характеристик через певні встановлені рівні. Аналіз цих даних може надати можливість виявлення шкідливого зовнішнього впливу при відсутності чітких характеристик такого впливу та без попереднього опису притаманних такому впливові подій.

Сучасний стан проблеми. Попри різноманіття сучасних кібератак, найбільш небезпечні [4] мають принаймні дві фази: фазу інфікування та активну фазу. У випадку атак класу DDoS інфіковані елементи однієї підмережі (так звані боти) часто використовуються для атак на іншу мережу. У випадку атак класу Ransomware активна фаза атаки спрямована безпосередньо на користувачів власної мережі. Події останніх років довели слабку ефективність корпоративних систем захисту супротив атак класу Ransomware. Однією з причин є те, що виявлення пасивних ботів системами антивірусної перевірки дедалі ускладнює через змінність програмного коду, поведінку ботів, їхню різноманітність [5]. Така сама проблема і з ботами, що приймають участь в атаках DDoS. Але у зв'язку з іншою архітектурою активної фази таких атак, власник інфікованих пристроїв може не зазнати жодних втрат і навіть не

дізнатись про участь в атаці, а отже – і про наявність ботів у власній ІТС.

Отже, необхідні нові методи та інструменти виявлення ботів, які б збільшили вірогідність виявлення першого етапу атаки (інфікування), та якомога швидше ідентифікували ботів в той момент, коли вони вступають в активну фазу атаки. Такі методи можуть полягати в дослідженні групових відхилень від звичайного стану роботи різноманітних підсистем ІТС.

Аномалії можуть досліджуватись:

- в загальному трафіку та в трафіку окремих елементів ІТС (загальний обсяг, кількість сесій, групування по адресі джерела, адресі призначення, протоколах тощо);
- в активних елементах ІТС, насамперед в серверах, а також в окремих програмних процесах (кількість, час виконання, групування по запитах, відповідях тощо).

Для визначення атаки за характером трафіку, система повинна навчатись визнавати нормальну активність системи. Для цього потрібні дві фази: тренування, де будується профіль звичайної поведінки, та тестування, де поточний трафік чи картина подій порівнюється з профілем, створеним на етапі навчання. Аномалії виявляються кількома способами, найчастіше з використанням методів штучного інтелекту. Вже понад 10 років досліджується можливість повністю автоматичного поведінкового аналізу трафіку з метою виявлення атак.

В роботі [6] використовується навчання детектора «типовому» профілю трафіку та виявлення аномалій на осонові відстані Магалонобіса. В роботі [7] пропонуються методи виявлення вторгнень шляхом складання профілю легітимних користувачів.

Важливим напрямком захисту є виявлення аномалій в поведінці користувачів шляхом аналізу протоколів рівня застосувань. В роботі [8] запропоновано розширену напівмарківську модель для опису поведінки веб-користувачів. Щоб зменшити обсяг обчислень, пов’язану з просторовою складністю моделі, запропоновано модифікований алгоритм. У якості критерію вимірювання нормальності користувача використовується ентропія його HTTP-запитів. Веб-сайт, яки є потенційно атакованим, описується за допомогою Марківського простору станів.

Існує багато вдалих та широко вживаних рішень з автоматизації такого моніторингу, зокрема, для серверів на базі UNIX-подібних операційних систем. Збір, накопичення та збереження даних для такого аналізу можуть виконувати системи моніторингу

працездатності, наприклад Nagios, Cacti, Zabbix. Вони також можуть використовуватись для сигналізування про зміну поведінки елементів системи в разі розробки і підключення модуля для аналізу, який можна назвати аналізом поведінки. Наступний набір функцій для системи виявлення аномалій дозволить їй функціонувати та розвиватись:

1) збір даних;
2) навчання: визначення характеристик нормального стану;
3) моніторинг;
4) виявлення відхилень;
5) отримання негативного зворотного зв'язку;
6) коригування порогових значень для характеристик нормального стану.

Негативний зворотний зв'язок необхідний для навчання системи. Із плином часу характеристики нормального стану ІТС змінюються, отже для коригування порогових значень необхідно передбачити можливість автокорекції.

Висновки. Знання принципів і механізмів реалізаіцї кібератак дає можливість сформулювати критерії зміни стану системи під впливом кібератаки. Для виявлення таких змін стану можуть бути використані звичайні системи контролю працездатності (моніторингу) ІТС. Багатовекторний аналіз даних, зібраних цими системами, дозволить виявляти аномалії від подій, чіткого опису яких ще не існує. Це відкриває перпективу підвищення ефективності виявлення інфікованих об'єктів в ІТС.

Список використаних джерел

[1] Internet Security Threat Report 2016. [Електронний ресурс]. Доступно: https://www.symantec.com/security-center/threat-report. Дата звернення: Кві.20, 2017.

[2] M-Trends 2017: Trends from the year's breaches and cyber attacks. [Електронний ресурс]. Доступно: https://www.fireeye.com/current-threats/annual-threat-report.html. Дата звернення: Тра.1, 2017.

[3] The Ukrainian Power Grid Was Hacked Again. [Електронний ресурс]. Доступно: https://motherboard.vice.com/en_us/article/ukrainian-power-station-hacking-december-2016-report. Дата звернення: Кві.20, 2017.

[4] Комплексный подход по защите от направленных атак и вымогательского ПО типа Ransomware. [Електронний ресурс]. Доступно: https://habrahabr.ru/company/infosecurity/blog/331700/. Дата звернення: Лис.18, 2017.
[5] Ботнеты. [Електронний ресурс]. Доступно: https://securelist.ru/botnety/155/. Дата звернення: Лис.16, 2017.
[6] Ke Wang. Anomalous Payload-Based Network Intrusion Detection / Ke Wang, Salvatore J. Stolfo // RAID 2004: Recent Advances in Intrusion Detection. - Pages 203-222.
[7] A.S.Syed Navaz. Entropy based Anomaly Detection System to Prevent DDoS Attacks in Cloud / A.S.Syed Navaz, V.Sangeetha, C.Prabhadevi // International Journal of Computer Applications (0975 – 8887). – Vol.62. – No.15. – Jan.2013.
[8] Yi Xie. A Large-Scale Hidden Semi-Markov Model for Anomaly Detection on User Browsing Behaviors / Yi Xie, Shun-Zheng Yu // IEEE/ACM Transactions On Networking. - Vol. 17. - №1.- Feb. 2009.

DEFINING RELATIVE WEIGHTS OF DATA SOURCES DURING AGGREGATION OF PAIR-WISE COMPARISONS

S.V. Kadenko,
Institute for Information Recording of the National Academy of Sciences of Ukraine, Kyiv, Ukraine

A method for defining the relative data source weights in the process of aggregation of pair-wise comparisons is suggested. It is shown that in order to define the relative weights of data sources, providing data in the form of pair-wise comparison matrices, it is not enough to consider just a-priori weight estimate, or the so-called objective component of the indicator. It is reasonable to assume that data sources, providing more consistent, compete, and detailed information, should be assigned larger weights when data from several sources is aggregated. The suggested approach allows us to consider consistency, compatibility and completeness of pair-wise comparisons as well as the level of detail of pair-wise comparison scales (in addition to the objective component) while calculating the relative data source weights.

Four conceptual levels of data source weight definition are described (from top to bottom): subject domain, specific problem, specific pair-wise comparison matrix, and specific pair-wise comparison. On the subject domain and the problem levels it is suggested to use any of the previously-developed approaches, including pair-wise comparison-based methods. On the level of a specific pair-wise matrix, provided by a data source, it is suggested to assign it a rating, reflecting its inner consistency and compatibility with pair-wise comparison matrices, provided by other sources. Finally, on the level of each pair-wise comparison, it is suggested assign it a weight, depending on the number of grades in the scale, in which it has been provided.

The suggested approach allows us to improve the pair-wise comparison aggregation methods, particularly, combinatorial method (sometimes called enumeration of all spanning trees). Obtained experimental results confirm that all different aspects of data source credibility should be taken into consideration during content-analysis-based research and expert examinations, particularly when completeness, consistency and compatibility levels of estimates vary across data sources, and scales with different numbers of grades are used. On the whole, obtained results allow us to improve the existing pair-wise comparison aggregation methods and increase the credibility of results of

content-analysis-based research and group expert examinations, using pair-wise comparisons.

1 Introduction: why pair-wise comparisons?

Information space is characterized by a multitude of factors, both tangible and intangible. In [1] it is stated that information operations are influenced by many solely qualitative (for instance, socio-psychological) criteria, factors, and parameters. Their formal mathematical and analytical description is a challenging task. Authors also stress the impossibility of development and implementation of some universal methodology for modeling of information operations, first of all, because of weak formalization of related factors and concepts.

In [2] it is shown that information security in general and information operations in particular represent a weakly structured subject domain. Decision-making support technologies prove to be an effective analytical tool in other weakly-structured domains, so they, definitely, should be among the technological means for analysis and modeling of information operations. Beside that, in [1] it is stressed that information security strategy is an important component of the national security strategy. Tsyganok et al. in [3] show that decision support technologies are a powerful instrument of strategic planning, especially, in weakly structured domains, where other approaches are not as effective.

In any weakly-structured domain, that has to be analytically described, we often come to the point when a set of objects or factors has to be measured or compared. Pair-wise comparisons are a flexible way of representation of data on a set of objects that cannot be measured or described by absolute quantitative values. Any measurement, by the way, is a pair-wise comparison, in a sense, because the measured object is actually compared to some benchmark value (meter, pound, second, byte, etc.). However, if there are no benchmarks to compare the objects with, the only way to obtain at least some quantifiable data on these objects is to compare them among themselves.

In the contexts of information security, information policy, and information operations, the necessity to deal with pair-wise comparisons might arise in several situations.

1) Strategic planning (more general case). As information security, informational policy, and information operations represent weakly structured domains, decision-making support technologies prove useful as strategic planning tools in these areas. Strategic planning process using expert data is described in [3] and includes such steps as: formulation of

the main strategic goal by the decision-maker (DM), selection of experts, decomposition (by experts) of the main goal into sub-goals (criteria, factors) down to the level of projects, that can be directly influenced by the DM (building of a hierarchy of criteria), evaluation of relative impacts of criteria by experts, definition of the strategy as the most rational distribution of resources among projects. Under such a scenario, most of the data comes from experts, so they are the primary information source. Pair-wise comparisons come into play when experts evaluate relative impacts of criteria (factors) in the hierarchy.

2) Building some rating or ranking of a set of objects (or factors) based on comparing them among themselves (concepts, topics, product brands, political parties, candidates etc) during some informational-analytical research (more particular case). Just like in the previous case, the data may come from experts, however, it may also come from analysts, that monitor information sources (online publications, meta-search results, blogs, or even comments in social networks). For example, if an online publication says something like "Brand A has much stronger standing at the market than Brand B" or "Candidate X is following close behind Candidate Y, according to the polls", such statements do provide the basis for pair-wise comparisons in some scale. Ordinal pair-wise comparison scales include only two values (such as "better" or "worse", "more" or "less" etc). Cardinal pair-wise comparison scales include particular quantitative values. For instance, a popular scale used by Saaty [4] includes the following values: 1 – no preference, 2 – weak preference, 3 – moderate preference, 4 – moderate plus, 5 – strong preference, 6 – strong plus, 7 – very strong preference, 8 – strong plus, 9 – extreme, as well as reciprocal values. The result of a session of pair-wise comparisons is a pair-wise comparison matrix (PCM), where each element shows ordinal or/and cardinal relation between the two respective objects.

Besides, pair-wise comparisons can be also derived from the frequencies, with which compared objects (again, product brands, political parties, electoral candidates etc) are mentioned in data sources.

If we need to build a unified rating or ranking of a set of objects or factors (that is a common task in both above-mentioned cases) and the data on these objects comes from several data sources, this data needs to be aggregated. However, before aggregation it is necessary to define the relative importance of these data sources, i.e., which of them should be assigned larger weights. Some weights can be assigned to the data sources by experts or analysts a priori, but there are also other important components of relative weights of data sources. It would be reasonable to assume, that the more complete, consistent, and detailed the data is, the

more credible the data source is and the greater weight should be assigned to it. In this paper we are going to address all the listed components of data source weight in greater detail.

2 Literature overview; existing approaches

When it comes to decision-making based on pair-wise comparisons (coming from several data sources), Tom Saaty (author of analytic hierarchy/network process (AHP/ANP) [4]) and his followers can, in a way, be called monopolists. During the last few decades this scientific school suggested a set of approaches to definition of expert competence indicators that can and should be used during aggregation pair-wise comparisons coming from several different sources. In [4] Saaty suggests building a hierarchy of such factors as "skills", "experience", "accomplishments", "persuasion", "efforts" etc and calculate expert competence as an aggregate estimate of the expert according to these criteria (again, based on eigenvectors of PCM, as prescribed by AHP method). If the DM is unable to evaluate experts according to these criteria, then the tasks of self-evaluation and mutual evaluation can be delegated to the experts themselves. At this phase Saaty considers data sources to be equally competent. Ramanathan and Ganesh [5] instead suggest compiling a matrix of alternative weights, obtained as eigenvectors of individual PCM $W = (\vec{w}_1,...,\vec{w}_m)$ (where m is the number of experts) and calculate relative weights of experts based on the equation $\lambda\vec{x} = W\vec{x}$. Another approach is suggested by Yang et. al in [6]. The authors use a kind of combination of two methods, AHP and TOPSIS (described, for instance, by Hwang and Yoon in [7]), and propose to calculate weights of expert judgments based on their distance from the best ones and the worst ones in the set ("ideal alternatives"). These approaches can be relatively easily extrapolated to the cases when data sources are not experts. Online or printed data sources, for instance, can be similarly evaluated by a set of criteria, and assigned respective relative weights.

Speaking of Ukrainian academic schools, we can mention Totsenko [8] who, in a way, summarized the earlier achievements of soviet researchers. Totsenko suggested calculating relative competence of experts as product of self-estimate and the weighted sum of mutual estimate and objective estimate (see formula (1) below). Gnatienko and Snytiuk [9] set forth two approaches for defining the expert's competence: 1) based on the questionnaire with questions of different

types; 2) based on PCMs, provided by the experts themselves. PCM is approximated by a "polyhedron" of weight coefficients. The more its volume is, the less consistent the PCM is, and the lesser weight should be assigned to the respective expert. In the case, when data sources are not experts, we cannot apply the concept of self-estimate (and mutual estimate as well) or talk about completion of questionnaires. In other aspects, the approaches specified in this passage can be extrapolated to the case when experts are not the primary data source. Rather promising attempts to apply decision support methods for processing of data that does not come from experts, but is derived from content-monitoring results have been made by Andriichuk in [10????].

The brief literature overview, provided above, indicates, that every existing approach to calculation of weights of data sources takes into account just one particular aspect of data source weight: objective estimate ([4,9]), mutual or self-estimate ([5]), mutual agreement (compatibility) ([6]), or inner agreement (consistency) ([9]) of data. Consequently, there is a need to develop a method for calculation of data source weights, that would take all the listed aspects into consideration: self-estimate (if applicable), mutual estimate (if applicable), objective a priori estimate of the weights, as well as consistency and compatibility of data. Beside that, it would be reasonable to assume, that the data source weight should depend on the degree of detail, with which the data is presented (the more detailed the data is the greater weight should be assigned to its source). In other words, the data source weight should reflect both a priori estimate and a posteriori estimate (the latter depending on the quality of information, coming from the source). Further we will try to take all the listed aspects into consideration.

3 Problem statement and solution idea

Data source weight calculation problem can be formulated as follows. Let us assume, that at some phase of strategic planning procedure, described in [3], or in the course of some independent content research m data sources are used to compile a rating of n objects. The data from these sources is represented in the form of m PCMs: $\{A^{(k)}; k = 1..m\} = \{a_{ij}^k : k = 1..m; i, j = 1..n\}$. ***We should find*** the relative weights of data sources that would most thoroughly reflect the quality of quality of information coming from these sources (in view of aspects listed in previous sections of this paper). These weights are to be further

used for aggregation of data, coming from all the sources: $c_l : l = [1..m]$: $\sum_{l=1}^{m} c_l = 1$.

Solution idea. Until recently the author of this paper (being the advocate of expert data-based decision support methods) would have preferred to use the approaches suggested by Totsenko [8] to define data source weights. In an expert group the key components of l-th member's competence c_l are (as we mentioned above) self-estimate, mutual estimate, and objective component ($c_{l,self}, c_{l,obj}, c_{l,mutual}$, respectively):

$$c_l = c_{l,self}(x_1 c_{l,obj} + x_2 c_{l,mutual}) \quad (1)$$

In (1) x_1, x_2 are the relative weights of objective and mutual estimates. In the case when data sources are not experts, only the objective component applies.

The results of recent research (such as improvement of combinatorial pair-wise comparison aggregation method [11] and development of a multi-scale approach to data representation [3]) of the academic school the author belongs to call for revision of existing approaches to definition of data source weights based on the requirements, specified in the previous section of this paper.

If in the course of some analytic research data from several sources is used and represented in the form of PCM, then the weights of these sources can be considered at several conceptual or contextual levels (from more general to more specific): 1) level of overall credibility of the data source within the subject domain in general; 2) level of the specific analytic research; 3) level of a specific series of pair-wise comparisons (resulting in completion of one PCM); 4) level of a specific pair-wise comparison, presented in some selected scale. Let us address each of the levels in particular.

1) Data source weight at the level of the *subject domain in general* is reflected by the components, presented in formula (1): self-estimate, mutual estimate, and objective estimate. In the case when the data source is a member of an expert group, all three components apply, if the data source is not an expert, just the objective component remains. It can be set a priori by the analyst (knowledge engineer) who conducts monitoring and research of online or printed data sources.

2) Data source weight in the context of *specific analytic research or examination* can be defined based in keywords, describing the main goal of the research. The principles of calculation of data source weights in terms of basic keywords (BKW) (again, based on the three listed components) were set forth by Totsenko in [8]. Totsenko (similarly to Saaty) suggested building a hierarchy of keywords. The non-normalized weight of the *j*-th data source with respect to *h*-th keyword, located at *i*-th level of the hierarchy k^*_{jih} should be calculated as follows.

$$k^*_{jih} = v_{jih}(x_1 o_{jih} + x_2 w_{jih}) \qquad (2)$$

In (2) x_1, x_2 – are, respectively, the relative weights of objective and mutual estimates of expert group members' competences (if data comes from experts), while $v_{jih}, o_{jih}, w_{jih}$ are, respectively, self-estimate, objective, and mutual estimates of experts' competence according to the given keyword. If experts are not involved and data comes from online or printed sources, then only the objective component remains. If experts are involved then self-estimates, mutual estimates and objective estimates of individual experts' competences, as well as relative weights of components x_1, x_2 can be defined using the methods, listed in the introductory section of this paper.

3) Relative weight of a data source, providing the basis for a *specific PCM* can be defined using the approach described in [12]. According to the approach, the data source should be considered more credible, if the PCM is more consistent within itself and, at the same time, compatible with the respective PCM, based on data from other sources.

At this point we should remind that weights are assigned to data sources in order to ensure more "balanced" procedure of aggregation of PCMs, provided by different sources. Combinatorial method of PCM aggregation ([11, 13]) (sometimes called enumeration of all spanning trees) proves to be one of the most effective data aggregation methods. Within this method, in order to exploit the redundancy of information most thoroughly, we decompose PCM, provided by each given data source, into basic pair-wise comparison sets (that can be represented by connected graphs called spanning trees). Each spanning tree is used to build (reconstruct) an ideally consistent PCM (ICPCM) [11, 12]. Each of these ICPCM is assigned a rating, reflecting the weigh of the data source

at the level of the respective *basic (informatively meaningful) set of pair-wise comparisons* of objects [12].

Weight or "rating" of a certain ICPCM depends on its proximity to original PCM, provided by data sources. It is reasonable to assume, that the more distant from original PCM the ICPCM is, the lower its rating should be (because this "distance" or "difference" is the indicator of inconsistency and incompatibility of the PCM). In [12] it is suggested to calculate ICPCM rating as shown in formulas (3) and (3a).

$$R_{kql} = \frac{c_k c_l}{\ln(\sum_{u,v} \left| a_{uv}^{kq} - a_{uv}^{l} \right| + e)} \tag{3}$$

$$R_{kql} = \frac{c_k c_l}{\log_2(\sum_{u,v} \left| a_{uv}^{kq} - a_{uv}^{l} \right| + 2)} \tag{3a}$$

In formulas (3) and (3a) R_{kql} is the rating (weight) of the ICPCM, "reconstructed" from the q-th informatively-significant set of pair-wise comparisons, taken from original PCM, provided by k-th data source, that is compared to the original PCM, provided by the l-th data source; a_{uv}^{kq} and a_{uv}^{l} are the respective elements of q-th ICPCM from k-th source and the original PCM from the l-th source; c_k, c_l are a priori estimates of data source weights (representing the first two of the above-mentioned conceptual levels). Logarithm operator is used to avoid large differences between PCM ratings and ensure that the rating fall within the same order of magnitude.

In [12] only the case of additive pair-wise comparisons is addressed. That is, the transitivity condition is formulated as $\forall i, j = 1..n : a_{ij} = w_i - w_j = (w_i - w_k) + (w_k - w_j) = a_{ik} + a_{kj}$, where n is the general number of compared objects, w_i, w_j are the relative weights of objects with the respective numbers. For multiplicative pair-wise comparisons (when the transitivity (consistency) condition of

PCM is formulated as $\forall i, j = 1..n : a_{ij} = \frac{w_i}{w_j} = \frac{w_i}{w_k} \times \frac{w_k}{w_j} = a_{ik} \times a_{kj}$) we can suggest alternative formulas for calculation of ICPCM ratings, where sum operation is replaced by product operation:

$$R_{kql} = \frac{c_k c_l}{\ln(\prod_{u,v} \max(\frac{a_{uv}^{kq}}{a_{uv}^{l}}; \frac{a_{uv}^{l}}{a_{uv}^{kq}}) + e - 1)} \quad (4)$$

$$R_{kql} = \frac{c_k c_l}{\log_2(\prod_{u,v} \max(\frac{a_{uv}^{kq}}{a_{uv}^{l}}; \frac{a_{uv}^{l}}{a_{uv}^{kq}}) + 1)} \quad (4a)$$

The denominators in relations (3), (3a), (4), (4a) reflect the differences between the respective PCMs. The greater the difference, the less consistent (and compatible) the data source is. ICPCM, based on data from the given source are compared to original PCM, provided by both this same source and other sources ($k, l = 1..m$). This allows us to consider both consistency and compatibility of the estimates. Based on the look of formulas (3), (3a), (4), (4a), we can make the following conclusion. If the DM (research organizer) needs the results of analytic research to reflect the differences in the data source weights most thoroughly, then (s)he should choose the multiplicative "version" of the formula to calculate the PCM rating, and set smaller logarithm base.

4) At the level of a *specific pair-wise comparison* the weight of the data source should depend on the degree of detail of the scale, in which the comparison is provided. The data source, providing the comparison value in the more detailed scale, should be considered more credible in regard to the respective pair of compared objects. In order to take this component of data source weight into account, the authors of [3] suggest assigning to every pair-wise comparison, provided in some given scale, a coefficient, proportional to the amount of information, "contained" by the scale. The amount of information in the scale of N grades is calculated based on Hartley's formula ([14]): $S_N = I = \log_2 N$.

Consequently, the rating (or weight) of a basic set of pair-wise comparisons of n objects (described above) can be calculated as the

geometric mean of ratings of its elements. Thus, the average non-normalized weight of q-th basic pair-wise comparison set, based on PCM, provided by k-th data source, will amount to:

$$s^{kq} = (\prod_{u=1}^{n-1} \log_2 N_u^{(kq)})^{\frac{1}{n-1}} \quad (5)$$

According to Caley's theorem on trees ([15]), the total number of basic pair-wise comparison sets, derived from a complete PCM of n objects is n^{n-2}. If the matrix is incomplete (comparisons of some objects are not provided by the data source), then the number of such basic sets (spanning trees) is $T \in (0,..,n^{n-2})$. In formula (5) the range of q is $\{q = 1..T\}$. Consequently, when it comes to the whole PCM, provided by data source number k, its rating (or weight) in terms of the degree of detail or scales, the matrix elements are provided in, can be calculated as follows.

$$s^k = (\prod_{\substack{u,v=1 \\ v>u}}^{n} \log_2 N_{uv}^{(k)})^{\frac{2}{n(n-1)}} \quad (6)$$

By definition, the matrix is reciprocally symmetrical, so we can consider only the elements lying above the principal diagonal of the matrix ($v > u$).

All the listed components of relative weights of data sources should be taken into consideration during aggregation of pair-wise comparisons, coming from these sources. So far we have addressed only separate conceptual aspects of the problem solution. In the next section we are going to complete the problem solution process and present a method for data source weight calculation that should be used for aggregation of data from different sources.

4 Application of the approach for improvement of combinatorial method of pair-wise comparison aggregation

Based on considerations, set forth in the previous sections, we can introduce certain improvements into combinatorial method of pair-wise comparison aggregation, described in [11, 12], which will allow us to

make the results of its application more compliant with the actual level of data sources' credibility levels.

Combinatorial method is used in the process of strategic planning ([3]) when experts are estimating the relative impacts of factors that influence the main goal of the strategic plan. Beside that, it can be used to compile a rating of a set of objects based on data from several sources, expressed in the form of complete or incomplete PCMs. We have chosen combinatorial approach from among various pair-wise comparison aggregation methods, because it allows us to use redundant data from different sources most thoroughly, and because of advantages of the approach (such as stability and others), demonstrated in [13].

We should remind that in combinatorial method basic pair-wise comparison sets (so-called spanning trees) are selected from PCMs, provided by each data source. Normalized weights (in Saaty's terms, priorities) of compared objects can be calculated based on each of these sets: $\{w_j^q; j=1..n; q=1..T\}$. If we neglect the differences in the relative credibility of data sources and calculate aggregate values of weights of n objects ($w_j^{aggregate}; j=1..n$) based on PCMs, provided by m sources using the geometric mean formula, we will get the following expression.

$$w_j^{aggregate} = \left(\prod_{q=1}^{T} w_j^q\right)^{\frac{1}{T}}; j=1..n; T \in [1..mn^{n-2}] \quad (7)$$

In order to consider consistency and compatibility of PCMs during aggregation of object weight values, it is reasonable to assign each ICPCM (and the respective weight vector), derived from the individual PCM, provided by a specific data source, a rating (as explained in the previous section). In order to introduce a rating, reflecting both consistency of an individual PCM and compatibility of PCMs, provided by different sources, it is suggested to produce *m* copies of every ICPCM. As a result, we will get $T^*=mT$ ICPCMs ($m \le T^* \le m^2 n^{n-2}$). Beside the level of agreement of estimates, the rating should depend on the weights of scales pair-wise comparisons are provided in. (4-th conceptual level in section 3 of this paper) and a priori estimate of relative weight of the data source (levels 1 and 2). In view of these considerations, we suggest the following formula for calculation of the rating of each single ICPCM (and the respective priority vector) from among all $T^*=mT$ ICPCMs.

$$R_{kql} = \frac{c_k c_l s^{kq} s^l}{\ln(\sum_{u,v} \left| a_{uv}^{kq} - a_{uv}^l \right| + e)} \quad (8)$$

$$R_{kql} = \frac{c_k c_l s^{kq} s^l}{\ln(\prod_{u,v} \max(\frac{a_{uv}^{kq}}{a_{uv}^l}; \frac{a_{uv}^l}{a_{uv}^{kq}}) + e - 1)} \quad (8a)$$

In formulas (8) and (8a) k,l are the numbers of data sources ($k,l = 1..m$) that provide PCMs being compared; c_k, c_l are weight coefficients, representing the levels of subject domain and particular analytic research; again, we should stress that k and l can be equal or different, i.e. ICPCM, based on original PCM, provided by data source number *k*, can be compared with original PCM, provided by this source (inner agreement or consistency), and with PCMs, provided by other sources (mutual agreement or compatibility); q is the number of ICPCM "copy" ($q = 1..mT_k$); s^{kq} is the average relative weight of scales, in which pair-wise comparisons from the given spanning tree were provided (formula (5)); s^l is the average weight of scales, in which the respective PCM by data source number *l* was provided (formula (6)). If we incorporate ICPCM ratings into formula (7), we get the following expression for aggregated object weights.

$$w_j^{aggregate} = \prod_{k,l=1}^{m} (\prod_{q^k=1}^{T_k} (w_j^{(kq_k l)})^{\frac{R_{kq_k l}}{\sum_{u,p,v} R_{upv}}}); j = 1..n \quad (9)$$

5 Experimental research and numeric results

The modified combinatorial method of pair-wise comparison aggregation was tested on multiple numeric examples, using the prototype application in MS Excel. The experiment indicated that aggregation results, that incorporated the weights of data sources, significantly differed from results that did not incorporate these weights, although the same PCMs were used. This fact confirms the necessity to take into

consideration all the aspects of relative importance of data sources when aggregating pair-wise comparisons, because in this case the final result reflects the credibility of these sources more adequately.

Naturally, if all data sources provide data in the same scale, and all comparisons are ideally consistent, then the respective conceptual levels will never come into play (as we can see from the previous sections of this paper). However, in the real world the situation is, usually, quite the opposite.

In this section we are going to consider a numeric example, which illustrates the differences between priority values, derived from pair-wise comparisons, respectively, with and without taking of data source weights into consideration. Let us say, 3 experts (E_1, E_2, E_3) compare 4 alternatives (A_1, A_2, A_3, A_4), and use integer scales with numbers of grades from 2 to 9 for pair-wise comparisons. In order to make the example more illustrative, let us assume that a priori estimates of experts' relative competence levels are equal $(c_1 = c_2 = c_3 = 1,$ if normalized – $c_1 = c_2 = c_3 = 1/3)$. In this example we will use additive PCM rating function (formula (8)).

Ordinal PCMs are shown in Table 1. Let ordinal estimates be perfectly consistent, i.e. alternatives can be sorted in such a way, that all pair-wise comparisons above the principal diagonal of every PCM exceed 1. Let us assume the 1st expert is unable to compare the 1st alternative with the 3rd one. This means that the respective PCM is incomplete, while all spanning trees, including the respective element from the 1st expert's PCM, are informatively insignificant, and, consequently, the respective multipliers in formula (9) equal 1.

Table 1 Ordinal pair-wise comparison matrices, provided by three experts

	E_1				E_2				E_3			
	A_1	A_2	A_3	A_4	A_1	A_2	A_3	A_4	A_1	A_2	A_3	A_4
A_1	1	>	*	>	1	>	>	>	1	>	>	>
A_2		1	>	>		1	>	>		1	>	>
A_3			1	>			1	>			1	>
A_4				1				1				1

Table 2 features the total numbers of grades in the scales, selected by the 3 experts for input of pair-wise comparisons. Table 3 contains the numbers of specific grades in the respective scales. Table 4 contains PCMs of the 3 experts, brought to the most detailed scale with 9 grades

(see section 1 of this paper) according to the rules, proposed in [3]. The unified value (brought to the more detailed scale from the less detailed one) does not always coincide with exact grade value. Pair-wise comparisons, skipped by the expert, are replaced by 1 («no preference»).

Table 2 Number of grades in scales, selected by experts for input of comparisons

	E_1				E_2				E_3			
	A_1	A_2	A_3	A_4	A_1	A_2	A_3	A_4	A_1	A_2	A_3	A_4
A_1	1	9	*	7	1	3	4	5	1	9	9	8
A_2		1	6	5		1	6	7		1	3	9
A_3			1	4			1	8			1	7
A_4				1				1				1

Table 3 Numbers of specific grades in the scales, selected by experts

	E_1				E_2				E_3			
	A_1	A_2	A_3	A_4	A_1	A_2	A_3	A_4	A_1	A_2	A_3	A_4
A_1	1	2	*	7	1	3	4	5	1	2	4	8
A_2		1	3	4		1	2	4		1	2	4
A_3			1	2			1	2			1	3
A_4				1				1				1

Table 4 Pair-wise comparison values, brought to the most detailed scale

	E_1				E_2				E_3			
	A_1	A_2	A_3	A_4	A_1	A_2	A_3	A_4	A_1	A_2	A_3	A_4
A_1	1	2	1	8 5/6	1	7 1/2	8 1/6	8 1/2	1	2	4	9
A_2	1/2	1	3 8/9	6 1/2	1/7	1	2 2/7	4 5/6	1/2	1	3 1/2	4
A_3	1	1/4	1	2 5/6	1/8	3/7	1	2	1/4	2/7	1	3 1/2
A_4	1/8	1/6	1/3	1	1/8	1/5	1/2	1	1/9	1/4	2/7	1

Based on matrices from table 4, 48 ICPCM are built ($mn^{n-2} = 3 \times 4^2 = 48$). In order to compare each of these ICPCMs with original PCMs, 3 copies of eeach ICPCM are made (according to the number of experts). That is, the general number of ICPCMs being processed (and, respectively, the number of multipliers in formula (9)) is $T = m^2 n^{n-2} = 144$. Ratings of 144 matrices are calculated according to formula (8) (if the spanning tree lacks the pair-wise comparison, skipped

by the 1st expert, then the rating of the respective matrix equals 0). After that, the ratings are normalized by their sum (see power index in formula (9)). At the same time, alternative weight vectors (priorities) are calculated based on every ICPCM. Any basic set of pair-wise comparisons (for instance, the last row of an ICPCM) can be used as non-normalized weight vector. Finally, non-normalized aggregate priorities are calculated according to formula (9) and normalized.

Normalized priority values, calculated according to formula (9) and according to formula (7) (i.e. not taking expert competence into consideration) are shown in table 5.

Table 5 Priority vectors, incorporating and not incorporating expert competence values

	W_1	W_2	W_3	W_4
Competence taken into consideration (formula (9))	0.533	0.264	0.133	0.071
Competence not taken into consideration (formula (7))	0.547	0.269	0.129	0.054
Difference (%)	2.6	1.9	3.0	23.9

As we can see, in spite of perfect ordinal and fairly high cardinal agreement of original data, and in spite of usage of additive (and not multiplicative) expression for ICPCM rating calculation, the results obtained using simple and weighted average mean are different. The greatest difference is witnessed between the smallest priority values.

The example confirms that if the level of agreement, completeness, and detail of initial data from a given source is low, then the final aggregation results will be significantly influenced by the source's relative weight. We should also stress, that if data sources are experts, then feedback can be organized to improve completeness, consistency, and compatibility of the data. However, if data sources are not experts (humans), then feedback cannot be organized, and data source weight components, reflecting the listed aspects, will play far greater role than in the cases when experts are involved.

6 Conclusions

It has been shown that during aggregation of data from different sources it is necessary to take their relative weights into account. Moreover, in order to calculate the relative weight of a data source it is

not enough to just estimate it a priori. The weight of a data source should incorporate completeness, agreement, and degree of detail of information, provided by the source. A method for calculating the relative weights of data sources has been suggested. Beside a priori estimates, the data source coefficient incorporates the degree of detail of scales, in which the data is provided, as well as consistency and compatibility of the data. The method has been tested on a large number of numeric examples, where initial data was input in the form of individual PCMs and aggregated. Obtained experimental results illustrate significant differences between aggregate priority values, respectively, incorporating and not incorporating data source weights. The method's key advantage is its universal nature: it allows us to take all the above-mentioned aspects of data source credibility into account, not just some single aspect. The original data source weight calculation mechanism provides the basis for improvement of pair-wise comparison aggregation methods and for increasing the credibility of results of their application.

This paper is prepared as part of project #F73/23558 "Development of Decision-making Support Methods and Means for Detection of Information Operations". The project won the contest #F73 for grant support of scientific research projects held by The State Fund for Fundamental Research of Ukraine and Belarusian Republican Foundation for Fundamental Research.

References

1. Горбулін В.П., Додонов О.Г., Ланде Д.В. Інформаційні операції та безпека суспільства: загрози, протидія, моделювання: монографія – К., Інтертехнологія, 2009 – 164 с.
2. Kadenko S.V. Prospects and Potential of Expert Decision-making Support Techniques Implementation in Information Security Area / S.V.Kadenko // CEUR Workshop Proceedings – 2016 – pp. 8-14.
3. Tsyganok V.V. Using Different Pair-wise Comparison Scales for Developing Industrial Strategies / V.V.Tsyganok, S.V.Kadenko, O.V.Andriichuk // Int. J. Management and Decision Making. – 2015. – Vol. 14, No 3. – pp. 224–250.
4. Saaty, T.L., 1994. Fundamentals of Decision Making and Priority Theory with The Analytic Hierarchy Process. / T.L.Saaty; RWS Publications, Pittsburgh PA, 204220.
5. Ramanathan, R. Group Preference Aggregation Methods Employed in AHP: An Evaluation and Intrinsic Process for Deriving Members' Weightages / R.Ramanathan, L.S.Ganesh// European Journal of Operational Research 79 – 1994, pp. 249-265.

6. Yang Q, Du P-a, Wang Y, Liang B. A rough set approach for determining weights of decision makers in group decision making. PLoS ONE 12(2) – 2017: e0172679. doi:10.1371
7. Hwang, C. L. Multiple attribute decision making: methods and applications : a state-of-the-art survey / C.L. Hwang, K. Yoon ; Berlin ; New York : Springer-Verlag, 1981 – 259 p.
8. Totsenko, V. G. Determination of Relative Competence of Group Members in the Subject under Discussion on Group Decision Making / V.G.Totsenko // JAutomatInfScien.v34.i4.30 – 2002. DOI: 10.1615
9. Гнатієнко Г.М. Експертні технології прийняття рішень / Г.М.Гнатієнко, В.Є.Снитюк. – К.: ТОВ "Маклаут", 2008 – 444 с.
10. Андрійчук О.В. Побудова баз знань систем підтримки прийняття рішень при виявленні інформаційних операцій / О.В.Андрійчук, Д.В.Ланде // Реєстрація, зберігання і обробка даних. Щорічна підсумкова наукова конференція – Київ, 2017 – с. 101-103.
11. Циганок В.В. Комбінаторний алгоритм парних порівнянь зі зворотним зв'язком з експертом / В.В.Циганок // Реєстрація, зберігання і обробка даних. – 2000. – Т.2, №2. – с.92-102.
12. Циганок В. В. Метод обчислення ваг альтернатив на основі результатів парних порівнянь, проведених групою експертів / В.В.Циганок // Реєстрація, зберігання і обробка даних. – 2008. – Т.10, №2. – с. 121-127.
13. Tsyganok V. Investigation of the aggregation effectiveness of expert estimates obtained by the pairwise comparison method / V. Tsyganok // Mathematical and Computer Modeling, 52(3-4) – 2010 – pp. 538–544.
14. Hartley R.V.L. Transmission of information. / R.V.L.Hartley // Bell System Technical J. 1928 – 7, pp.535-63.
15. Cayley A.A Theorem on Trees. / A.Cayley // Quarterly Journal of Mathematics 1889 – Volume 23, pp. 376-378.

ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ АНАЛІЗУ КЛІЄНТСЬКОЇ БАЗИ АБОНЕНТІВ ТА ПРОГНОЗУВАННЯ ЇХ ПОВЕДІНКИ

Кузнєцова Н.В.
Інститут прикладного системного аналізу Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського», м. Київ

Вступ

Сучасний світ неможливо уявити без мобільних пристроїв, Інтернету, планшетів та комп'ютерів. Сьогодні клієнт, який не користується послугами мобільного зв'язку чи Інтернету, стає майже виключенням. Це скоріш за все клієнти, які змінюють оператора зв'язку або перебувають в Україні досить короткий час і потребують тимчасового тарифного пакету. Телекомунікаційні компанії зосереджують свої зусилля на розробці інформаційних технологій, що використовують сучасні методології інтелектуального аналізу даних та моделюють поведінку клієнтів-користувачів послуг телекомунікаційних компаній. Основною метою є виявлення уподобань клієнтів та утримання їх як абонентів, розробляючи та пропонуючи їм нові послуги та тарифні пакети згідно їх потребам.

Постановка задачі

Стандартні підходи та методи дозволяють побудувати математичні моделі, які будуть прогнозувати безпосередньо подію – можливий відток клієнтів. Метою даного дослідження стало розроблення математичних моделей аналізу клієнтської бази та короткострокове та довгострокове прогнозування поведінки клієнтів за рахунок виявлення часового проміжку та групи абонентів з усієї бази клієнтів, які замислюються найближчим часом (від 1 місяця до 3 місяців) у зміні оператора.

Загальні припущення теорії аналізу виживання

Для аналізу даних використовується вибірка (популяція), яка характеризується певними ознаками: по кожному об'єкту відомий результат події (загибель чи виживання). Для цього здійснюється один з видів цензурування (відсікання). Спостереження називаються цензурованими, якщо спостережувана залежна змінна представляє момент настання термінальної події, а тривалість дослідження обмежена за часом. Можливі механізми цензурування змінних:

фіксоване цензурування (спостереження відбувається протягом фіксованого проміжку часу) та випадкове цензурування (спостереження відбувається протягом проміжку часу, який настає після того часу, коли елементи вибірки пережили певну подію) [1-3].

При розробці математичних моделей враховувались коваріанти, тобто параметри, що характеризують поведінку клієнтів, як статичні параметри – характеристики клієнта, так і динамічні параметри його поведінки (обсяг трафіку, кількість хвилин дзвінків тощо).

Функція виживання визначається як $S(t) = P(T > t)$, а функція ризику $h(t) = \lim_{\Delta \to 0} \frac{P(t < T < t + \Delta \mid T > t)}{\Delta}$, $h(t) = -\frac{\frac{dS(t)}{dt}}{S(t)}$.

Найпростіша функція, яка визначає, що ризик є константою в часі: $h(t) = \lambda$, або що еквівалентно $\log h(t) = \mu$.

Оскільки $S(t) = \exp[-\int_0^t h(u)du]$, то після підстановки та інтегрування отримуємо: $S(t) = e^{-\lambda t}$, а $f(t) = \lambda e^{-\lambda t}$. Це функція щільності ймовірності з добре відомим експоненційним розподілом з параметром λ. Таким чином, сталий ризик передбачає експоненційний розподіл для часу, поки не наступить подія (або час між подіями) [4-5].

Модель виживання може будуватись з горизонтом часу і ймовірність відтоку клієнта в наступний період може бути обчислена таким чином [2]:

$$PO(t \mid x) = P(t \le T < \mathrm{t} + \mathrm{b} \mid \mathrm{T} \ge t, X = x) =$$

$$\frac{P(T < t + \mathrm{b} \mid X = x) \; - \; P(T \le t \mid \mathrm{X} = \mathrm{x})}{P(T \ge t \mid X = x)} = \quad , \qquad (1)$$

$$= \frac{F(t + b \mid x) - F(t \mid x)}{1 - F(t \mid x)} = 1 - \frac{S(t + b \mid x)}{S(t \mid x)}$$

де t – час спостереження обслуговування клієнту, а x - значення коваріаційного вектору X для цього клієнта, тобто параметри самого клієнта, його тарифного плану та його поведінки.

Для розподілу часу життя можна прийняти узагальнену лінійну модель [7]:

$$P(T \le t \mid X = x) = F_\theta(t \mid x) = g(\theta_0 + \theta_1 t + \theta^{\mathrm{T}} x),$$

де $\theta = (\theta_2, \theta_3, \ldots, \theta_{p+1})^{\mathrm{T}}$ p-вимірний вектор, g - відома функція зв'язку, така як логістична чи пробіт-функція. Таким чином, ця модель характеризує умовний розподіл часу обслуговування абоненту телекомунікаційною компанією Т в термінах невідомих параметрів. Як тільки ці параметри будуть оцінені, отримаємо оцінку функції умовного розподілу, $F_{\hat{\theta}}$ і, нарешті, оцінка PO (probability of outlet) може бути обчислена шляхом включення цієї оцінки у рівняння (1), тобто

$$PD^{GLM}(t \mid x) = \frac{F_{\hat{\theta}}(t+b \mid x) - F_{\hat{\theta}}(t \mid x)}{1 - F_{\hat{\theta}}(t \mid x)} = 1 - \frac{S_{\hat{\theta}}(t+b \mid x)}{S_{\hat{\theta}}(t \mid x)},$$

де $\hat{\theta} = \hat{\theta}^{GML}$ є оцінкою максимальної правдоподібності вектору параметрів.

Розглянемо одновимірний випадок. У такому випадку $\theta = \theta_2$ і умовний розподіл задається моделлю $F(t \mid x) = g(\theta_0 + \theta_1 t + \theta_2 x),$ зі щільністю $f(t \mid x) = \theta_1 g'(\theta_0 + \theta_1 t + \theta_2 x).$ Оскільки зазвичай задана випадкова цензурована справа вибірка, то умовна функція правдоподібності представляє собою добуток членів, що включають умовну щільність, для нецензурованих даних та умовної функції виживання для цензурованих даних:

$$L(Y, X, \theta) = \prod_{i=1}^{n} f(Y_i \mid X_i)^{\delta_i} (1 - F(Y_i \mid X_i))^{1-\delta_i},$$

де Y_i – строк обслуговування і-го клієнту в телекомунікаційній компанії і δ^i є індикатором відтоку для і-го клієнту.

Таким чином, логарифмічна функція правдоподібності визначається [2]:

$$l(\theta) = \ln(L(Y, X, \theta)) = \sum_{i=1}^{n} [\delta_i \ln(f(Y_i \mid X_i)) + (1 - \delta_i) \ln(1 - F(Y_i \mid X_i))] =$$

$$= \sum_{i=1}^{n} [\delta_i \ln(\theta_i g'(\theta_0 + \theta_1 Y_i + \theta_2 X_i)) + (1 - \delta_i) \ln(1 - g(\theta_0 + \theta_1 Y_i + \theta_2 X_i))] =$$

$$= \sum_{i=1}^{n} \delta_i [\ln(\theta_1) + \ln(g'(\theta_0 + \theta_1 Y_i + \theta_2 X_i))] + \sum_{i=1}^{n} (1 - \delta_i) \ln(1 - g(\theta_0 + \theta_1 Y_i + \theta_2 X_i))$$

І, нарешті, оцінка знаходиться як максимізація функції логарифмічної правдоподібності:

$$\hat{\theta}^{GML} = \arg\max_{\theta} l(\theta).$$

Моделі пропорційних ризиків Кокса

Відома модель Кокса, запропонована в 1972 році, інтенсивно використовується в самих різних областях, особливо в медицині і страхування, для оцінки умовного ризику захворювання при заданих значеннях вихідних ознак [1-3]. Модель Кокса заснована на припущенні, що функцію ризику можна факторизувати, тобто представити у вигляді добутку двох функцій:

$$h_i(t) = h_0(t) \cdot \psi(X_{i1},..., X_{ik}),$$

де $h_0(t)$ – базова функція інтенсивності, що включає фактор часу, але не включає коваріанти, а $\psi(X_{i1},...X_{ik})$ – лінійна функція досліджуваних ознак, яка не включає фактор часу.

Досить часто модель записують у наступному вигляді :

$$h_i(t) = h_o(t) \cdot e^{\{\beta_1 X_{i1} + ... + \beta_k X_{ik}\}},$$

$$\ln h_i(t) = \ln h_o(t) + \beta_1 X_{i1} + ... + \beta_k X_{ik},$$

де $\beta_1,...\beta_k$ – невідомі параметри.

Аналіз поведінки клієнтів за допомогою SAS-технологій

Для поставленої задачі моделювання використовували статистичні дані клієнтської бази телекомунікаційної компанії за 2014-2016 роки [6]. Групи абонентів розділялись на корпоративних та індивідуальних клієнтів (враховувалась також стать клієнтів). Моделювались окремо клієнти кожної групи для прогнозування можливого відтоку клієнті та періоду, в який це може відбутись.

Були використані можливості інформаційних технологій SAS Enterprise Miner для побудови і оцінювання параментрів та якості узагальненої лінійної моделі, написані власні коди на SAS Base для побудови моделей виживання з відповідними розподілами (моделі пропорційних ризиків Кокса та напівпараметричної моделі). Були отримані наступні коефіцієнти узагальненої лінійної моделі:

Source	DF	Type I SS	Mean Square	F Value	Pr > F
ONNET_MINS	1	2668.403703	2668.403703	18684.3	<.0001
OMO_MINS	1	250.054005	250.054005	1750.89	<.0001
PSTN_MINS	1	20.909361	20.909361	146.41	<.0001
INTERN_MINS	1	59.120719	59.120719	413.97	<.0001
GPRS_USG_MB	1	460.615988	460.615988	3225.26	<.0001
SUBSCRIPTION_TYPE_CO	2	25.330932	12.665466	88.68	<.0001
time	1	6285.412857	6285.412857	44010.8	<.0001

Після цього модель була оцінена на основі ROC-кривої та індексу GINI (рис.1).

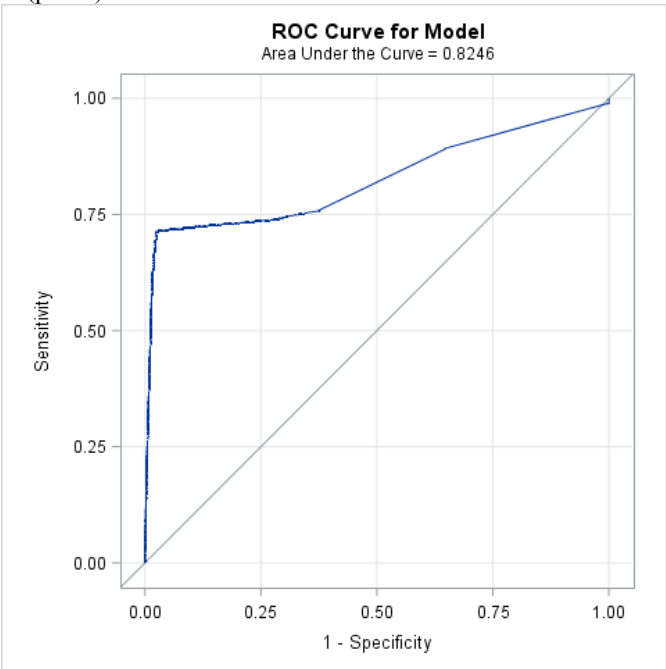


Рисунок 1 – ROC-крива для узагальненої лінійної моделі

За площею під ROC-кривою (AUC=0,8246) обраховуємо значення індексу GINI: $GINI = 2 * AUC - 1 = 0,6492$. Це говорить про прийнятні предикативні якості моделі, тобто дозволяє спрогнозувати саму подію відтоку клієнту (чи буде відток, чи ні). Однак ця модель не може бути використана для прогнозування самого періоду можливого відтоку. Для цього були побудовані моделі виживання та фукція втрат (збитковості) для моделі Кокса (рис. 2 та рис 3.) [6,7].

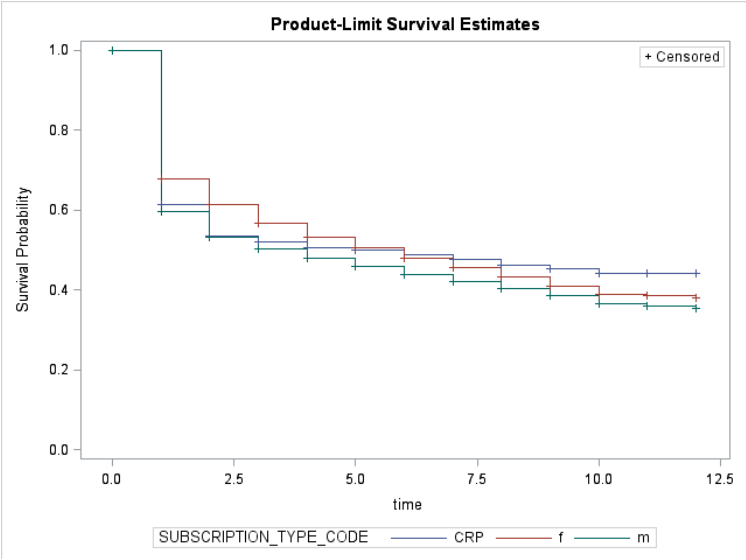


Рисунок 2 – Графіки функцій виживання для згрупованих даних

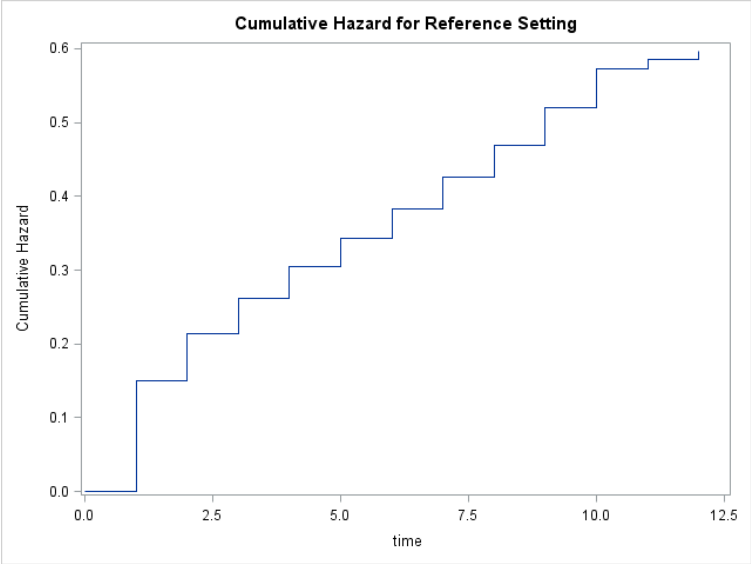


Рисунок 3 – Графік функції втрат для моделі Кокса

За цими моделями ми можемо визначити прийнятний ступінь ризику та час, в який момент ризик переходить в катастрофічний, а також рівень втрат, які в цей момент буде нести телекомунікаційна компанія з точки зору недоотримання доходу через відток клієнтів (абоненттів).

Висновки

Проведене моделювання підтвердило доцільність використання методів з теорії виживання, оскільки враховуються навіть спостереження з невідомим результатом, тобто ті, по яких не встановлений факт відтоку і вони досі обслуговуються оператором, що значно розширює і наближає вибірку до реальних статистичних даних. Окрім цього, побудовані моделі дозволяють включати навіть прогнози описаних факторів, динаміку поведінки, що дозволяє будувати динамічні моделі, які є більш точними та функціональними. І, нарешті, побудовані моделі дозволяють здійснювати прогнозування фактору ризику та можливих втрат з урахуванням часу, тобто на певний період вперед.

Література

1. Cox D. R. Regression models and life-tables / D. R. Cox, S. Society, S. B. Methodological // 2007. — Vol. 34, No. 2. — P. 187–220.
2. Cao R., Vilar J.M., Devia A. Modelling consumer credit risk via survival analysis / SORT 33 (1) January-June 2009, p.3-30.
3. Marimo M. Survival analysis of bank loans and credit risk prognosis master of science mathematical statistics / M. Marimo // [Електронний ресурс]. — Режим доступу : http://wiredspace.wits.ac.za/jspui/bitstream/10539/18597/1/Mercy%20Marimo%20Thesis_Survival%20Analysis_28.03.%202015_v1.pdf.
4. Stepanova M. Survival analysis methods for personal loan data / M. Stepanova, L. C. Thomas // Operations Research. — 2002. — Vol. 50, No. 2. — P. 277–289.
5. Fleming, T.R., Harrington, D. P. Counting Processes and Survival Analysis. - John Wiley & Sons - New York. 1991.
6. Кузнєцова Н.В. Моделювання фінансового ризику в телекомунікаційній сфері / Н.В. Кузнєцова, П.І. Бідюк // Наукові вісті НТУУ “КПІ”. – 2017. – №5. – С. 51–58.
7. Бидюк П. И. Анализ временных рядов / П.И. Бидюк, В. Д. Романенко, О. Л. Тимощук. – Киев: Политехника, 2013. – 600 с.

METHOD OF MACHINE LEARNING BASED ON STOCHASTIC AUTOMATA IN PROBLEMS OF DATA CONSOLIDATION FROM OPEN SOURCES

Kuzminykh V.A., Koval A.V., Osipenko M.V.
NTUU Igor Sikorsky Kyiv Polytekhnik Institute
Kyiv, Ukraine

Data consolidation is considered as the complex of methods and procedures directed to the extraction of data from diverse sources, transformation to a uniform format, in which they can be loaded in the storage facility of given or analytical system. Usually, consolidation described a process of search, selection, analysis, structuring, transformation, storage, cataloging and granting to the consumer of the information on given topics is understood. The task of the information consolidation is one of the urgent tasks of processing of big volumes of data [1].

In the base of consolidation procedure process of the collection and the organization of data storage as, convenient from the standpoint of their processing on particular platform lies. An important feature of data acquisition on the basis of unsealed sources is instability of information fullness of these sources, the absence of reliable a priori information about their content and his relevance, low accuracy, and efficiency of expert evaluations of their condition and conformity of these sources to the topics and inquiries parameters [2].

Therefore, for data processing from unsealed sources of the information it is the necessary execution of effective consolidation of data with use of specialized software supplying:

1. adaptive choice of the sources of the information the most relevant for request conformity;
2. accumulation of the information of the condition of the sources during the performance of the request;
3. the account of the opportunity for a change of condition of the sources at repeat request;
4. analysis of perspective from the standpoint of sources relevance;
5. construction of information evaluation of sources promising from the standpoint of relevance.

At the development of the model and algorithm of consolidation of the information from diverse sources with plenty of documents, it is set the task of choice of maximum quantity of documents relevant to the request

at a minimum quantity of processed (checked to relevance to the request) documents. It has been made possible through the expense of revealing the sources of the information the most appropriate to the request and their further use at further choice.

There is a range of information resources presenting information retrieval objects:

– mass media are various sort news sites (RSS channels) and semantic sites (or electronic mass media versions).

– electronic libraries are distributed information system enabling to preserve reliably and effectively to use electronic documents through global networks.

– databases are the set of files organized by special image, documents grouped on topics and with spreadsheets which are united to groups.

– sites are Internet-resources devoted some organization, company, enterprise, particular problem or person. They differ by the completeness of the information, from pure fact-finding, surface to highly professional, which lights all parties of activity.

– information portals are the groups of sites to which it is possible to take advantage of various service services. They can contain a various scientific, political, economic and other information, as well as electronic letter boxes, catalogs, dictionaries, directories, weather forecast, TV-program, exchange etc. As a rule, their updating of which takes place in a real of the time.

There are numerous models of information retrieval, on the basis of which modern information-analytical and information retrieval systems are built.

Among them it is possible to allocate following the most common types of the models of information retrieval [3]:

1. Boolean model,
2. The model of indistinct plenty,
3. Vector model,
4. Latent-semantic model,
5. Probability model.

Boolean the model uses the apparatus of mathematical logic and the theory of plenty [4]. Advantages of the model are the simplicity of the model and her realization. Model deficiencies - the complexity of construction of inquiries without knowledge by Boolean algebra, is impossible to rank automatically documents and to scale search.

The traditional model of indistinct plenty is based on the theory of indistinct plenty and admits a partial element accessory to plenty [5]. In

such model, all document file is described as the set of indistinct plenty of terms. The advantage of the model - opportunity to rank results, the simplicity of realization, is unnecessary big memory size. Model deficiencies are big computational expenditures, then in the Boolean model, and smaller accuracy.

In vector model [6] documents and users' inquiries are presented as n-dimensional vectors in -dimensional vector space. At the same time space dimension corresponds total of various terms in all documents. Advantages of the model are the simplicity of construction and opportunity of results ranking. Model deficiencies - it is required big volumes of data processing it is necessarily accurate of coincidence with request parameters.

Latent-semantic model [7] is constructed on the basis of latent-semantic analysis (Latent Semantic Analysis - LSA). This method of extraction and representation of the values of words dependent on context with the aid of statistical processing of big volumes of textual documents. The advantage of the model - less space dimension, than vector model, is not provided with an accurate coincidence of request parameters, is not required difficult set-up. To the model deficiencies it is possible to attribute - plenty of calculations and absence of the rules of choice of dimension, from which results efficiency depends.

The specific place is taken by probability models. These models of the search are based on the application of methods of the theory of probability and use statistics which define the probability of the document relevance to search request [8]. In the base of these models, the principle of probability ranking on the decrease of the probability of their relevance to the user's request lies. These models it differs considerably from each other by computational procedures. Advantages of the model are the opportunity of adaptation of the model of use, the absence of dependence of volume of calculations from a number of inquiries parameters, high efficiency at work with the sources of the information which are constantly updated. Model deficiencies are the necessity of constant training of the system during work of the model and regarding low efficiency on initial stages of work.

To date there are no models which would have obvious quantitative and qualitative advantages of some above others [3,9], however the largest interest is presented by the decision constructed on the basis of stochastic approaches which allow successfully to realize the principles of machine training, successfully to be adapted to varied conditions, is simple in construction and realization.

To meet the challenges of search and consolidation of the information from unsealed sources the most promising is the use of methods of machine training based on operational analysis of the condition of the sources of the information during consolidation problem solving.

Machine training is considered usually as a separate section of the theory of artificial intellect studying methods of construction of algorithms, capable to be adapted and to learn. Machine training is on the joint of mathematical statistics, methods of optimization and classical mathematical disciplines, but has as well its own specific features connected with the problems of computational efficiency, adaptation, and conversion training. Almost none research in machine training does not dispense with an experiment on the model or real data verifying practical serviceability of method which is developed [3].

Use of stochastic model allows effectively to realize a machine training in two ways [10]:

1. Training with reinforcement (reinforcement learning). At the same time, the role of objects is played by the "situation and accepted decision". Results are the value of functional quality characterizing correctness of accepted decisions (reaction of the environment), that is, averaged evaluation of relevance. Here essential role is played by the time factor. The examples of applied tasks: choice of documents from bibliographic and abstract bases, information retrieval in open information resources, for example, scientific RSS-channels, self-learning of robotized systems, etc.

2. Dynamic training (online learning). Specific features that precedents arrive with the flow. It is required to make a decision immediately on each precedent and simultaneously re-training the model of dependence in view of new precedents. Here essential role is played by the time factor that assumes opportunity of constant adaptation of the model of choice to changes of the environment during work of the system.

The structure of interaction during consolidation contains following basic elements – the sources of the information, parameters of the content of the request, documents packs.

The sources of the information – it information resources, important and are considered in the construction of particular information system, is considered in the determined case of the search of the information.

Parameters of the content of the request in most instances are difficult terms simple both, parameters determining various signs of the time, places (occurrences, edition, storage etc.), and many other specifications determining features of determined request.

Documents packs are particular sets of information units (articles, records, tables, reports, newspapers, journals, books, the theses of reports, news, reviews and other electronic data).

We will consider basic specifications of the description of consolidation of the information pursuant to the stochastic model of choice of the sources of the information on the basis of the theory of stochastic automatic devices [11,12]. For the evaluation of the relevance of the sources of the information to the inquiries, model is used on the basis of the theory of indistinct plenty [5], which takes into account communications of parameters of the request and the sources of the information.

Such model admits partial conformity between request and the source of the information. This conformity is evaluated by the value within the range of [0,1]. It is formed on the basis of the definition of coincidence between request parameters and specifications of information units entering the particular source of the information.

After execution of sample and the evaluations of a few units of the information from determined source information average pursuant to the amount of chosen documents units.

The evaluation of conformity *i-th* source of the information *j-th* parameter of the request defines his partial conformity, she corresponds range [0,1] and can be determined as

$$k_{ij} = \frac{1}{V}\sum_{l=1}^{V} q_{ijl} \qquad (1)$$

where q_{ijl} - quantitative evaluation of conformity q_{ijl} *l-th document (for* l *= 1,...,* V*)* with *i-th* source of the information *j-th* parameter of the request at one sample for the evaluation relevance of the source ;

$q_{ijl} = 0$ -when the *j-th* parameter isn't present in *l-th* document *i-th* source of the information;

$q_{ijl} = 1$ - when *j-th* parameter is present in *l-th* document *i-th* source of the information.

V - the volume of one sample from one source of documents should correspond following limitations:

$$V \ll S_i \text{ для } i = 1,...,n \qquad (2)$$

where S*i* - a number of information units (documents) in each of *n-th* the sources of the information, are considered for given request.

Then the evaluation of relevance *i-th* source of the information will be defined as

$$R_i = \frac{1}{m}\sum_{j=1}^{m} k_{ij} v_j \qquad (3)$$

where $0 < v_i < 1$

v_j - weight coefficient *j-th* parameter of the topic of the request, is defined on expert evaluations or on the basis of priorities, can define independently the customer of the query.

At such consideration, the evaluation of relevance *i-th* source to the determined request will be defined in the interval [0,1].

For the organization of procedures of an effective choice of information sources of relevant pursuant to inquiries algorithms constructed on the basis of the theory of stochastic automatic devices can be used. Such by the stochastic automatic device is stochastic automaton Moore type [12]. It is automatic device, for which it satisfies following conditions for conventional density of probability

$$p(u', y / u, x) = p(u'/ u, x)\, p(y / u) \qquad (4)$$

Considering type automatic device Moore, as automatic device with discrete time (discrete stochastic automatic device), where the moments of transition are defined as the number of iterations of process of information retrieval at ultimate the amount of iterations, can be recorded that

$$p(u(t+1), y(t) / u(t), x(t)) = p(u(t+1) / u(t), x(t))\, p(y(t) / u(t)) \quad (5)$$

For greater presentation stochastic type automatic device Moore can be described in canonical kind

$$u(t+1) = F(u(t), x(t+1)) \qquad (6)$$

$$y(t) = f(u(t)) \qquad (7)$$

where t is the variable which defines time, that is the moments of operation of the automatic device. This time is defined as integer parameter, that is, $t = 1,.., N$, where N - given a number of iterations of information retrieval, on each of which is executed choice V documents from one of the sources of the information elected on this iteration (D1,..., Dn).

From the standpoint of the system of information retrieval constructed on above-described rules, these values can be defined so:

$u(t)$ - the condition of the system on current iteration which defines probabilities of choice of the sources of the information (D1,..., Dn) on this iteration.

$u(t+1)$ - the condition of the system on the following iteration which defines probabilities of choice of the sources of the information (D1,..., Dn) on the following iteration.

$x(t)$ - the input data of the system on current iteration determining results of the evaluation of a sample of size V from elected on current iteration the sources of the information,

$y(t)$ - the output data of the system on current iteration determining the selected source of the information (D1,..., Dn).

Realization of such automatic device can be constructed in such a manner that change of condition of *u(t + 1)* automatic device is defined as regular dependence, and his exit *y(t)* is defined in the kind of stochastic process.

The algorithm of the information consolidation constructed on the basis of the use of such automatic device consists of following steps:

1. Initial state of stochastic *u*(*t*) automatic device for *t =1* as probabilities vector is installed (with equal probabilities or on the grounds of the previous sessions of information retrieval)

$$P(t) = \{p_1(t), p_2(t), p_3(t),..., p_i(t),..., p_{n-1}(t), p_n(t)\} \quad (8)$$

$$\sum_{i=1}^{n} p_i(t) = 1 \quad (9)$$

2. Evenly distributed random variable *w* on the interval [0,1] is generated.

3. Depending on value random variable *w* is a defined interval, the appropriate to one of *n* the sources of the information.

At the same time, if

$$\sum_{i=0}^{z-1} p_i(t) < w < \sum_{i=1}^{z} p(t) \quad (10)$$

that realization of the automatic device will correspond the source of the information with number z.

4. Sample and processing V the units of documents from the source of the information on number z are executed.

5. The evaluation R_i (3) is led of relevance for V documents from the source of the information on number z.

6. On given algorithm recalculation of the values of probabilities is produced. Development of such algorithm is a separate investigation phase. At present, a few algorithms with one, two and three level adaptations are developed. For simplification of account as an algorithm, recalculation table is presented.

Table 1. Table of recalculation of the vector

R_i	<0.2	<0.4	<0.6	<0.8	<1.0
k_R	0.5	0.75	1	1.5	2

then

$$p_z(t+1) = p_z(t)k_R \qquad (11)$$

A calculation for the vector *p(t + 1)* in such a manner that

$$D(t) = 1 - p_z(t) \qquad (12)$$

$$D(t+1) = 1 - p_z(t+1) \qquad (13)$$

$$p_i(t+1) = \frac{p_i D(t+1)}{D(t)} \qquad \text{для } i = 1,...,n \text{ при } i \neq z \qquad (14)$$

7. We go to the step 2 for a further sample of documents from the sources of the information or we finish the process of choice in the case of performance of the conditions of his ending.

In the case of repeat information retrieval on the same request, it is possible to use already present probability model of choice of the most relevant sources of the information that considerably reduces the time of the search. At the same time and occurrence of other sources more appropriate to the request is not excluded opportunity of a change of the situation in due course. Selection of parameters, methods, and optimization of their specifications with the use of real-life sources of the information practically is impossible, as access to real sources of the information containing necessary documents, for example, scientific articles, is limited to a considerable extent both on a number of inquiries, and at their cost. Therefore, was developed test program environment on Python (Anaconda), which allows conducting research of models for the definition of parameters of adaptive stochastic models and comparisons of their efficiency with other models.

During the performance of the program, virtual sources of data are created in which preset percent of valid models of documents are located. These values and distinction for all sources are specified before the beginning of the generation. Valid model of the document is formed from beforehand present request copy. Those parameters are used which can be specified by the user as a part of search request, for example, the keywords, the name of the author, the date of publication. Not valid

documents are formed by the way of exception from the initial documents of part of parameters, the appropriate to the request.

Use of test environment enables to obtain the appropriate data for definition of the most effective method of choice of the sources of the information, definition, and set-up of his parameters for maximization of the amount of choice of relevant documents at the limited amount of inquiries to the sources of the information. Test environment can be used and for the set-up of parameters of any other annexes requiring analysis of plenty of inquiries to diverse sources having probabilities of the specification.

Table 2. Results of comparison of the models on the test environment.

A number of processed documents	Boolean model	Vector model	Stochastic model
200	7	10	14
400	15	16	33
600	19	22	51
800	26	31	72
1000	38	43	98

The example of testing of the algorithm in a test environment is presented consisting of ten sources of the information generated pursuant to request each of which contains one thousand generated documents. In each of the sources, they were located from 1% to 10% of relevant documents, enables the evaluation of finding of the determined amount of relevant documents with the use of various search algorithms. The example of results of testing they are listed in the table 2. The described stochastic algorithm at processing of plenty of documents (up to 1000) gave the twice best ratings on amount chosen relevant documents than algorithms of direct search constructed on the basis of Boolean and vector model.

The presented approach in the development of algorithms with use of the stochastic automatic device for data consolidation allows creating the complex of software for the decision of tasks of consolidation of data for various systems. One of the promising directions of use of the algorithm of consolidation based on the use of stochastic automatic devices is being used for their opportunities for construction of the systems of the search of scientific and technical information from various scientific bibliographic and abstract bases and other unsealed sources. One of the modifications of the described algorithm was tested during the

development of a pilot project of a geocoding system on request and showed positive results [13].

References

[1] Cherniak L. Big Data - a new Theory and Practice—Open systems.RDBMS—M.: Open systems, 2011— № 10—ISSN 1028-7493 (in Russian)

[2] Shakhovs'ka N.B. Methods of processing of consolidated data using data space — Modelling problems—2011— № 4—p.72-84 (in Ukrainian).

[3] Christopher D. Manning, Prabhakar Raghavan, Hinrich Schutze. Introduction to Information Search (translate from English) — M.: JSC "I.D.Williams", 2011—p.504 (in Russian).

[4] Yaglom I.M. Boolean structure and its models —M: Soviet radio, 1980 —p.192 (in Russian).

[5] Ukhobotov V.I. Selected chapters of the theory of fuzzy sets —Tutorial—Chelyabinsk: Publishing house of Chelyabinsk State University, 2011—p.245 (in Russian).

[6] Dubinsky A.G. Some questions of application of vector model of representation of documents in information search — Control systems and machines — 2001 — №4 — p.77–83 (in Russian).

[7] Bondarchuk D.V. The use of latent-semantic analysis in the case of classification of texts by emotional coloring— Bulletin of research results—2012—2(3)—p.146—151 (in Russian).

[8] Robertson S. E. The probabilistic ranking principle in IR— Journal of Documentation — 1977 — № 33 — p.294–304.

[9] Lande D. V., Snarskiy A. A., Bezsudnov I. V. Internetika. Navigation in complex networks: models and algorithms—Moscow: Bookhouse "Librocom"— 2009—p.264 (in Russian).

[10] Witten I.H., Frank E. Data Mining: Practical Machine Learning Tools and Techniques (Second Edition). — Morgan Kaufmann, 2005.

[11] O.V. Koval, V.O. Kuzminykh, D.V. Khaustov. Using stochastic automation for data consolidation - Research Bulletin of NTUU "KPI".Engineering. – 2017. – №2. – C. 29-36.

[12] Rastrigin L.A., Ripa K.K. Automated random search theory—Riga: Zinatne,1973.— p.344 (in Russian).

[13] Kuzminykh V.O., Boichenko O.S. The system of automatic geocoding of user requests— Environmental security clusters: energy, environment, information technology—Kyiv:"MP Lesia", 2015— NTUU "KPI" – pp. 217–222 (in Ukrainian).

АНАЛІЗ ВІДКРИТИХ НАБОРІВ ДАНИХ НА ПРИКЛАДІ ОЦІНЮВАННЯ СВІТОВОЇ ЕКОЛОГІЧНОЇ ПРОБЛЕМИ ІЗ КРИТИЧНИМ РІВНЕМ CO_2 В АТМОСФЕРІ

Кузьмичов А.І.
Інститут проблем реєстрації інформації НАН України, Київ, Україна

Світові ЗМІ підкреслюють серйозну і небезпечну екологічну проблему – різке збільшення рівню діоксиду вуглецю (carbon dioxide, CO_2) в атмосфері, з-за чого вже виникли непередбачені наслідки. За даними Всесвітньої метеорологічної організації рівень забруднення атмосфери викидами CO_2 перевищив важливий поріг і, можливо, не опуститься нижче ще багато поколінь. Мова йде про критичний поріг у $400:10^6$ молекул в атмосфері, який людство перетнуло 2015 року. За наявною інформацією на душу населення припадає по кілька тон CO_2, загальні обсяги якого вимірюють в мільйонах тон. Є ще парниковий газ, який прямо впливає на зміни клімату, це сукупність 7 газів, обсяги якого вимірюється в тисячах тон, є й інші, усе це – «продукт» сучасної життєдіяльності людства. Й у будь-кого виникають природні запитання: скільки цих викидів в Україні усього й скільки їх припадає на кожного з нас, порівняння оцінок стану цієї ситуації у нас й на світовому рівні, які прогнози нас чекають на найближчі роки.

Вступ

Для отримання обґрунтованої відповіді на поставлені запитання, тобто, «знань», за схемою «дані – інформація - знання», у першу чергу звертаємось до власних (державних) джерел, а для уточнення і узгодження – до відкритих наборів даних (open datasets), що зберігаються у певних інформаційних сховищах поза державних установ. Отримавши необхідні дані, користуємось інструментальними засобами сучасної дата-аналітики, розгорнутими на доступних платформах, в прикладі, в середовищі Excel.

Зовнішні джерела даних Зовнішніми називають джерела відкритих даних, що зберігаються поза робочого комп'ютера, не зважаючи на географічне розташування джерела.

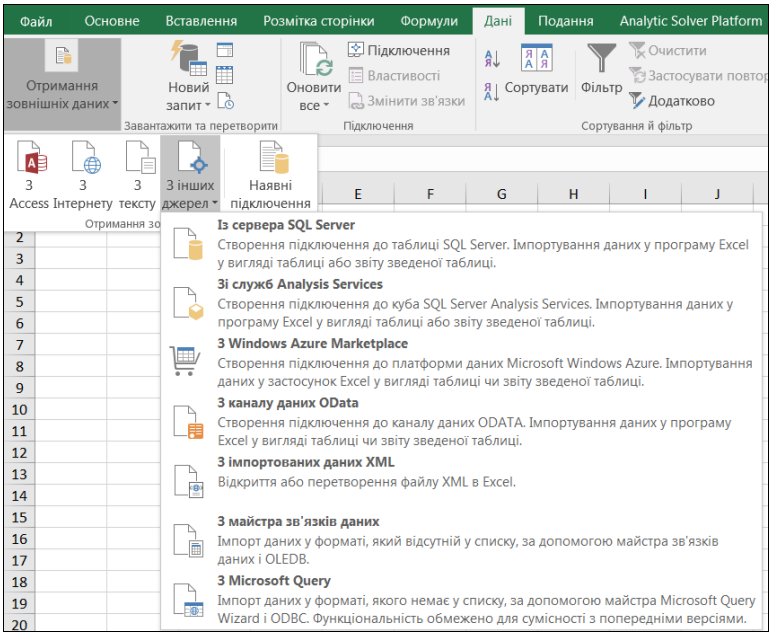
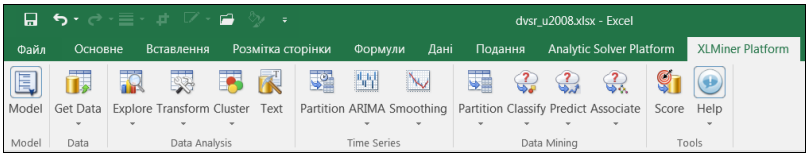


Рис. 1. Зовнішні джерела даних

Дані Зареєстровані дані у відкритих наборах даних, особливо великих обсягів, часто вважаються «сирими» з-за різних способів їх реєстрації, представлення і тривалого зберігання, часових періодів їх виникнення, передачі і організації зберігання, застосованих технічних засобів комунікації, типів і форматів, і, зрозуміло, комп'ютерної грамотності і кваліфікації виконавців.

Інформація Тож для отримання інформації шляхом обробки даних здійснюють комплекс попередніх процедур розуміння отриманих даних, їх дослідження і підготовки до наступної обробки. Це, зокрема, процедури візуалізації, фільтрації, сортування, отримання узагальнених статистичних оцінок та взаємозв’язків між елементами записів баз даних, це пошук відсутніх чи зайвих значень, будь-яких помилок, зміна типу чи формату тощо. Враховуючи солідні розміри цих наборів як баз даних, розміри яких вимірюються тисячами записів і десятками полів, ця робота – складова майнінгу даних – має здійснюватися виключно із застосуванням спеціального програмного інструментарію (наприклад, надбудови XLMiner середовища Excel):



Знання Лише маючи очищені дані, можна розраховувати на отримання якісної інформації й очікуваних «знань» засобами поглибленого статистичного аналізу – кореляційного, регресійного, дисперсійного, дискримінантного тощо, оптимізаційного і імітаційного моделювання й, за необхідності, інструментами машинного навчання зі сфери штучного інтелекту.

Рішення Кінцевий результат – сутність («знання», інсайт) процесу – приховані закономірності чи зв'язки між складовими необроблених масивів даних – що надана обробкою отриманої інформації, це, скажімо, математична модель типу рівняння регресії чи описова статистика, ці знання надалі використовуються у прийнятті організаційних та управлінських рішень.

Дейтамайнінг Апарат майнінгу даних є складовою сучасної науки Data science і аналітичної практики Data analytics, яку на професійному рівні здійснюють фахівці за відповідними новітніми спеціальностями і яку, бажано, мають вміти робити кожен із дослідників. Саме так нині отримують відповіді хоча б на прості запитання, типу наведених вище, так формують зважені рекомендації для формування й прийняття відповідальних рішень на місцевому чи глобальному рівнях.

Для отримання відповідей на поставлені запитання у першу чергу звертаємось до державного органу – Держстату України (*ukrstat.gov.ua*), де зберігаються усі статистичні дані, які віддзеркалюють функціонування держави в усіх його проявах.

Тут ланцюгом *Статистична інформація → Навколишнє середовище* знаходимо потрібний офіційний документ (Рис. 2), звідки дізнаємось, що викиди CO_2 реєструються лише з 2004 року, за 2014-2016 рр. вони неповні з-за відсутності даних із тимчасово окупованих територій.

Тобто, маємо **10** років реєстрації, останній офіційний показник – Україна викинула в атмосферу **230,7** млн. т у 2013 році. Важко сказати, багато це чи мало, як було і що буде із цими викидами, хто їх продукує, до чого треба готуватися, який вплив і на що мають ці тонни і, нарешті, чому ми дізнаємось від ООН чи ВВС, а не від державних служб, бо вони є.

Викиди забруднюючих речовин та діоксиду вуглецю в атмосферне повітря

	Обсяги викидів забруднюючих речовин			Крім того, викиди діоксиду вуглецю		
	усього,тис.т	у тому числі		усього,млн.т	у тому числі	
		стаціонарними джерелами	пересувними джерелами[1]		стаціонарними джерелами	пересувними джерелами[1]
1990	15549,4	9439,1	6110,3	...	...	...
1991	14315,4	8774,6	5540,8	...	...	...
1992	12269,7	8632,9	3636,8	...	...	...
1993	10015,0	7308,3	2706,7	...	...	...
1994	8347,4	6201,4	2146,0	...	...	...
1995	7483,5	5687,0	1796,5	...	...	...
1996	6342,3	4763,8	1578,5	...	...	...
1997	5966,2	4533,2	1433,0	...	...	...
1998	6040,8	4156,3	1884,5	...	...	...
1999	5853,4	4106,4	1747,0	...	...	...
2000	5908,6	3959,4	1949,2	...	...	...
2001	6049,5	4054,8	1994,7	...	...	...
2002	6101,9	4075,0	2026,9	...	...	...
2003	6191,3	4087,8	2103,5	...	...	...
2004	6325,9	4151,9	2174,0	126,9	126,9	...
2005	6615,6	4464,1	2151,5	152,0	152,0	...
2006	7027,6	4822,2	2205,4	178,8	178,8	...
2007	7380,0	4813,3	2566,7	218,1	184,0	34,1
2008	7210,3	4524,9	2685,4	209,4	174,2	35,2
2009	6442,9	3928,1	2514,8	185,2	152,8	32,4
2010	6678,0	4131,6	2546,4	198,2	165,0	33,2
2011	6877,3	4374,6	2502,7	236,0	202,2	33,8
2012	6821,1	4335,3	2485,8	232,0	198,2	33,8
2013	6719,8	4295,1	2424,7	230,7	197,6	33,1
2014[2]	5346,2	3350,0	1996,2	194,7	166,9	27,8
2015[2]	4521,3	2857,4	1663,9	162,0	138,9	23,1
2016[2]	3078,1	3078,1	...	150,6	150,6	...

Рис. 2. Офіційний документ Держстату України

Інше джерело – Мінприроди України, тобто, Міністерство екології та природних ресурсів (*menr.gov.ua*) – є уповноваженим органом, що забезпечує інвентаризацію антропогенних викидів джерелами та абсорбції поглиначами парникових газів на національному рівні, щороку розробляє і подає у відповідні організації ООН Національний кадастр викидів, починаючи із базового 1990 року й, зрозуміло, в Держстат України (який запровадив лише у 2004 році облік викидів 130 забруднюючих речовин, у тому числу, CO_2).

Для отримання повної картини звертаємось до джерел міжнародних організацій.

Перше з них – EDGAR (Emissions Database for Global Atmospheric Research) при European Commission (*edgar.jrc.ec.europa.eu*), де отримано csv-файл

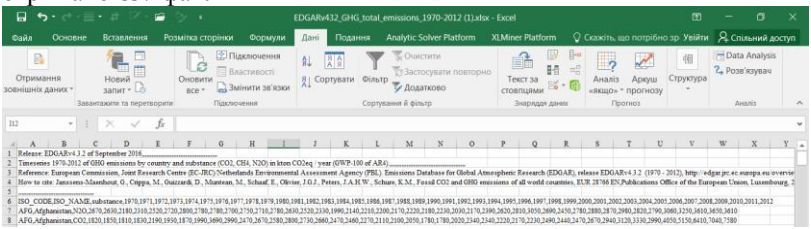


Рис. 3. Відкриті дані від EDGAR

який після певних перетворень в Excel став табличною базою даних із 25200 записів, де містяться дані про викиди 4-ох видів для 208

держав, зареєстровані за 55 років (1960-2015), дані для України надані за період 1970-2014 рр. (аналогічно, для усіх республік СРСР, 1970-1990 рр.)

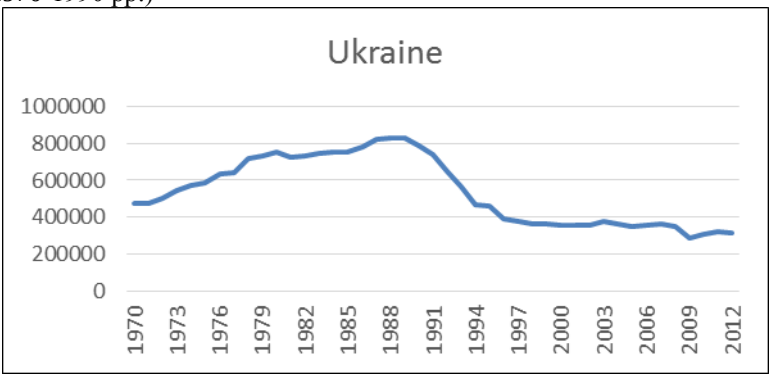


Рис. 3. Перший результат для України

Друге джерело: OECD (Organisation for Economic Co-operation and Development), потрібні дані – база даних на 25000 записів у csv-форматі щодо 6 видів викидів для 138 держав світу за період 1960-2015 рр. на сайті *data.oecd.org*.

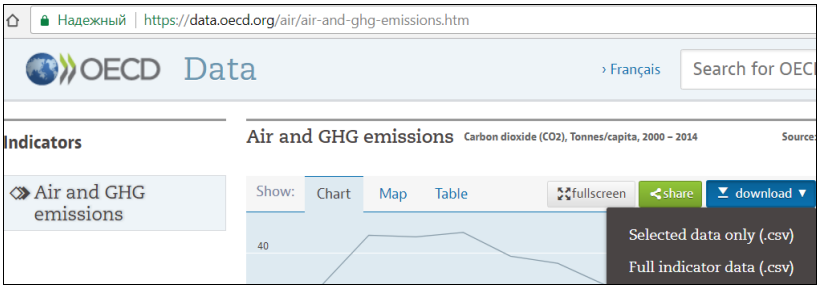


Рис. 4. Дані в csv-форматі від OECD

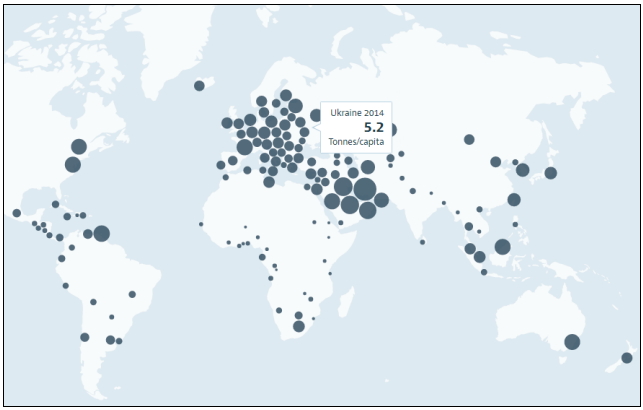


Рис. 5. OECD: Українці, дихайте глибше! 5,2 тони CO_2 кожному українцю

Третє джерело: UN (OOH), csv-файл бази даних на 2500 записів стосовно викидів CO_2 для 215 держав світу за період 1990-2011 рр. Усі отримані дані щодо України зведені у таблицю, за якою побудовано графік:

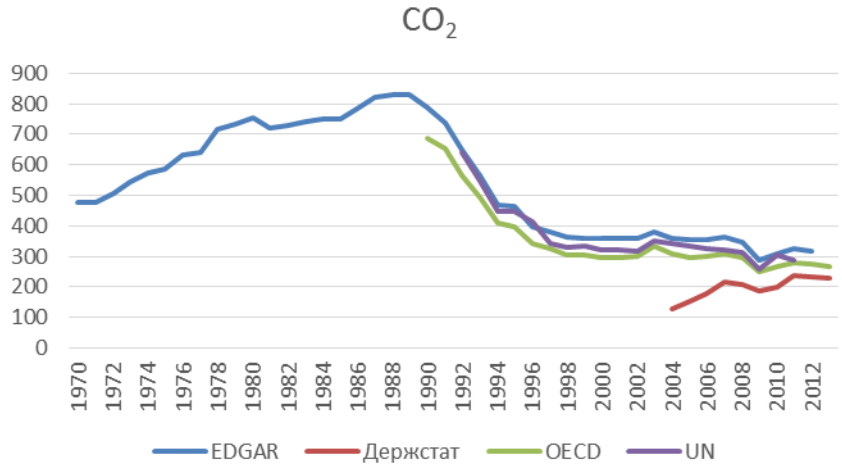


Рис. 6. Результати порівняльного аналізу

1	CHN	CO2	2014	9086.96
2	USA	CO2	2014	5176.21
3	OEU	CO2	2014	3391.65
4	IND	CO2	2014	2019.67
5	RUS	CO2	2014	1467.55
6	JPN	CO2	2014	1188.63
7	DEU	CO2	2014	723.27
8	KOR	CO2	2014	567.81
9	IRN	CO2	2014	556.09
10	CAN	CO2	2014	554.8
11	SAU	CO2	2014	506.59
12	BRA	CO2	2014	476.02
13	ZAF	CO2	2014	437.37
14	IDN	CO2	2014	436.53
15	MEX	CO2	2014	430.92
16	GBR	CO2	2014	407.84
17	AUS	CO2	2014	373.78
18	ITA	CO2	2014	319.71
19	TUR	CO2	2014	307.11
20	FRA	CO2	2014	285.68
21	POL	CO2	2014	279.04
22	TWN	CO2	2014	249.66
23	THA	CO2	2014	243.52
24	**UKR**	CO2	2014	236.54
25	ESP	CO2	2014	231.99
26	KAZ	CO2	2014	223.69
27	MYS	CO2	2014	220.52
28	ARG	CO2	2014	192.41
29	ARE	CO2	2014	175.43
30	EGY	CO2	2014	173.27

EDGAR

1	China	2011	9019.5
2	United States	2011	5305.6
3	India	2011	2074.3
4	Russian Federation	2011	1808.1
5	Japan	2011	1187.7
6	Germany	2011	729.5
7	Korea, Republic of	2011	589.4
8	Iran (Islamic Republi	2011	586.6
9	Indonesia	2011	564.0
10	Saudi Arabia	2011	520.3
11	Canada	2011	485.5
12	South Africa	2011	477.2
13	Mexico	2011	466.5
14	United Kingdom	2011	448.2
15	Brazil	2011	439.4
16	Italy	2011	398.0
17	Australia	2011	369.0
18	France	2011	338.8
19	Turkey	2011	320.8
20	Poland	2011	317.3
21	Thailand	2011	303.4
22	Ukraine	2011	286.2
23	Spain	2011	270.7
24	Kazakhstan	2011	261.8
25	Malaysia	2011	225.7
26	Egypt	2011	220.8
27	Argentina	2011	190.0
28	Venezuela	2011	188.8
29	United Arab Emirate	2011	178.5
30	Viet Nam	2011	173.2

OECD

№	ISO_NAME	2012
1	China	9920
2	United States of	5200
3	India	2100
4	Russian Federati	1780
5	Japan	1300
6	Germany	806
7	Korea (the Repu	604
8	Iran (Islamic Rep	593
9	Canada	564
10	Mexico	484
11	Brazil	473
12	United Kingdom	470
13	Saudi Arabia	462
14	Indonesia	446
15	Australia	429
16	Italy and San Ma	400
17	South Africa	386
18	France and Mon	356
19	Turkey	340
20	**Ukraine**	316
21	Poland	309
22	Spain and Ando	279
23	Taiwan (Provinc	273
24	Thailand	263
25	Kazakhstan	256
26	Egypt	229
27	Malaysia	217
28	Argentina	197
29	Venezuela (Boliv	191
30	United Arab Emi	189

UN

Рис. 7. Рейтинг України

Висновки

Дані Держстату не завжди узгоджуються зі світовою тенденцією до зменшення викидів CO_2. У часовому періоді 2004-2014 рр. наші показники майже втричі нижчі за показники світових реєстраторів стану екології без будь-яких пояснень.

Навіть поверхневий порівняльний аналіз відкритих даних щодо екологічної проблеми світового масштабу й його тривалої дії свідчить, що державні організації мали б першими інформувати про небезпеку суспільство і владу, маючи кваліфіковано розроблену інформацію і доносити її для підготовки серйозних управлінських рішень, направлених на реальне зниження шкідливих викидів, зокрема, парникових газів та CO_2 у їх складі, загалом їх 130 видів, які викидаються у повітря, а поглинаються рослинами, водами, ґрунтами й, зрозуміло, живими організмами. Основні «гравці»: хімічна промисловість, теплоенергетика, металургія, чия продукція йде на експорт. Отримані результати свідчать, що в сучасних умовах якісного інформаційного забезпечення суспільства кінцеве рішення

має узгоджуватися із якісними показниками світових екологічних організацій.

Література

1. UKRAINE'S GREENHOUSE GAS INVENTORY 1990-2015 (draft). Annual National Inventory Report for Submission under the United Nations Framework Convention on Climate Change and the Kyoto Protocol. – The Ministry of Environment and Natural Resources of Ukraine, 2017. – 518 p.
2. Shmueli G. et al. Data Mining for Business Analytics. Concepts, Techniques, and applications, 3-ed. – Wiley, 2017. – 549 p.
3. Witten I. et al. Data Mining. Practical Machine Learning Tools, 4-ed. – Elsevier, 2017. – 655 p.
4. Analytic Solver Platform. XLMiner Platform. Data Mining User Guide. – Frontline Systems, Inc., 2016. – 188 p.
5. Abbott D. Predictive Analytics. Principles and Techniques for Professional Data Analyst. – Wiley, 2014. – 433 p.
6. Kenny P. Business Decision from Data. Statistical Analysis for Professional Success. – Apress, 2014. – 260 p.
7. Кузьмичов А.І. Економетрія. Моделювання засобами MS Excel. – К.: Ліра-К, 2015. – 212 с.

ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ ПРОГНОЗУВАННЯ ТА МОДЕЛЮВАННЯ КІБЕРБЕЗПЕКИ

Кунченко-Харченко В.[1], Огірко І.[2], Огірко О.[3]
[1]Черкаський державний технологічний університет, м.Черкаси, Україна,
[2]Українська академія друкарства, м. Львів, Україна,
[3]Львівський державний університет внутрішніх справ, м. Львів, Україна

Сучасні інформаційні технології потребують організації високого рівня захисту даних [1-4]. Об'єктами комп'ютерної безпеки є інформаційні ресурси, канали інформаційного обміну і телекомунікації, механізми забезпечення функціонування телекомунікаційних систем і мереж та інші елементи інформаційної інфраструктури. Об'єктом захисту в інформаційній системі є інформація з обмеженим доступом, яка циркулює та зберігається у вигляді даних, команд, повідомлень, що мають певну обмеженість і цінність як для її власника, так і для потенційного порушника технічного захисту інформації. Стандартність архітектурних принципів побудови, обладнання та програмного забезпечення персональних комп'ютерів, висока мобільність програмного забезпечення і ряд інших ознак визначають порівняно легкий доступ професіонала до інформації, що знаходиться в ПК. Для захисту персональних комп'ютерів використовуються програмні методи, які значно розширюють можливості по забезпеченню безпеки інформації, що зберігається. Серед стандартних захисних засобів персонального комп'ютера найбільше поширення отримали: засоби захисту [3-7] обчислювальних ресурсів, що використовують парольну ідентифікацію і обмежують доступ несанкціонованого користувача; застосування різних методів шифрування, що не залежать від контексту інформації; засоби захисту від копіювання комерційних програмних продуктів; захист від комп'ютерних вірусів і створення архівів. Компютерна безпека – це сукупність проблем у галузі зі телекомунікацій та інформатики, пов'язаних з оцінкою і контролюванням ризиків, що виникають при користування комп'ютерними мережами. Основними технічними складовими комп'ютерної безпеки[6-9] є:

Конфіденційність – означає, що у неавторизованих користувачів не буде доступу до інформації. Наслідки, які можуть бути викликані

прогалинами в конфіденційності, можуть варіюватися від незначних до руйнівних;

Цілісність – означає, що інформація захищена від неавторизованих змін, що не відноситься до авторизованих користувачам. Загрозу цілісності баз даних і ресурсів, як правило, представляє хакерство;

Аутентифікація – сервіс контролю доступу, який здійснює перевірку реєстраційної інформації користувача. Іншими словами це означає, що користувач – це є насправді той, за кого він себе видає;

Доступність – означає те, що ресурси доступні тільки авторизованим користувачам. Іншими важливими компонентами, яким приділяється велика увага професіоналами в області комп'ютерної безпеки, є контроль над доступом і суворе виконання зобов'язань.

Для користувачів Інтернету найбільш важливою складовою є конфіденційність, тому що більшість користувачів думають, що їм нема чого приховувати або інформація, яку вони надають при реєстрації на сайті, не є секретною. В Інтернеті інформація швидко поширюється і зібрана інформація з різних джерел може багато сказати про людину. Тому можливість контролю інформації, для чого вона збирається, хто і як може нею скористатися – є дуже серйозним і важливим питанням в контексті комп'ютерної безпеки. Захист інформації [10-14]– сукупність організаційно-технічних заходів і правових норм для запобігання заподіянню шкоди інтересам власника інформації чи автоматизованій системі та осіб, які користуються інформацією. Сьогодні захист інформації все більше і більше пов'язується з безпечним функціонуванням комп'ютерних систем у всіх галузях суспільної діяльності. Актуальна проблема захисту інформації від різних загроз: — несанкціонований доступ – 2%; — укорінення вірусів – 3%; — технічні відмови апаратури мережі – 20%; — цілеспрямовані дії персоналу – 20%; — помилки персоналу – 55%. Однією з потенційних загроз для інформації в інформаційних системах слід вважати цілеспрямовані або випадкові дії персоналу, оскільки вони становлять 75% усіх випадків. Для вирішення проблеми рекомендується проводити постійне ознайомлення користувачів з наявною політикою безпеки. Інформаційна безпека є складовим компонентом загальної проблеми інформаційного забезпечення людини, держави і суспільства. Робота присвячена засобам, стратегіям і принципам забезпечення безпеки в інформаційних системах, а також правовим і технічним аспектам діяльності в цій

галузі. Кібербезпека — це процес застосування заходів безпеки з метою забезпечення конфіденційності, цілісності та доступності даних. Кібербезпека забезпечує захист ресурсів. Кібербезпека покликана захистити дані на етапі їх обміну та збереження. Створення безпечних комп'ютерних систем і додатків є метою діяльності програмістів, а також предметом теоретичного дослідження як у галузі інформатики. Забезпечення безпеки зводиться до управління ризиком: визначення потенційних загроз, оцінка ймовірності їхнього настання та оцінка потенційної шкоди, із наступним ужиттям запобіжних заходів в обсязі, що враховує технічні можливості й обставини. Поділ міжмережевих екранів залежно від з'єднань, які відслідковують:проста фільтрація — без відслідкування поточних з'єднань, а фільтрація потоку даних виключно на основі статичних правил;фільтрація з урахуванням контексту — з відслідковуванням поточних з'єднань та пропуском лише таких пакетів, що задовольняють логіці й алгоритмам роботи відповідних протоколів та програм. Це дозволяє ефективніше боротися з різноманітними DDoS-атаками та вразливістю деяких протоколів мереж.

Вимоги для прогнозування та моделювання кібербезпеки:володіння методами та засобами програмування мовами високого та низького рівня; - володіння спеціалізованими програмними пакетами; - здатність планувати й реалізувати відповідні заходи, щодо захисту інформації в інформаційних і комунікаційних системах; - знання правових основ дослідницьких робіт і законодавства України в галузі інформаційної безпеки; - здатність до ділових комунікацій у професійній сфері, знання основ ділового спілкування, навички роботи в команді;володіння типовими підходами до проектування захищених об`єктів інформаційної діяльності ;формулювати технічні вимоги до спеціальних технічних і програмно-апаратних засобів захисту, обробки та управління інформацією в інформаційно-комунікаційних системах;розробляти та оцінювати моделі і політику безпеки на основі використання сучасних принципів, способів та методів теорії захищених систем; виконувати аналіз ризиків та джерел загроз, розробляти модель загроз, розробляти модель порушника; - виконувати аналіз і вибір дисципліни обслуговування заявок для комп’ютерних систем з врахуванням режимів роботи, вимог стосовно обслуговування заявок, інтенсивності потоків заявок, дисперсії часу очікування; обгрунтовувати та реалізовувати систему захисту інформаційних ресурсів з обмеженим доступом на об'єктах інформаційної

діяльності; уміти проводити атестацію зон, приміщень тощо в умовах додержання режиму із зафіксуванням результатів у відповідних документах; здатність використовувати професійно-профільовані знання й практичні навички з математики, фізики, онов теорії кіл, сигналів та процесів у комп'ютерних системах і мережах, метрології та вимірювання, теорії інформації та кодування; здатність здійснювати правове забезпечення інформаційної безпеки, прогнозування та моделювання в соціальній сфері; здійснювати оцінку відповідності системи захисту інформації автоматизованої системи своєму призначенню відповідно до вимог діючих стандартів; вміти аналізувати ризики для оцінки реальних загроз порушення захисту та охорони; на основі аналізу обирати систему захисту та проводити розробки пропозицій по її удосконаленню.

Для передавання веб-сторінок, що належать до захищеної частини сайту, замість протоколу HTTP використовується протокол HTTPS, тобто поєднання протоколів HTTP та SSL. Захищені сайти зазвичай вимагають введення імені користувача та пароля. У браузерах реалізовані останні досягнення в галузі шифрування та технологій захисту й безпеки.

Працюючи в Інтернеті, важливо знати про загрози, які існують в мережі. Основні типи шкідливих програм та принципи їх дії. Хробаки — програми, що самостійно поширюються мережею, не інфікуючи інші файли. Трояни — програми, що поширюються під виглядом нешкідливих програм і виконують несанкціоновані дії: викрадають інформацію і передають злочинцям через Інтернет, самостійно відкривають сайти для здійснення хакерських атак тощо. Скрипт-віруси — програми, що потрапляють у комп'ютер через електронну пошту, маскуючись під вкладені документи. Дропери — виконувані файли, що самі не є вірусами, але призначені для встановлення шкідливих програм. Боти — програми, що дають можливість зловмиснику таємно керувати вашим комп'ютером. Рекламні програми — програми, що зазвичай встановлюються на комп'ютер разом із безкоштовними програмами й збирають конфіденційну інформацію або демонструють нав'язливу рекламу. Загрозу також становить фішинг як різновид інтернет-шахрайства: виманювання конфіденційної інформації через підробні сайти, які копіюють сайти відомих банків, інтернет-магазинів тощо, або за допомогою спаму. Для запобігання інтернет-загрозам між комп'ютером і мережею встановлюють перешкоди — міжмережеві екрани. Брандмауери захищають комп'ютер від зловмисного проникнення або потрапляння шкідливих програм. Але не

запобігають витоку конфіденційної інформації користувача та завантаженню вірусів. Браузери Mozilla Firefox, Safari, Opera, Google Chrome мають багато вбудованих засобів захисту. Одним із найпопулярніших браузе-рів для комп'ютерів, телефонів і планшетів є Google Chrome, який:попереджає про відкриття сайта із загрозою фішингу або шкідливих програм;ізольовано відкриває веб-сторінки, що в разі загрози приводить до закриття лише однієї шкідливої веб-сторінки;дозволяє вимкнути збереження конфіденційних даних;надає можливість налаштувати показ спливних вікон. Браузери можуть забезпечувати захист даних за допомогою електронного цифрового підпису — цифрового аналогу звичайного підпису, яким можна скріпити всі електронні документи. Це надасть їм юридичної сили, гарантуватиме цілісність. Електронний цифровий підпис та підтверджувальний сертифікат можна отримати в будь-якому Центрі сертифікації ключів, подавши відповідну заяву та пакет документів. Для забезпечення інформаційної безпеки в Інтернеті недостатньо захистити дані на комп'ютері-клієнті або комп'ютері-сервері. Захист даних забезпечується спеціальним криптографічним протоколом шифрування даних під час їхнього передавання. Захищений сайт — це сайт, який використовує для обміну даними протоколи захищеного зв'язку.Для захисту загроз можна встановити антивірусну програму, яка буде відслідковувати переходи на фішингові сайти, однак повного захисту, покладаючись тільки на програму, не буде. Це обумовлено тим, що база фішингових сайтів дуже швидко оновлюється, а серленя тривалість життя подібного сайту зазвичай не перевищує кількох днів. Тому, якщо фішинговий сайт з'явився зовсім недавно, жодна антивірусна програма не зможе точно визначити, чи несе загрозу користувачеві перехід на даний ресурс. Єдиним варіантом попередження можуть служити бази, в яких вбиті адреси звідки ведуть з вами переписку і у разі виявлення підозрілого листа з'явиться повідомлення про це. Загалом програми - аналізатори пакетів призначені для контролю за мережею, проте вони ж використовуються хакерами для несанкціонованого збирання інформації. Хробак схожий на вірус тим, що розмножується, роблячи власні копії, але на відміну від останнього він не потребує носія й існує сам по собі. Часто хробаки передаються через електрону пошту. Нині щорічно з'являються тисячі вірусів. Якщо 1990 року їх було десь між 200 та 500, то у 2017 році — вже 60 000.Нові віруси та інші методи вторгнення до вашої комп'ютерної системи виникають майже щодня.

Крім програм, за допомогою намагаються проникнути до системи, існують також засоби, що застосовуються для спостереження за вами. Це насамперед програмне забезпечення, яке називають adware та spyware, шпигунські програми, програми для батьківського контролю, блокуючи програми тощо. Таке програмне забезпечення має багато функцій. Воно може відстежувати ваші звички стосовно мандрування Інтернетом, надсилати комусь дані без вашого дозволу, змінювати адресу домашньої сторінки вашого браузера і навіть змінювати системні файли комп'ютера.Інформацію про відвідувані веб-сторінки також можна отримати із cookie-файлів.Програми типу spyware без вашого дозволу надсилають комусь інформацію про те, що ви робите в Інтернеті. Зазвичай це здійснюється в рекламних цілях. Програмне забезпечення типу spyware також сповільнює роботу системи і навіть призводить до її збоїв.

Висновок

Існує безліч причин, з яких певні особи застосовують програми, що стежать за нашими діями, аналізують вашу електронну пошту та фіксують адреси відвідуваних вами веб-сторінок. Простіші програми створюють журнальні файли, де фіксується інформація про те, коли, хто і який сайт відвідував. Більш складні програми здатні відстежувати кожне натискання клавіші на комп'ютері та надсилати цю інформацію особі, що здійснює стеження. Існує багато засобів, які утруднюють несанкціоноване отримання персональної інформації. На деяких інтернет-порталах, зокрема на MSN, також є засоби блокування доступу до подібних інтернет-ресурсів. Існує багато програм, які містять функції блокування. Крім того, у більшість браузерів вбудовано функції, що дозволяють користувачеві підключатися до певних сайтів лише після введення паролю. Усі такі програми діють майже однаково. Програма встановлюється на комп'ютер. Коли користувач вводить адресу сайту, програма її перевіряє, звертаючись до бази даних заборонених сайтів. Якщо ця адреса є в базі даних, програма блокує доступ до сайту, і користувач не зможе до нього підключитися, доки не введе пароль. Якщо ж адреси в базі даних немає, програма сканує сам сайт у пошуку певних заборонених слів і тільки після цього надає користувачеві доступ до сайту. Більшість подібних програм щомісяця оновлюють свою базу даних, завдяки чому ця інформація завжди актуальна, незважаючи на зростання кількості інтернет-ресурсів.

Правила безпеки, яких слід дотримуватися під час передавання інформації Інтернетом. Не надавайте більше інформації, ніж потрібно. Захищені сайти зазвичай вимагають введення імені користувача та пароля. Робіть його довжиною щонайменше вісім символів, комбінуючи букви та числа. Зручно мати два пароля: один для так званих розважальних сайтів, тобто ігор, чатів тощо, а інший для більш важливих дій, наприклад для придбання товарів. Тоді зменшиться ймовірність, що ваші важливі дії піддаватимуться ризику. І ніколи не використовуйте для паролів такі дані, як дата народження, номер телефону чи ім'я. Користуйтеся останньою версією браузера. У новіших браузерах реалізовані останні досягнення в галузі шифрування та інших технологій захисту й безпеки. Уважно читайте правила безпеки сайту. Саме чати та системи обміну миттєвими повідомленнями ці особи обирають для налагоджування контактів з молодими людьми, оскільки почуваються там безпечно. З початком широкого використання міжнародних мереж передачі даних загального користування темпи росту мережної злочинності зростають в геометричної прогресії. За оцінками експертів Міжнародного центру безпеки Інтернет (CERT) кількість інцидентів пов’язаних з порушенням мережної безпеки зросла в порівнянні з 2000 роком майже у 10 разів. Основними причинами, що провокують ріст мережної злочинності є недосконалі методи і засоби мережного захисту, а також різні уразливості у програмному забезпеченню елементів, що складають мережну інфраструктуру. Основними джерелами небезпек для користувачів Інтернету являється діяльність хакерів, вірусів та спамів. Існують засоби, що утруднюють доступ до чужих комп'ютерів – це брандмауери, антивірусне та антиспамове програмне забезпечення. Велике значення також має дотримання користувачами правил безпеки під час роботи в Інтернеті. Для отримання персональної інформації, небезпечні особи використовують чати, системи обліку миттєвими повідомленнями та сайти знайомств. Дотримання простих правил в спілкуванні через мережу Інтернет, дозволить захистити користувача від недобрих намірів зловмисників. Комп'ютерна безпека є важливою практично для всіх технологічних галузей, що функціонують за участю комп'ютерних систем. Авіаційна галузь є особливо важливою з точки зору комп'ютерної безпеки, оскільки пов'язані із нею ризики стосуються життя людей, цінного обладнання й вантажів, а також авіатранспортної інфраструктури. Безпека системи може опинитися під загрозою через неправильне функціонування апаратного і програмного

забезпечення, помилкові дії оператора чи неполадки, пов'язані із середовищем, в якому працює комп'ютерна система. Причиною таких загроз, що реалізуються через уразливість комп'ютерних систем, може бути саботаж, шпіонаж, промислова конкуренція, тероризм, технічні неполадки та людський фактор. Наслідки успішних навмисних чи ненавмисних злочинних дій із комп'ютерною системою в авіації можуть бути різноманітними, від розголошення конфіденційної інформації до порушення цілісності системи, що може призвести до таких серйозних проблем, як витік інформації, простій у роботі мережі чи системи управління повітряним рухом, а це, у свою чергу, може призвести до призупинення роботи аеропорту, втрати повітряного судна, загибелі пасажирів. Ще більшому ризикові піддаються військові системи, що керують майном та обладнанням військового призначення.Під час проектування комп'ютерної системи необхідно взяти до уваги чимало характеристик, безпека є серед них однією з найважливіших. За даними опитування, проведеного корпорацією Symantec у 2010 році, 94% організацій, що взяли участь в опитуванні, планували вжити заходів з підвищення безпечності їхніх комп'ютерних систем, а 42% зазначили, що вважають неналежний рівень кібербезпеки за основний ризик. Незважаючи на те, що більшість організацій вдосконалюють системи інформаційного захисту, чимало кіберзлочинців знаходять шляхи їхнього обходу і продовжують свою діяльність. Нині спостерігається зростання числа майже усіх типів кібератак. В опитуванні 2009 року щодо комп'ютерних злочинів і кібербезпеки, проведеному Інститутом комп'ютерної безпеки, респонденти відзначили суттєве зростання числа атак шляхом застосування зловмисних програмних засобів, DoS-атак, викрадання паролів та дефейсу сайтів.

Литература

1. Огірко І., Огірко О. Інформаційні технології безпекометрії в поліграфії. ІІІ-я Міжнародна науково-технічної конференції «Захист інформації і безпека інформаційних систем». Національний університет «Львівська політехніка», 5—6 червня. — Львів, 2014. — С. 46—51.

2. Гавловский В. Інформаційна безпека : захист інформації в автоматизованих системах -організаційно-правовий аспект [Електронний ресурс] / В. Гавловский // Портал : bezpeka.com. — Режим доступу www URL: http://www.bezpeka.com/ru/

lib/spec/law/information-security-automated-systems.html. — Заголовок з екрана, доступ вільний, 27.03.2015.

3 Черкун О. М. Сучасні технології комп'ютерної безпеки / О. М. Черкун // Сучасні технології комп'ютерної безпеки : колективна монографія. — Рівне : МЕГУ, 2012. — 90 с.

4. Кунченко-Харченко В. І., Огірко І. В. Оптимізація пошукової моделі відбору даних з використанням Ku простору // Обробка сигналів і негаусівських процесів. Праці V міжнародної науково-практичної конференції. Черкаський державний технологічний університет. — Черкаси, 2015. — С. 103—107.

5. Орлик О. В. Економічна безпека підприємства: властивості, стратегія та методи за- безпечення / О. В. Орлик // Економічна безпека в умовах глобалізації світової економіки :— Дніпропетровськ : «ФОП», 2014. — Т. 2. — С. 176-182.

6. Корольов М. В. Проблематика дослідження питань інформаційної безпеки у держав- ному управлінні // М. В. Корольов, О. О. Скопа / Вісник Східноукраїнського національного університету імені Володимира Даля. — Луганськ : СНУ ім. В. Даля. — 2013. — №15(204). — Ч. 1. — С. 88-93.

7. Орлик О. В. Система загроз економічній безпеці суб'єктів господарювання / О. В. Орлик // Вісник соціально-економічних досліджень : зб. наук. праць. — Одеса : ОНЕУ. — 2014. — Вип. 1(52). — С. 250-257.

8. Ших Ю. А., Огірко І. В. Віртуальне паломництво — інтернет технології для задоволення духовних потреб. Поліграфія і видавнича справа. — Львів : УАД. — № 2 (58). — 2012. — С. 77—81.

9. Огірко І. В., Ясінський М. Ф., Пілат О. Ю. Інформаційна система оцінки якості електронних видань // Комп'ютерні технології друкарства: наук.-техн. зб. — Львів : Українська академія друкарства. — № 31. — С. 96—104.

10. Йона О. О. Світові тенденції боротьби з кіберзлочинністю / О. О. Йона, Н. Ф. Казакова // Вісник Східноукраїнського національного університету імені Володимира Даля. — 2013. — № 15(204). — Ч. 1. — С. 59-62.

11. Огірко І. В., Огірко О. І. Математична модель оцінювання якості та захисту WEB. Науково-технічна конференція професорсько-викладацького складу, наукових працівників і аспірантів (16–19 лютого 2016 р.). Українська академія друкарства. Тези доповідей. — 2016. — С. 133.

12. Орлик О. В. Методи управління фінансово-економічною безпекою / О. В. Орлик // Сборник научных трудов SWorld. — 2014. — Т. 28. — №. 1. — С. 37-41.

13. Казакова Н. Ф. Дослідження та застосування в системах захисту інформації кореля- ційного критерію подібності графічних структур / Н. Ф. Казакова, О. О. Фразе- Фразенко // Системи обробки інформації. — 2014. — № 2 (118). — Т 2. — С. 246.

14. Фразе-Фразенко О. О. Спосіб регуляризації некоректно поставленої задачі розпізна- вання у системах телебачення замкнутого контуру / О. О. Фразе-Фразенко // Східно-Європейський журнал передових технологій. — 2012. — № 6/4(8). — Спецвипуск (4). — С. 19-20.

КЕРУВАННЯ ПЕРЕВАНТАЖЕННЯМ КОМП'ЮТЕРНОЇ МЕРЕЖІ

Кучеров Д.
Національний авіаційний університет,
м. Київ, Україна

Перевантаження мережі є однією з основних проблем, з якою періодично стикаються користувачі комп'ютерних мереж. Вона викликає зниження перепускної здатності мережі, збільшення часу проходження пакетів або їх втрату. Саме це явище призводить до припинення дії деяких мережевих служб, таких як VoIP, інтерактивні додатки, чат, доступ до віддалених ресурсів та інших. Останнім часом, коли спостерігається експоненціальне зростання мереж, ця проблема при їх експлуатації стає найбільш загостреною.

Одним з чинників перевантаження вважається надлишкова буферизація каналу передавання [1]. Буфер необхідний при передачі даних по лінії зв'язку, якщо у відправника та отримувача різний темп оброблення. При цьому в буфері затримується передавання пакетів на час приймання та первинного оброблення отримувачем. Наповнення буферу приводить до втрат переданих пакетів. Це явище може спостерігатися в маршрутизаторах, бездротових точках доступу, мостах, шлюзах, пристроях супутникового зв'язку.

Виключення перевантаження вирішується декількома способами. Запобігти перевантаженню буфера можна досягти завдяки керуванню чергами та методами передбачення. Методи керування перевантаженням контролюють перевантаження після того, коли воно виникає, в той час, як методи передбачення усувають перевантаження шляхом контролю інтенсивності передавання даних у мережі, що дозволяє передбачити перевантаження та запобігти його в типових "вузьких місцях" мережі. Основним засобом мережевих операційних систем із запобігання перевантажень в Cisco є застосування алгоритмів зваженого випадкового раннього виявлення (WRED) [2].

В дуплексному режимі контролю портів комутатора можлива реалізація механізму зворотного зв'язку, який введений для мереж Ethernet специфікацією IEEE 802.3x. Механізм реалізується введенням підрівня керування рівнем MAC, на якому вводиться параметр часу зупинки передачі кадрів іншим вузлам. Час вимірюється в 512 бітовими інтервалами конкретної реалізації

Ethernet, діапазон можливих варіантів зупинки знаходиться в інтервалі 0 – 65535. Після завершення часу зупинки передача відновлюється.

Перевантаження може бути виключеним, якщо вводиться резервування перепускної здатності на підставі двійкових методів. При цьому користувачу додатками QoS надається частина перепускної здатності каналу, інша частина резервується іншим користувачам, що здійснюється за рахунок використання логічного з'єднання. Але це відбувається на рівні обладнання, яке вводить пріоритети певним користувацьким додаткам, наприклад, таким як відеоконференція. Таким чином, все мереже обладнання, що є в каналі передачі, повинне підтримувати цю технологію, але це не завжди виконується.

Мобільні мережі типовим рішенням завдання перевантаження вважається реконфігурація мережі. В статті проводиться аналіз можливості відтворення перепускної здатності комп'ютерної мережі, формування трафіку в різних умовах мережевих підключень, визначається конфігурація з меншим навантаженням, встановлюються відношення між падінням трафіку та його відновленням завдяки резервуванню.

Постановка проблеми

Будемо розглядати комп'ютерну мережу, вузли якої здатні обробляти інформацію та обмінюватися даними, крім того дозволяється зміна конфігурації системи. За способом управління мережа відповідає архітектурі «клієнт-сервер».

Мережа має певну топологію, яка описується зваженим графом $G = (V, E)$, в якому V – вузли, число вузлів $V(G) = N$, а E – зважені дуги, число дуг $G(E) = M$. Кожній дузі графу (i, j) ставиться у відповідність число $w_{ij} > 0$, яке зветься вагою дуги (i, j), $i = 1..N$, $j = 1..M$. В разі, коли $(i, j) \notin G$, $w_{ij} = \infty$.

Шлях від вузла s до вузла f довжиною l називають впорядковану послідовність $(E_s, \ldots, E_f)_l = (E_s, E_i), \ldots, (E_{i+q}, E_{i+q+1}), \ldots, (E_l, E_f)$.

Зваженою довжиною шляху від s до f називають число

$$L(s, f) = \sum_{(j, j+1) \in (E_s, \ldots, E_f)} w_{i,j} \cdot \qquad (1)$$

Введемо множину усіх припустимих шляхів Ξ_{sf}, тоді найкоротший шлях визначається виразом

$$L'_{sf} = \min_{(E_s,...,E_f)\in\Xi_{sf}} \sum_{(j,j+1)\in(E_s,...,E_f)} w_{i,j} \cdot \quad (2)$$

Ставиться завдання визначити можливість щодо визначення найкоротшого шляху в комп'ютерній мережі в умовах впливу природних чи штучних перешкод на систему (1).

Задача маршрутизації. Передача даних в мережі відбувається за правилами, що регламентуються застосовуваними протоколами, згідно яких встановлюється адреса, розмір і структура пакету, що передається, швидкість передачі даних.

Адреса може бути індивідуальною, коли повідомлення посилається тільки одному агенту, груповою, коли повідомлення розсилається кожному агенту в групі, або деякій підгрупі агентів. Такий спосіб адресації подібний до адресації в комп'ютерних мережах. Передача даних відбувається від адреси джерела до адреси приймача.

Якщо кількість вузлів значна, тоді виникає необхідність передавати інформацію через транзитні вузли. Але при цьому стає необхідність прокладання маршруту щодо забезпечення мінімуму часу на доставку інформації, що досягається мінімумом транзитних вузлів або наявністю каналів з високою пропускною здатністю та надійністю ліній зв'язку. Зрозуміло, що невелике збільшення кількості транзитних вузлів, що мають велику пропускну здатність, є переважним над мінімальною кількістю вузлів з невеликою пропускною здатністю з точки зору часу передачі інформації. Надійні канали зв'язку характеризуються мінімальними втратами інформації, що передається.

Через вузол може проходити декілька підпотоків, їх відрізняють за адресом пункту призначення. Зрозуміло, щоб визначити маршрут, який би забезпечив рівний час потоків різного обсягу даних, необхідно враховувати швидкість передачі, яку мають окремі лінії, та надавати можливість проведення розпаралелювання підпотоків та їх збірку. Переключення вузлів для передачі підпотоків здійснюється мультиплексуванням вільних каналів. Дана задача складається у визначенні маршруту проходження інформації кінцевому агенту. Розрізнюють статичні та динамічні маршрути. Статичний маршрут задається одноразово або за певним розкладом та не змінюється в межах певного часу. Динамічні маршрути обчислюються відповідними алгоритмами в залежності від топології та стану

інформаційної мережі. До них відносяться алгоритми пошуку найкоротшої відстані в ширину, Дейкстри, Беллмана-Форда [9, 12].

Пошук в ширину. Проводиться обхід кожної вершини $V_i \in V$ графа G, запам'ятовується кількість пройдених дуг $E_i \in E$, Мінімальна відстань $L(s, f)$ між точками s та f відповідає найменшій кількості дуг, що з'єднують вершини s і f

$$L'(s, f) = \min_{(E_s,\ldots,E_f)\in\Xi_{sf}} \sum_{i=1}^{N} E_i \qquad (3)$$

Приклад застосування алгоритму наведений на рис.1.

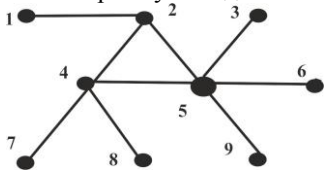


Рис. 1. Комп'ютерна мережа, що складається з 9 вузлів.

Найкоротший маршрут передачі інформації між вершинами 1-9 за алгоритмом (2) відповідно до рис. 1 складає $L(1, 9)=3$. Оцінкою продуктивності графа виступає часова складність $O(\cdot)$, що визначається кількістю операцій за алгоритмом (2). За цим алгоритмом усі вузли та ребра скануються одноразово, тому часова складність визначається їх кількістю, а саме $O(M+N)$.

Алгоритм Дейкстри є процедурою пошуку найкоротшого шляху на зваженому орієнтованому графі. Алгоритм використовується протоколами маршрутизації OSPF та IS-IS в IP-мережах [13, 14]. Ребра графа мають вагу $\omega(i, j)$ таку, що

$$w_{ij} = \begin{cases} 0, \text{ якщо } i = j, \text{ тобто та сама вершина}, \\ w > 0, \text{ якщо } |i - j| < 2 \text{ тобто сусідні вершини}, \\ \infty, \text{ якщо інакше}. \end{cases} \qquad (4)$$

Довжина шляху на кожному кроці k від вершини s визначається правилом

$$L_k = \min_V [L(k), L(V) + w_{ij}] \text{ для } V \notin G, L(s)=0. \qquad (5)$$

Правило (4) не дозволяє робити проходи по дугам графа G з великою вагою. Таким чином, множина вершин в графі G являє собою впорядковану послідовність зв'язаних між собою вузлів, яка

містить найкоротший шлях від вершини *s* до *k*. Цей шлях показаний на рис. 2 стрілками.

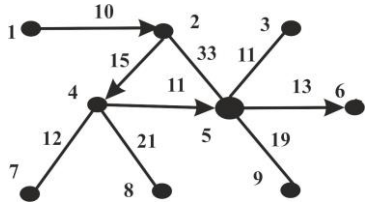


Рис. 2. Структура зваженого графа.

В кожний розрахунковий момент часу на маршрутизатор *i* поступає сумарний потік інформації, призначений для передавання кожному маршрутизатору *j*. Цей потік визначає таблицю маршрутизації, яку можна визначити матрицею $P(t)$ розмірності $N \times N$ з нульовою головною діагоналлю

$$P(t) = \sum_{i,j} c_{ij}(t), \quad k \in K, \tag{6}$$

де K – множина потоків інформації, яка зв'язана з відповідними IP-адресами, а c_{ij} – пропускна здатність. Компанія Cisco в маршрутизаторах визначає ваги дуг за формулою

$$w_{ij} = \frac{10^8}{c_{ij}(t)}, \tag{7}$$

Оскільки алгоритм є ітераційним, то число ітерацій визначається кількістю вершин графа, тому часова складність алгоритму $O(N)$. В межах кожної ітерації, відбувається нове проходження з врахуванням нової (j+1) вершини. При цьому вершини з найбільшою вагою вивільняються, а довжина шляху з новими вершинами поновлюється, кращий результат запам'ятовується. Це те ж оцінюється кількістю вершин. Загальна продуктивність алгоритму оцінюється величиною $O(N^2)$. Таким чином, алгоритм Дейкстри є ресурсоємним, але завдяки знанням топології мережі і шляху до потрібної вершини, маршрутизатор завжди знаходить альтернативний шлях до потрібного вузла мережі у випадку виникнення проблем у будь-якому вузлу визначеного шляху.

Резервування перепускної здатності повинно враховувати навантаження мережі. Для контролю маршрутизатор доповнюється засобами вимірювання навантаження, які будуть створювати аналогічно (6) матрицю навантажень $X=|x_{ij}|$. Тоді алгоритм з резервуванням трафіку може бути записаний як

$$\bar{c}_{ij}(t) = \begin{cases} c_{ij} + \Delta, & \text{якщо } c_{ij} < x_{ij}, \\ c_{ij}, & \text{якщо інакше,} \end{cases} \tag{8}$$

де Δ - частка, що компенсує перевантаження.

Алгоритм Беллмана-Форда. За суттю цей алгоритм нагадує попередній (Дейкстри). На відміну від алгоритму Дейкстри цей алгоритм не відкидає ребр з великою вагою та ітераційно розраховує довжину усі шляхи в графі, запам'ятовуючи мінімальний шлях, використовується протоколом маршрутизації RIP. Кількість ітерацій, як і алгоритмі Дейкстри визначається кількістю вершин, а кількість розрахунків в межах ітерації кількістю ребр, то часова складність алгоритму оцінюється величиною $O(V \cdot E)$. Результат цього алгоритму для графа рис.2 співпадає з алгоритмом Дейкстри. При однакових розмірах графа алгоритм Дейкстри є менш ресурсоємним ніж Беллмана-Форда, а значить більш швидкий. Зменшення ресурсоємності можна досягти зменшенням кількості ребр, тобто переходом до розрідженого графу.

Зміна структури системи. Під реконфігурацією інформаційної системи будемо розуміти зміну будови системи, яка стосується розміру чи її топології. Особливістю реконфігурації є перебудова структури та топології системи з метою виключення перевантажень та збоїв в роботі системи. Вважається, що мережа виконує свої функції, поки між вузлами здійснюється обмін інформації. Приклад реконфігурованої системи показаний на рис.3.

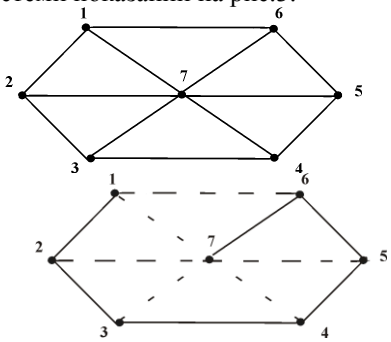


Рис. 3. Зміна будови інформаційної системи в результаті втрати зв'язку.

Таким чином, внаслідок перевантажень чи збоїв топологія локальної повнозв'язної мережі змінюється до «шини».

Аналіз реконфігурації комп'ютерної системи. Відповідно до загального уявлення про комп'ютерну мережу (1) вона є системою з обмеженою кількістю можливих станів. Тому її поведінку можна промоделювати за допомогою математичного апарату аналізу марківських ланцюгів.

Будемо вважати, що в процесі роботи комп'ютерна мережа, яка складається з кінцевої множини елементів N, обмінюється інформацією між всіма елементами за протоколами OSPF або RIP, що відповідає умовам нормального функціонування системи. Початковий стан мережі позначимо S_1. Якщо за якимись не випадковими причинами починають виходити з ладу елементи мережі чи може пропадати обмін інформацією, то відбувається перехід системи до іншого стану. Будемо вважати, що елементи системи не перевантажуються одночасно, а по одному, тому здійснюються послідовні переходи зі стану S_1 у стани S_2, S_3, …, S_i через певні інтервали часу Δt, i означає номер стану. Послідовний набір станів комп'ютерної мережі та переходів між ними утворює ланцюг Маркова. Оскільки ланцюг послідовний, то функціонування системи можна подати у вигляді схеми «загибель-розмноження» [15].

Нехай комп'ютерна мережа складається з n вузлів, тому відповідно до підходу за схемою «загибель-розмноження» введемо стани S_i, i=1..n+1, де S_1 – стан мережі, що відповідає функціонуванню всіх вузлів без перевантажень. За умови впливу зовнішніх і внутрішніх факторів погіршується перепускна здатність каналів в мережі з фіксованою інтенсивністю λ. При цьому здійснюється перехід в стани, коли вузли поступово втрачають пакети, послідовно один за другим. Таким чином, здійснюється перехід в стан S_2, коли не працює 1 вузол, та так далі, а відповідно S_{n+q} – комп'ютерна мережа перестала виконувати завдання у зв'язку з виходом з ладу $(n - q)$-вузлів. Система також може вживати заходи щодо нарощування перепускної здатності, що відбувається з інтенсивністю $\mu > \lambda$. При цьому з інтенсивністю μ відбуваються послідовні переходи зі станів S_i у стани S_{i-1}. Знайдемо ймовірності p_i знаходження комп'ютерної мережі в кожному з кінцевих станів S_i та проаналізуємо їх. Ймовірності знаходження системи у фінальних станах знаходяться за формулами [15]

$$p_i = p_0 \frac{\prod_{k=1}^{i} \lambda^k}{\prod_{k=1}^{i} \mu^k}, \quad i \neq 0, \qquad p_0 = \left(1 + \sum_{i=1}^{n} \frac{\prod_{k=1}^{i} \lambda^k}{\prod_{k=1}^{i} \mu^k} \right)^{-1}. \tag{9}$$

Розглянемо цю задачу для випадку n = 5, λ = 0,5, μ_1 = 0,8, μ_2=1,6, μ_3=2,4. Результати розрахунку за формулами (9) наведені на рис.4. Аналіз рисунку показує, що чим більше μ, тим краще система справляється с пакетами, більше ймовірність знаходження її у стані з мінімальними затримками та менше ймовірність знаходження у стані з найбільшими затримками.

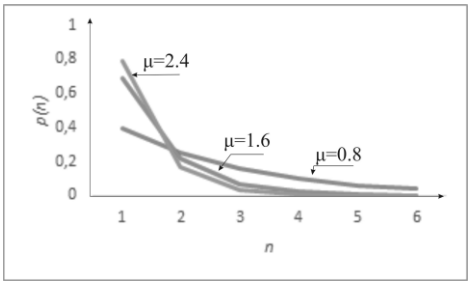


Рис. 4. Деградація комп'ютерної мережі в умовах інтенсивних навантажень.

Як видно з рисунку, якщо $\mu > (2..5)\ \lambda$, то система стає більш стійкою до навантажень різних типів.

Висновки. На підставі проведеного аналізу протоколів та алгоритмів маршрутизації встановлено, що ефективність комп'ютерної мережі визначається можливістю її функціонування в умовах перевантажень та збоїв, що є наслідком надлишкової буферизації системи. Одним з ефективних способів зменшення впливу перевантажень на мережу є резервування перепускної здатності каналів та компенсація її частки в каналах, які піддаються найбільшому впливу. Відповідно до проведеного аналізу мережі за схемою «загибель-розмноження» встановлена, що дія система функціонує, якщо відношення інтенсивності перевантажень до інтенсивності приросту перепускної здатності не перевищує значення 0,2…0,5.

Подальші дослідження функціонування комп'ютерної мережі планується зосередити на аналізі її динамічних властивостей.

Література

1. Gettys J. Bufferbloat: Dark Buffers in the Internet / J. Gettys // Internet Computing. – IEEE Computer Society, 2011. – Vol. 15. – № 3. – P. 96-95.
2. Руководство по технологиям объединенных сетей / Пер. с англ. – М.: ИД «Вильямс», 2005. – 1040 с.
3. Давиденко И.Н. Способ определения местоположения агентов домена в мобильных сетях / И.Н. Давиденко, И.А. Жуков // Проблеми інформатизації та управління. – №1 (23). – 2008. – С. 61-64.
4. Ластовченко М.М. Метод анализа эффективности реконфигурации топологии беспроводных мультисерверных сетей повышенной помехозащищенности / М.М. Ластовченко, Е.Е. Зубарева, В.О. Саченко // УСиМ. – №6. – 2009. – С. 79-86.
5. Кузьмичов А.І. Оптимізаційні моделі мережевих структур / А.І. Кузьмичов, Є.О. Додонов // Реєстрація, зберігання і обробка даних. – Т.19. - № 2. – 2017. – С. 24 – 35.
6. Kucherov D.P. Control System Objects with Multiple Stream of Information / D.P. Kucherov, A. N. Kozub // Proceedings 2015 IEEE 3rd International Conference "Actual Problems of Unmanned Aerial Vehicles Developments (APUAVD)", October 13-15, 2015. – P. 290-293.
7. Горбунов И. Э. Методология анализа и синтеза реконфигурируемых топологий мобильных сетей связи / И. Э. Горбунов // Математичні машини і системи. – №2. – 2006. – С. 48-59.
8. Кучеров Д.П. Реконфігурація мультисенсорної системи за умови впливу дестабілізуючих факторів / Д.П. Кучеров // Сенсорна електроніка і мікросистемні технології. – Т.13. – №2. - 2016. – С. 101-112.
9. Столлингс В. Современные компьютерные сети / В. Столлингс. – СПб.: Питер, 2003. – 783 с.
10. Олифер В.Г. Компьютерные сети: принципы, технологии, протоколы / В.Г. Олифер, Н.А. Олифер. – СПб.: Питер, 2006. – 958 с.
11. Таненбаум Э. Компьютерные сети / Э. Таненбаум, Д. Уэзеролл. – СПб.: Питер, 2012. – С. 960.

12. Кормен Т. Алгоритмы. Построение и анализ / Т. Кормен, Ч. Лейзерсон, Р. Ривест, К. Штайн. – М.: ИД «Вильямс», 2005. – 1296 с.
13. Кузнецов Н.А. Алгоритм Дейкстры с улучшенной робастностью для управления маршрутизацией в IP-сетях / Н.А. Кузнецов, В.Н. Фетисов // Автоматика и телемеханика. - № 8. – 2008. – С. 80-85.
14. Fortz B. Optimizing OSPF / IS-IS weights in a changing world / B. Fortz, M. Throup // IEEE Journal on selected areas in communications, June, 2002. – P. 1-31. - [електронний ресурс]. – Режим доступу: DOI: 10.1109/JSAC.2002.1003042
15. Вентцель Е.С. Исследование операций: задачи, принципы, методология / Е.С. Вентцель. – М.: Наука, 1988. – 208 с.

APPLICATION OF DECISION-MAKING SUPPORT, NONLINEAR DYNAMICS, AND COMPUTATIONAL LINGUISTICS METHODS DURING DETECTION OF INFORMATION OPERATIONS

Lande D.V.[1] (dwlande@gmail.com),
Andriichuk O.V. [1] (oleh.andriichuk@i.ua),
Hraivoronska A.M. [1] (nastya_graiv@ukr.net),
Guliakina N.A. [2] (guliakina@bsuir.by)
[1]Institute for Information Recording of National Academy of Sciences of Ukraine, Kyiv, Ukraine
[2]Belarusian State University of Informatics and Radioelectronics, Minsk, Republic of Belarus

Introduction

Sources of information have a substantial influence on people. The last several years show that mass media sources can be used efficiently for propagating misinformation. Furthermore, social experiments show that people often believe in the unconfirmed news and disseminate them. For example, in [1] the authors present a review of known false beliefs and misinformation in American society. In [2] the author describes his experiments. He examined belief in political rumors surrounding the health care reforms enacted by Congress in 2010. Depending on the details presented, 17-20% of the respondents believed in them, and 24-35% of the recipients did not have a specific opinion, and 45-58% of the respondents rejected the rumors.

Consequently, investigating processes in the information space is a topical issue. The problems of high significance are processing information streams, identifying trends and anomalies, and detecting critical and meaningful events in real-time mode.

Let us define the information operation [3-4]. The information operation (IO) is the complex of informational events (news on the Internet and media, comments on social networks, forums, etc.), aimed to change the public opinion about a particular object (person, organization, institution, country, etc.). Most of IO have the typical structure (Fig. 1). If the presented IO has the following phases: «background publications» – «calm» – «preparatory bombardment» – «calm» – «attack», then by the first three phases it is possible to predict future events with high probability.

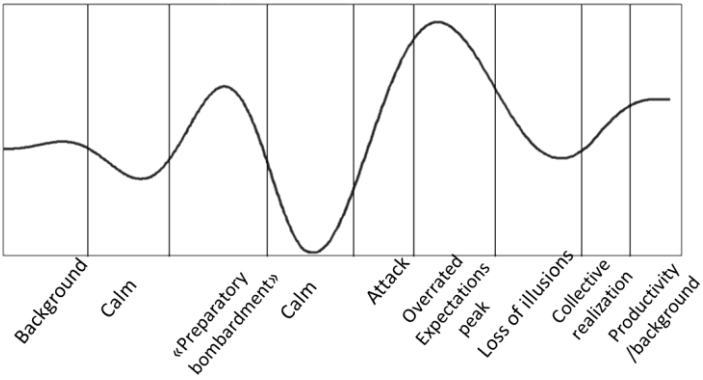


Fig. 1. Information Operation Roadmap.

To illustrate techniques and approaches presented in the article we use the "Brexit" topic. As you may know, a referendum was held on Thursday 23 June 2016, to decide whether the UK should leave or remain in the European Union. Leave won by 51.9% to 48.1%. Naturally, this event is connected with many informational processes on the Internet. Brexit is currently a topical issue that is widely researched by the scientific community [5].

Nonlinear Dynamics Methods for Analyzing Informational Streams

When investigating thematic information streams, we consider how the number of publications changes over time. A content monitoring system provides time series data for a particular topic. In this case, time series is a sequence of numbers of publications dedicated to the topic per day during a specific time. For example, we gathered the time series for the "Brexit" topic (Fig. 2).

To deal with time series data, we use wavelet-analysis [6]. A wavelet is a function that is well localized in time. In practice, we often use the Mexican Hat wavelet:

$$\psi(t) = C(1-t^2)\exp\left\{\frac{t^2}{2}\right\},$$

and the Morlet wavelet:

$$\psi(t) = \exp\left\{ikt - \frac{t^2}{2}\right\}.$$

The essence of the wavelet-transform is to identify regions of the time series which are similar to the wavelet.

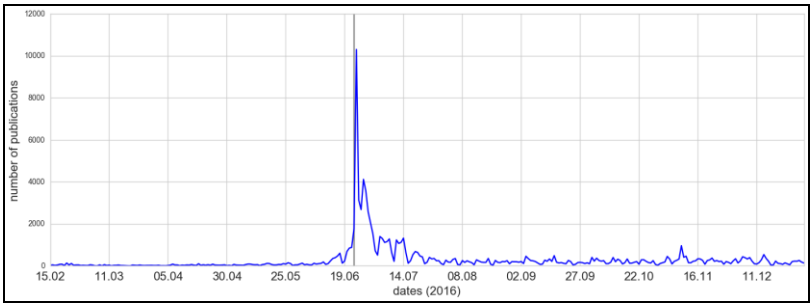


Fig. 2. The Number of Publications on "Brexit" during 2016.

To explore different parts of the original signal with various degrees of detail the wavelet is transformed by stretching/squeezing and moving along the time axis. Therefore, the continuous wavelet-transform has the location parameter (l) and the scale parameter (s). By definition, the continuous wavelet transform of the function $x \in L^2(R)$ is:

$$W(s,l) = \frac{1}{\sqrt{|s|}} \int_{-\infty}^{\infty} x(t)\psi^*\left(\frac{t-l}{s}\right)dt,$$

where $l,s \in R$, $s \neq 0$; ψ^* is complex conjugate of ψ, the values $\{W(s,l)\}$ is called the coefficients of the wavelet transform or the wavelet-coefficients. The wavelet-coefficients are visualized in the plot with a locations axis and a scale axis.

The reason to use the Mexican hat wavelet and the Morlet wavelet is a possibility to detect spikes in time series. The wavelet-coefficients for the "Brexit" time series are shown in Fig. 3 (Mexican hat) and 4 (Morlet). In both cases, the spike in the time series on the second half of June is highlighted strongly. One can also notice smaller spikes in the second part of time series.

Fig. 3. The Wavelet-coefficients for the "Brexit" Time Series Using the Mexican Hat Wavelet.

Fig. 4. The Wavelet-coefficients for the "Brexit" Time Series Using the Morlet Wavelet.

A wavelet must meet the mathematical requirements to be used in the continuous wavelet-transform. Generally, we can explore the correlation of the time series with a pattern of our choice. For example, if we aim to detect the IO, we can use the pattern shown in Fig. 5. The shape of the pattern matches stages of the IO. Now we use the number of points in the pattern instead of the scale.

A pattern can be moving along the time axis in the same way as a wavelet. To calculate each wavelet-coefficient we use the entire time series, but in this case, we need *k* points of the series and a pattern with *k* points to calculate the correlation coefficient by the formula:

$$C(l,k) = \frac{\sum_{i=1}^{k} \left(x(t+i) - \langle x \rangle\right)\left(s(i) - \langle s \rangle\right)}{\sqrt{\sum_{i=1}^{k} \left(x(t+i) - \langle x \rangle\right)^2 \sum_{i=1}^{k} \left(s(i) - \langle s \rangle\right)^2}}.$$

To visualize the results we make the similar plot as for the wavelet-coefficients (Fig. 6). When applying this method, the small spikes in the second part of the time series are highlighted more noticeably.

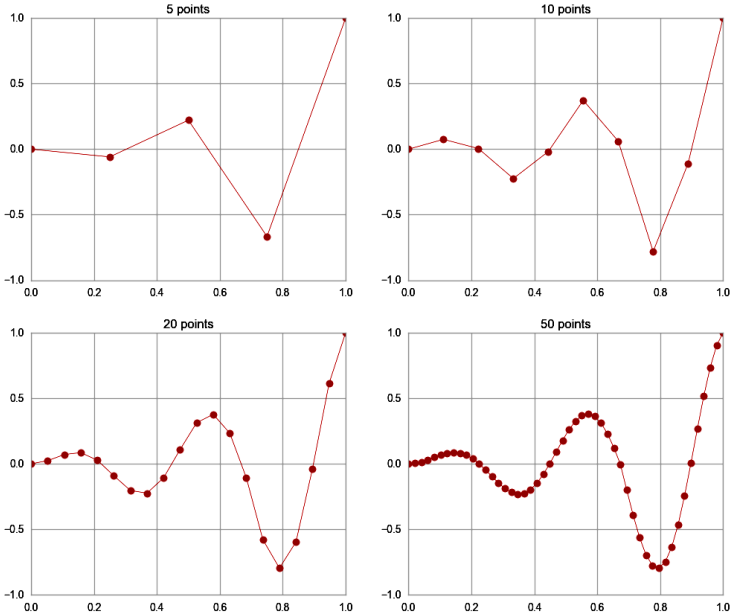


Fig. 5. Pattern with Different Numbers of Points.

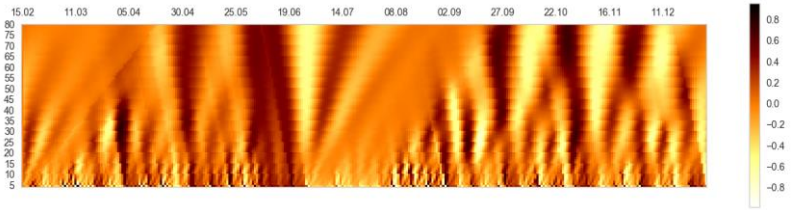


Fig. 6. Correlation Coefficients for the "Brexit" Time Series and the Pattern shown in Fig.4.

Another approach to analyzing time series is the ΔL-method. The ΔL-method is based on the DFA (Detrended Fluctuation Analysis) method [7]. The essence of the approach is to determine and visualize the absolute deviation of the accumulated time series from the corresponding values of linear approximation.

First, let us fix a length of a segment *s*. We split up the time series into overlapping segments. For the point x_t we choose the segment with the

length s and the center at the point t (or at the point $t-1$ if s is even). For each segment fit the points in it with a linear function. Denote the value of local approximation at the point t by $L_{t,l,s}$. Next, calculate the absolute deviation of x_t from the approximation line, as follows $\Delta_{t,l,s} = | x_l - L_{t,l,s} |$. According to the method we calculate values $\Delta_{t,l,s}$ for all $l = 1,...,T$ and $s = 1,...,[T/4]$. Finally, we calculate standard deviations:

$$E(l,s) = \sqrt{\frac{1}{s}\sum_{t=1}^{T} | x_t - L_{t,l,s} |^2} = \sqrt{\frac{1}{s}\sum_{t=1}^{T} \Delta_{t,j,s}{}^2}.$$

Coefficients $E(l,s)$ are plotted in the same way as wavelet-coefficients. We apply the ΔL-method to the Brexit time series and present the results in Fig. 7.

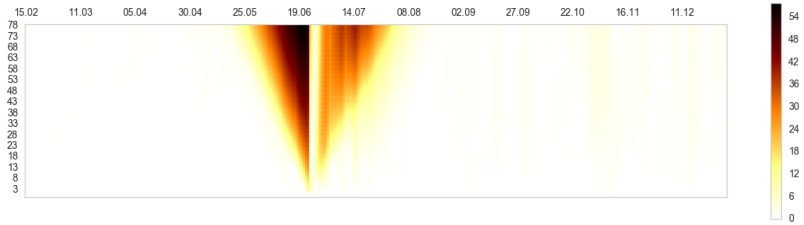


Fig. 7. Coefficients Obtained by the ΔL-method for the "Brexit" Time Series.

The continuous wavelet-transform, the correlation with a pattern, and the ΔL method help to identify spikes, edges, periodicity, and local features of the time series. Plots or scalograms obtained by these methods are used to visualize special features of time series.

Computational Linguistics Methods for Analyzing Informational Streams

In the previous section, we discuss the methods from nonlinear dynamics for analyzing the time series associated with the information stream. In that case, we use only numbers of publications dedicated to a particular topic. In this section, we are going to introduce some

techniques of computational linguistics to extract useful information from the text of the publications.

Many publications and messages on social networks have strong sentiment. When analyzing the information stream, it is useful to determine the attitude of the text. Therefore, we apply methods of sentiment analysis. In the simple case, we classify whether a publication is positive, negative or neutral, using lists of words commonly associated with having a specific sentiment.

The second useful task is to select significant words from the text. For the task to be done, one can apply widely accepted TFIDF score. To calculate TFIDF scores, a text consisting of N words is divided into equal parts of M words. Then for each *i*-th word in the text TF(*i*) is the frequency of the word in the document, and DF(*i*) is the frequency of parts in which the word occurs. The TFIDF score for each word is calculated as following:

$$\mathrm{TFIDF}(i) = \mathrm{TF}(i) \log\left(\frac{N}{M \cdot \mathrm{DF}(i)} \right).$$

Another estimate of word significance is a standard deviation estimate [8]. Let us numerate all words in the text from 1 to *N*. Denote by $A_k(n)$ layout of a certain word *A*, where *k* is the number of occurrence of this word in the text, and *n* is a position of this word in the text. The distance between occurrences of the word *A* is $\Delta A_k = A_{k+1}(m) - A_k(n) = m - n$, where *m* and *n* are the positions of the *k +1*- th and *k* -th occurrences of the word A in the text, respectively. Denote be ΔA the vector $\Delta A_1, \Delta A_2, ..., \Delta A_K$. The standard deviation estimate is:

$$\sigma_A = \frac{\sqrt{\left\langle \Delta A^2 \right\rangle - \left\langle \Delta A \right\rangle^2}}{\Delta A}.$$

Using the TFIDF score or the standard deviation estimate, one can match each word in the text with a numerical value or weight. The next step might be to transform these pairs into a horizontal visibility graph [9]. The horizontal visibility graph (HVG) allows constructing the

keyword network. Such network leads to detecting primary and secondary essential words for understanding the text and significant collocations.

The process of constructing the language network using HVG consists of two stages. At the first step, the traditional HVG is constructed. At the second step, all the nodes corresponding to a single word are combined into a single node and the connections of these nodes are also combined. It was proved that the largest-degree nodes reflect the meaning of the text; thus, these nodes are of informational significance. The keywords can be used to identify subtopics connected with the main topic of interest.

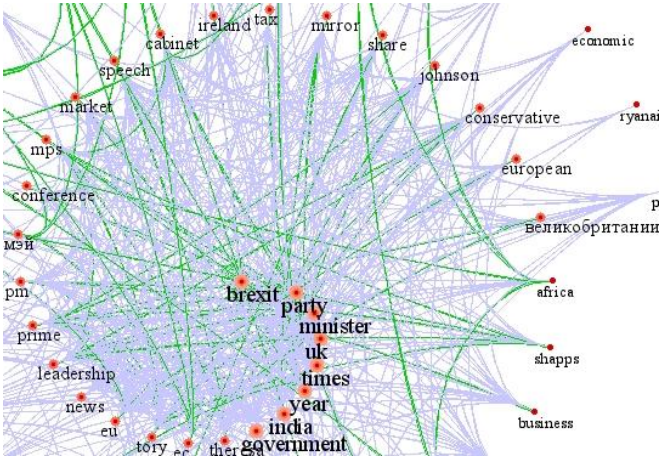


Fig. 1. Keywords for the “Brexit” Topic.

Decomposition of Information Operations Topics and Their Rating Calculation

In publications [10-12] it is shown that IO belong to weakly-structured domains. In order to deal with such domains, decision support systems (DSS) are used [13]. DSS produce recommendations for a decision maker. To do this, they are modeling weakly structured subject domains in their knowledge base (KB).

As part of the “goal hierarchy graph” model, a hierarchy of goals, or KB, represented in the form of an oriented network-type graph (an example for the Brexit is shown in Fig. 8), is built by experts. Nodes (vertices) of the graph represent goals or KB objects. Edges reflect the impact of one set of goals on achievement of other goals: they connect sub-goals to their immediate “ancestors” in the graph (super-goals). Goals can be quantitative and qualitative.

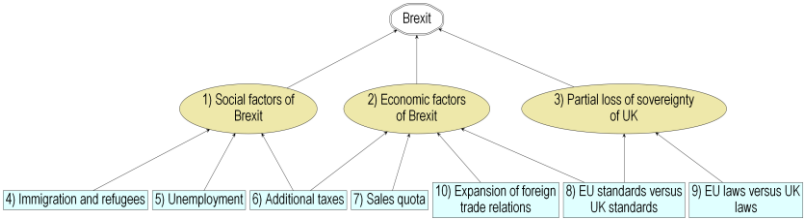


Fig. 8. Structure of Goals Hierarchy of KB for Brexit

For building of the goal hierarchy the method of hierarchic target-oriented evaluation of alternatives is used [13]: a hierarchy of goals is built; then the respective partial impact coefficients are set, and the relative efficiency of projects is calculated. First, the main goal of the problem solution is formulated, as well as potential options of its achievement (projects) that are to be estimated at further steps. After that a two-phase procedure of goal hierarchy graph construction takes place: "top-to-bottom" and "bottom-to-top" [13]. "Top-to-bottom" phase envisions step-by-step decomposition of every goal into sub-goals or projects, that influence the achievement of the given goal. The main goal is to be decomposed into more specific components – sub-goals that influence it. Then these lower-level goals are further decomposed into even more specific sub-components that are, in their turn, also decomposed. When a goal is decomposed, the list of its sub-goals may include (beside just-formulated ones) goals (already present in the hierarchy) that were formulated in the process of decomposition of other goals. Decomposition process stops, when the sets of sub-goals, that influence higher-level goals, include already decomposed goals and decision variants being evaluated. Thus, when decomposition process is finished, there are no goals left unclear. "Bottom-to-top" phase envisions the definition of all upper-level goals (super-goals, "ancestors") for each sub-goal or project (i.e., the goals this project influences).

As it has been mentioned, the experts build a hierarchy of goals that is represented by an oriented network-type graph (Fig. 8). Its nodes are marked by goal formulations. Presence of an edge, connecting one node (goal) to the other indicates the impact of one goal upon achievement of the other one. As a result of the above-mentioned process of goal hierarchy building, we get a graph that is unilaterally connected, because from each node there is a way to the node, marking the main goal. Each goal is assigned an indicator of achievement level, form 0 to 1. This

indicator equals 0 if there is absolutely no progress in achievement of the goal, while if the goal is completely achieved it equals 1. Impact of one goal upon the other can be positive or negative. Its degree is reflected by the respective value – a partial impact coefficient (PIC). In the method of target-oriented dynamic evaluation of alternatives (MTDEA) the delay of impact is also taken into consideration [13]. In case of projects their implementation time is taken into account as well. PIC are defined by experts, and, in order to improve the credibility of expert estimation process, pair-wise comparison-based methods are used.

Application of the approach and the methodology set forth in [10-11], calls for availability of a group of experts. Work of experts is rather costly and requires considerable time. So, reduction of expert information usage during building of DSS knowledge base (KB) in the process of IO detection represents a relevant issue.

The essence of the methodology of DSS KB building during IO detection [14-15] is as follows:

1) Group expert examination is conducted in order to define and decompose the goals of the informational operation. Thus, the IO is decomposed as a weakly structured system. For this purpose, the means of the system for distributed collection and processing of expert information (SDCPEI) are used.

2) Using the DSS, the respective KB is built, based on the results of the expert examinations conducted by SDCPEI as well as on available objective information.

3) Analysis of dynamics of thematic information flow is performed by means of content-monitoring system (CMS). PIC are enter into the KB of DSS.

4) Recommendations of the decision-maker are calculated by means of the DSS, based on the KB already built.

The methodology is illustrated by the example of Brexit.

The system “Consensus-2”, intended for evaluation of alternatives by a group of territorially distributed experts, was used as SDEICP for group decomposition. The system represents an upgraded version of “Consensus” system [16].

Based on interpretation of the data base, formed in SDCPEI “Consensus-2”, the knowledge engineer created the respective KB of “Solon-3” DSS [17]. The structure of goal hierarchy of this KB is provided on Fig. 8.

After that, by means of InfoStream CMS [18] the analysis of dynamics of thematic information flow was conducted. For this purpose, in accordance to each component of the IO, queries were formulated in

the specialized language. Based on these queries, dynamics analysis of publications on the target topic was performed. During formulation of queries in accordance to the goal hierarchy structure (Fig. 8), the following rules were used:

1) when moving from top to bottom, respective queries of lower-level components of the IO were supplemented by queries of higher-level components (using "&" symbol), for clarification;

2) in cases of abstract, non-specific character of IO components, movement from bottom to top took place, while the respective queries were supplemented by queries of lower-level components (using "|" symbol);

3) in cases of specific IO components, the query was made unique.

Based on the results of fulfillment of queries to InfoStream system, particularly, on the number of documents retrieved, respective PICs were calculated for each of them. PICs were calculated under assumption that the degree of impact of IO component was proportional to the quantity of the respective documents retrieved. Obtained PIC values were input into the KB. Thus, we managed to refrain from addressing the experts for evaluation of impact degrees of IO components.

The structure of goals hierarchy of KB for the Brexit is shown in Fig. 8. Based on the KB, the "Solon-3" DSS provided recommendations. The following rating has been obtained for the information impact of the publications: "Immigration and refugees" (0.364), "EU standards versus UK standards" (0.177), "EU laws versus UK laws" (0.143), "Expansion of foreign trade relations" (0.109), "Additional taxes" (0.084), "Sales quota" (0.08), "Unemployment" (0.043).

Use Case Diagram of a Concept of the Information-Analytical System for the Detection of Information Operations

Shown in Fig. 9 is the use case diagram of a concept of the information-analytical system for IO detection. The information-analytical system is an implementation of described above methods via system integration of DSS, CMS, SDCPEI and expert estimation system (EES). Our concept of the information-analytical system has following actors:

- **Expert** is a specialist who is invited or hired to provide a qualified opinion, judgment or estimation on some issues.
- **Knowledge engineer** is a specialist who constructs the knowledge base and provides recommendations with the help of the decision-making support toolkit.

- **DSS** is a system that provides recommendations based on the data from its knowledge base (both objective and expert data).
- **CMS** is a system that collects information from websites automatically in real time. In addition, CMS performs structuring, grouping by semantic features, thematic selecting, providing access to information databases in search mode, and analyzing the dynamics of thematic information streams.
- **SDCPEI** is a system for remote group work of experts in the global network.
- **EES** is a complex of software tools for expert estimation. EES must adapt flexibly to the level of experts and allow getting full and exact knowledge from them. "Level" ("Riven") system [19] is an example of EES.

Use Cases:

- **Building a knowledge base.** The knowledge engineer builds a knowledge base of DSS for modeling of the weakly structured domain and producing the recommendations. In this process, both objective and expert information can be used.
- **Group decomposition.** Expects take part in the group decomposition. According to the target of the IO, the knowledge engineer forms a group of experts who are specialist in the field. The process of group decomposition is to be carried by dialogue with the experts of the group to decompose (or to divide) the target to the subparts. At each stage of the decomposition, experts are asked to formulate a set of goals (they may choose from the existing ones or formulate their own). These goals must directly affect the target. Further, the decomposition of the next goals is performed similarly. The decomposition process comes up to the end when we obtain the specific components of the IO. These components must have precise formulations that we will use as the queries for the content monitoring system.
- **Group expert estimation.** The process can be initiated by the knowledge engineer after the group decomposition process. During this process, the knowledge engineer identifies the type of the sub-target influences (qualitative or quantitative, positive or negative), as well as their degrees.
- **Objective Information Entering to the Knowledge Base.** The knowledge engineer enters the objective information (content-monitoring results) to the knowledge base.
- **Analyzing of the dynamics of the thematic information stream.** The process is initiated by the knowledge engineer. During this

process, CMS performs the search queries. The found publications are used to study the dynamics of number of publications on target topics.

- **Calculation of recommendations.** The process is aimed to calculate the relative efficiency of each component of the IO (information throw-in). The relative efficiency means the contribution of the component to achieving the target. The calculation is based on data from the knowledge base. In addition, we evaluate the aggregated estimation of the effectiveness of the components. Dynamics of the process is taking into account. Based on the past data is possible to change the quantitative values of the target.

The concept integrates all components presented above.

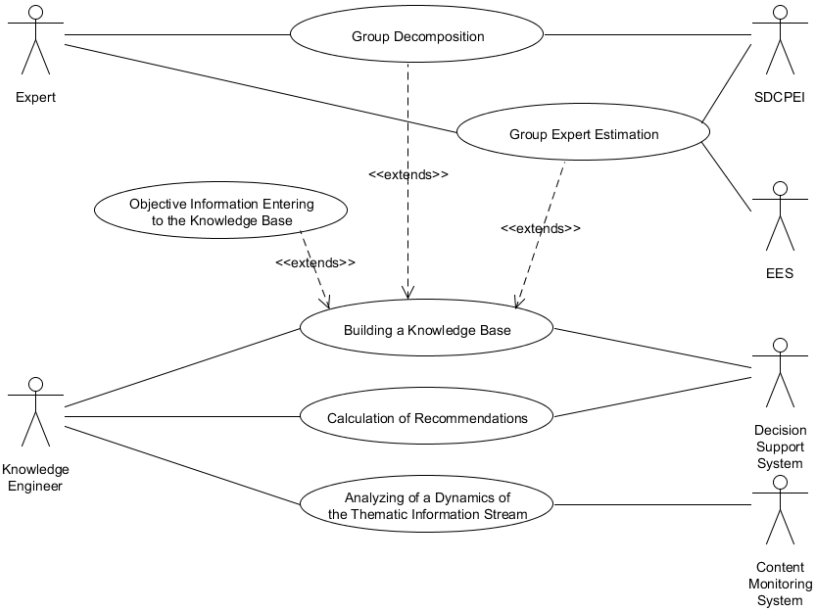


Fig. 9. Use Case Diagram of the Concept of Information-analytical System for Information Operations Detection

Conclusions

In order to detect IO, it is necessary to involve a complex of analytical techniques. The complex includes methods of nonlinear dynamics and computer linguistics.

The wavelet transform and its modifications are applied to detect periodicity, peaks, spikes, and behavioral patterns in the dynamics of the number of thematic publications.

To extract meaningful and significant words from the text of publications we use the TFIDF score, the standard deviation estimate, and horizontal visibility graphs.

Applying decision-making support methods, we provide the decomposition of the topics of the IO and assess rating of the effectiveness of these topics.

It is proposed a concept of a new information-analytical system for detection of IO through integration of a decision support system, a content-monitoring system, a system for distributed expert information collection and processing and an expert estimation system.

This paper is prepared as part of project #F73/23558 “Development of Decision-making Support Methods and Means for Detection of Information Operations”. The project won the contest #F73 for grant support of scientific research projects held by The State Fund for Fundamental Research of Ukraine and Belarusian Republican Foundation for Fundamental Research.

References

1. Lewandowsky S., Ecker U. K. H., Seifert C. M., Schwarz N., Cook J. Misinformation and Its Correction Continued Influence and Successful Debiasing, Psychological Science in the Public Interest, Department of Human Development, Cornell University, USA, 2012, Volum e 13 issue 3, pp. 106-131.

2. Berinsky A. J. Rumors and Health Care Reform: Experiments in Political Misinformation, British Journal of Political Science, Volume 47, Issue 2 April 2017, pp. 241-262.

3. Горбулін В.П., Додонов О.Г., Ланде Д.В. Інформаційні операції та безпека суспільства: загрози, протидія, моделювання: монографія – К., Інтертехнологія, 2009 – 164 с.

4. Ландэ Д.В., Додонов В.А., Коваленко Т.В. Информационные операции в компьютерных сетях: моделирование, выявление, анализ // МОДЕЛИРОВАНИЕ-2016: материалы пятой Международной конференции МОДЕЛИРОВАНИЕ-2016, Киев, 25-27 мая 2016 г. / ИПМЭ НАН Украины, 2016. - С. 198-201.

5. Veit Bachmann, James D. Sidaway Brexit geopolitics / Geoforum 77 (2016) pp. 47–50.

6. Addison Paul S. The illustrated wavelet transform handbook: introductory theory and applications in science, engineering, medicine

and finance. Boca Raton, FL: CRC Press, Taylor & Francis Group, 2016. 446 p.

7. Peng C.-K., Buldyrev S. V., Havlin S. et al. Mosaic organization of DNA nucleotides. Physical Review E, 1994. V. 49 E. 2. 1685-1689 pp.

8. Ortuño M., Carpena P., Bernaola P., Muñoz E., Somoza A.M. Keyword detection in natural languages and DNA // Europhys. Lett, – 57(5). – P. 759-764 (2002).

9. Lande D., Snarskii A. Compactified Horizontal Visibility Graph for the Language Network. arXiv preprint arXiv: 1302.4619, 2013. Andriichuk O.V., Kachanov P.T. Usage of expert decision-making support systems in information operations detection / Proceedings of the International Symposium for the Open Semantic Technologies for Intelligent Systems (OSTIS 2017), Minsk, Republic of Belarus, 16-18 Ferbruary, 2017, pp. 359-364.

10. Андрейчук О.В., Качанов П.Т. Методика применения инструментария экспертной поддержки принятия решений при идентификации информационных операций / Информационные технологии и безопасность: Материалы международной научно-технической конференции 1 декабря 2016 года – К: ИПРИ НАН Украины, 2016. – С. 141-155. // Oleh V. Andriichuk, Petro T. Kachanov A Methodology for Application of Expert Data-based Decision Support Tools while Identifying Informational Operations / CEUR Workshop Proceedings (ceur-ws.org). Vol-1813 urn:nbn:de:0074-1813-0. Selected Papers of the XVI International Scientific and Practical Conference "Information Technologies and Security" (ITS 2016) Kyiv, Ukraine, December 1, 2016. - pp. 40-47. [http://ceur-ws.org/Vol-1813/paper6.pdf]

11. Каденко С.В. Можливості та перспективи використання експертних технологій підтримки прийняття рішень у сфері інформаційної безпеки / Информационные технологии и безопасность: Материалы международной научно-технической конференции 1 декабря 2016 года – К: ИПРИ НАН Украины, 2016. – С. 31-42. // Sergii V. Kadenko Prospects and Potential of Expert Decision-making Support Techniques Implementation in Information Security Area / CEUR Workshop Proceedings (ceur-ws.org). Vol-1813 urn:nbn:de:0074-1813-0. Selected Papers of the XVI International Scientific and Practical Conference "Information Technologies and Security" (ITS 2016) Kyiv, Ukraine, December 1, 2016. - pp. 8-14. [http://ceur-ws.org/Vol-1813/paper2.pdf]

12. Andriichuk O.V., Kachanov P.T. Usage of expert decision-making support systems in information operations detection / Proceedings of the International Symposium for the Open Semantic Technologies for

Intelligent Systems (OSTIS 2017), Minsk, Republic of Belarus, 16-18 Ferbruary, 2017, pp. 359-364.

13. Тоценко В.Г. Методы и системы поддержки принятия решений. Алгоритмический аспект. – К.: Наукова думка, 2002. – 382 с.

14. Андрійчук О.В., Ланде Д.В. Побудова баз знань систем підтримки прийняття рішень при виявленні інформаційних операцій / Реєстрація, зберігання і обробка даних: зб. Наук. Праць за матеріалами Щорічної підсумкової наукової конференції 17-18 травня 2017 року / НАН України. Інститут проблем реєстрації інформації. – К.: ІПРІ НАН України, 2017. – С. 101-103.

15. Андрійчук О.В., Ланде Д.В. Застосування інструментарію підтримки прийняття рішень при виявленні інформаційних операцій / Актуальні проблеми управління інформаційною безпекою держави : зб. матеріалів наук.-практ. конф., (Київ, 24 трав. 2017 р.). – Київ : Нац. акад. СБУ, 2017. – С.161-163.

16. Цыганок В.В., Качанов П.Т., Андрійчук О.В., Каденко С.В. Свідоцтво про реєстрацію авторського права на твір № 45894 Державної служби інтелектуальної власності України. Комп'ютерна програма “Система розподіленого збору та обробки експертної інформації для систем підтримки прийняття рішень - «Консенсус»” від 03.10.2012.

17. Тоценко В.Г., Качанов П.Т., Циганок В.В. Свідоцтво про державну реєстрацію авторського права на твір №8669. Міністерство освіти і науки України державний департамент інтелектуальної власності. Комп’ютерна програма "Система підтримки прийняття рішень СОЛОН-3" (СППР СОЛОН-3) від 31.10.2003.

18. Григорьев А.Н., Ландэ Д.В., Бороденков С.А., Мазуркевич Р.В., Пацьора В.Н. InfoStream. Мониторинг новостей из Интернет: технология, система, сервис: Научно-методическое пособие – К.: Старт-98, 2007. – 40 с.

19. Циганок В.В., Андрійчук О.В., Качанов П.Т., Каденко С.В. Свідоцтво про реєстрацію авторського права на твір № 44521 Державної служби інтелектуальної власності України. Комп'ютерна програма “Комплекс програмних засобів для експертного оцінювання шляхом парних порівнянь «Рівень»” від 03.07.2012.

ПРИМЕНЕНИЕ ГРАФОВ ГОРИЗОНТАЛЬНОЙ ВИДИМОСТИ В ИНФОРМАЦИОННОЙ АНАЛИТИКЕ

Д.В. Ландэ[1,2], д.т.н., с.н.с. (dwlande@gmail.com),
А.А. Снарский[2,1], д.ф.-м.н., профессор (asnarskii@gmail.com)
[1]Институт проблем регистрации информации НАН Украины, Киев, Украина
[2]НТУУ «Киевский политехнический институт им. Игоря Сикорского», Киев, Украина

Описывается применение метода графов горизонтальной видимости (Horizontal Viability Graph – HVG) в информационной аналитике, при формировании модели предметной области, соответствующей специальному запросу к поисковой системе, а также при построении сети связей информационных источников, отражающих общую тему.

Введение

В настоящее время в естественных науках и технологиях получили широкое распространение методы исследования временных рядов, в основе которого лежит их преобразование в граф (сложную сеть). При таком отображении объединяются две развитые области исследований – нелинейные методы анализа временных рядов и методы теории сложных сетей. Появляется возможность применить развитые методы анализа сложных сетей [1, 2] к анализу временных рядов. Предложено несколько методов построения сетей на основе временных рядов [3, 4], в частности, так называемый граф горизонтальной видимости (Horizontal Visibility Graph – HVG) [5, 6].

Алгоритм построения графов видимости по временному ряду проиллюстрирован на Рис. 1. При построении графа горизонтальной видимости на горизонтальной оси (ось времени) отмечаются точки, соответствующие индексу моменту времени t_i, от которых в перпендикулярном направлении строятся отрезки высотой, равной значениям ряда в этих точках – $x(t_i)$. Узлами графа горизонтальной видимости являются внешние вершины построенных отрезков. Связь между вершинами в HVG считается существующей, если горизонтальная прямая, проведенная из одной из вершин пересекает отрезок другой вершины, не пересекая ни одного из построенных отрезков, находящихся между ними. Этот геометрический критерий можно записать следующим образом: два узла (элемента ряда),

например, t_m и t_n соединены связью, если (см. рис. 2) , $x(t_m)$, $x(t_n) > x(t_p)$ для всех p: $n < p < m$.

Работа посвящена применению концепции Horizontal Viability Graph к двум областям информационной аналитики – построению моделей предметных областей из ключевых слов, отражающих наиболее важные понятия в тематических информационных потоках, а также построению сети связей источников информации.

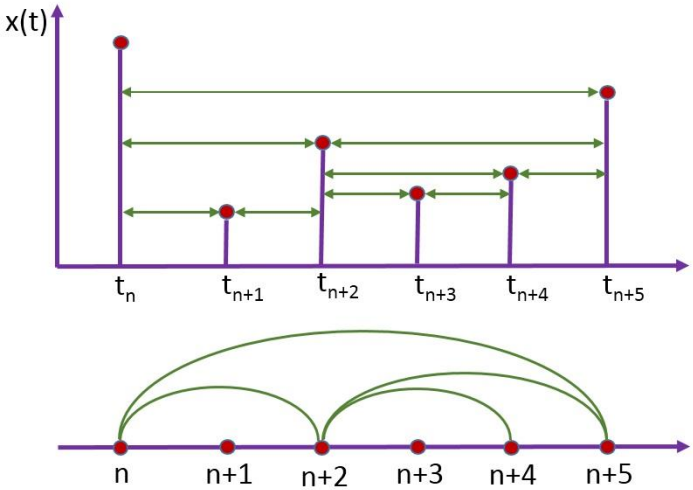


Рисунок 1 − Пример построения графа горизонтальной видимости

Сеть из ключевых слов – модель предметной области

Известно несколько подходов к построению так называемых сетей слов (Language Network) из текстов и способов интерпретации узлов и связей. Например, узлы могут быть соединены между собой, если соответствующие им слова стоят рядом в тексте, принадлежат одному предложению или абзацу, соединены синтаксически или семантически [7, 8]. В [9] был предложен метод построения сети слов путем анализа текстовых корпусов с применением графов видимости.

Поисковая система выдает по тематическому запросу документы, в каждом из которых выделены M ключевых слов, ранжированных в порядке важности, причем ранг определяется весовым значением ключевого слова, например, его относительная частота в документе (слова из стоп-словаря, естественно, не учитываются), TFIDF (TF –

частота слова в документе, IDF – величина, обратная количеству документов из массива, в котором встретился данный термин) и его модификации. Рассмотрим последовательность весовых значений ключевых слов из документов по некоторой узкой тематике, определяемых, например, запросом, следующих друг за другом в порядке их появления в информационном потоке (времени публикации). На Рис. 2 представлена цепочка документов и характеризующих их ключевых слов и их «видимость справа налево». Предполагается, что первый по времени документ в информационном потоке – D_1, за ним следуют D_2, D_3 и т.д. В пределах одного документа ввиду ранжированности по весу соседние ключевые слова связываются друг с другом (справа налево), и лишь лидирующие по весу ключевые слова могут связываться с ключевыми словами других документов (не обязательно с самими большими весовыми значениями).

Следует отметить, что одно и то же ключевое слово может присутствовать в разных документах и иметь в них различные весовые значения. При этом одно и то же ключевое слово в каждом документе может «видеть» различные ключевые слова, т.е. выходная степень его как узла графа горизонтальной видимости может превышать 1. Связь между некоторыми лидирующими по весу ключевыми словами из различных документов может ассоциироваться с «клубом богатых» [10, 11], феномену, присущему большинству сетей языка [12]. Традиционно связи в сетях языка строятся путем соединения тех узлов (ключевых слов), которые соответствуют одному фрагменту текста (документу, абзацу или предложению), т.е. в рассматриваемом примере степень каждого из узлов не может быть меньше M.

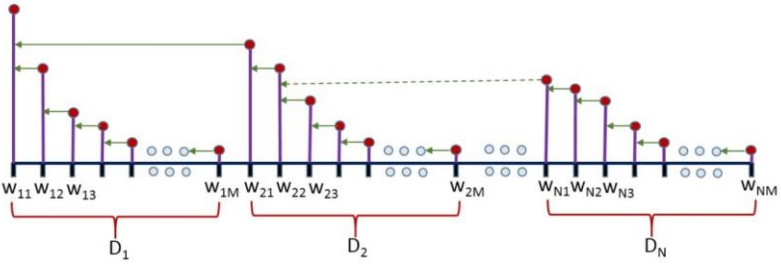


Рисунок 2 − Горизонтальная видимость ключевых слов w_{ij} из документов D_i

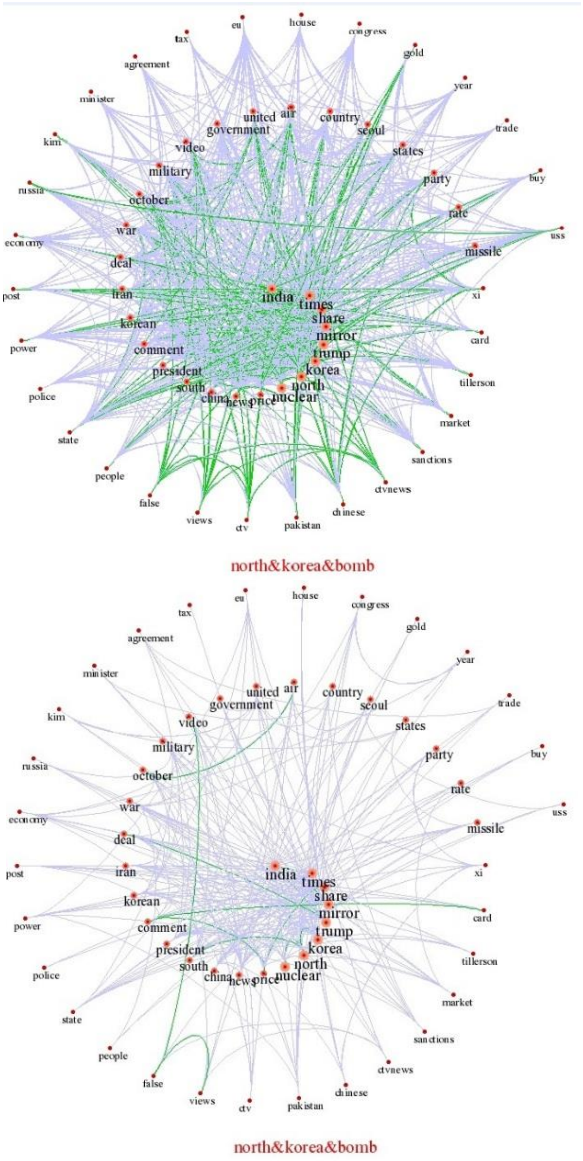
Оценить качество сети, построенной в соответствии с концепцией HVG могут лишь пользователи – аналитики информационной системы, но данный метод обеспечивает значительное сокращение количества связей между узлами (Рис. 3), оставляя наиболее значимые (между ближайшими по весу, что важно для ассортативных сетей [13, 14], в которых наряду с «феноменом клуба богатых» существенны связи между узлами, имеющими близкие степени). На рис. 3 представлены две сети связи терминов, первая – традиционная, а вторая построенная с применение предложенного метода. Как можно видеть, второй подход обеспечивает создание более наглядных сетевых структур.

Сеть взаимосвязи источников информации

При анализе тематических информационных потоков из различных источников, формируемых системами контент-мониторинга, существует проблема построения сети связей источников информации. При этом основанием для связи между двумя источникам может служить тот факт, что они часто публикуют совпадающие или близкие по теме документы. Построение такой сети позволяет определять, какие из информационных источников по данной тематике являются основными, наиболее влиятельными, какие подвержены определенным информационным влияниям.

Предлагаемый подход базируется на предположении, что источники информации заранее ранжированы по объемам публикаций исходя из опыта наблюдения за ними в течение продолжительного времени, т.е. им уже приписаны некоторые весовые значения.

Предполагается построение временного ряда, элементы которого будут соответствовать документам из тематического информационного потока в порядке их появления в системе контент-мониторинга сетевых ресурсов (публикации на веб-ресурсах, в социальных сетях). Значения ряда будут соответствовать весовому значению информационного источника, опубликовавшему соответствующий документ. Предполагается, что весовое значение, степень важности источника определены заранее. По данному ряду, как и в предыдущем случае строится граф горизонтальной видимости, причем «направление взгляда» как и в предыдущем примере направлено в прошлое.



a)

b)

Рисунок 3 − Сети языка из ключевых слов: a) традиционная; b) Horizontal Visibility Graph

Таким образом устанавливается связь источника информации с другим, более рейтинговым, если он существует и опубликовал информацию раньше. Построенная таким образом сеть (Рис. 5) отражает связи источников по заданной тематике, позволяет определять лидеров среди них, делать предположения относительно первоисточников информации. Например, на Рис. 5 показана сеть связей международных источников информации, отражающих тематику ядерных испытаний в Северной Корее в октябре 2017 г. по данным системы контент-мониторинга.

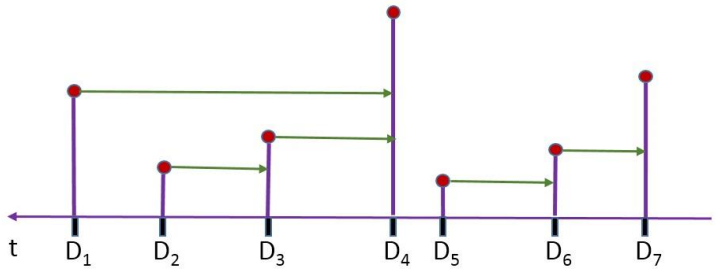


Рисунок 4 − Горизонтальная видимость документов D_i тематического информационного потока

Выводы

Предложены реализации концепции Horizontal Viability Graph при анализе тематических информационных потоков, которые, с одной стороны, расширяют сферу применения данного подхода, а, с другой стороны, позволяют решать задачи формирования и наглядной визуализации моделей предметных областей как сети терминов, соответствующих основным понятиям и связям между ними и сетей взаимосвязей источников информации.

Публикация содержит результаты исследований, проведенных при грантовой поддержке Государственного фонда фундаментальных исследований Украины по конкурсному проекту Ф73

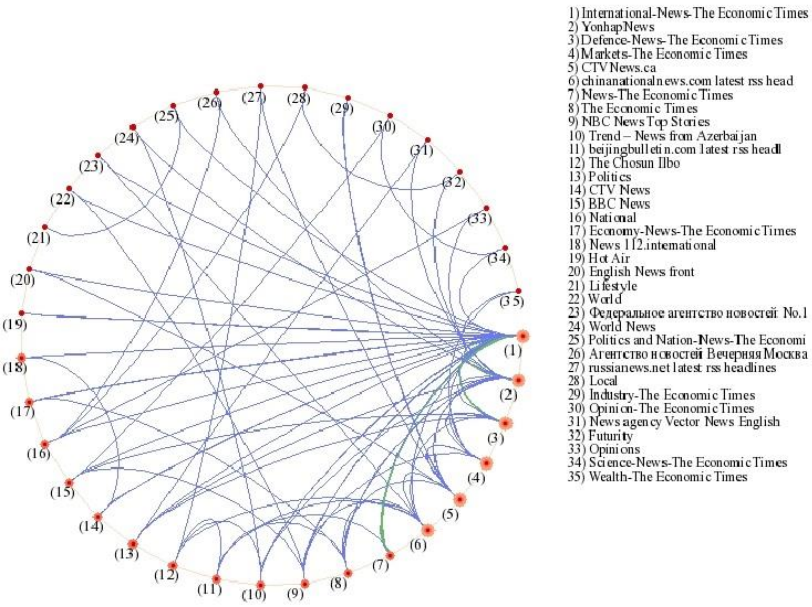


Рисунок 5 − Пример графа горизонтальной видимости, отражающего связи источников информации по заданной теме

.

Литература

1. Albert R., Barabási A.-L. Statistical mechanics of complex networks // Rev. Mod. Phys, 2002. – 74. – pp. 47-97.
2. Newman M.E.J. The structure and function of complex networks // SIAM Rev., 2003. – 45. – pp. 167-256.
3. Nunez A. M., Lacasa L., Gomez J. P., Luque B. Visibility algorithms: A short review // New Frontiers in Graph Theory, Y. G. Zhang, Ed. Intech Press, ch. 6. – pp. 119-152 (2012).
4. Bezsudnov I.V., Snarskii A.A. From the time series to the complex networks: The parametric natural visibility graph // Physica A: Statistical Mechanics and its Applications, 2014, 414, 53-60.
5. Luque B., Lacasa L., Ballesteros F., Luque J. Horizontal visibility graphs: Exact results for random time series // Physical Review E,– pp. 046103-1–046103- 11 (2009).
6. Gutin G., Mansour T., Severini S. A characterization of horizontal visibility graphs and combinatoris on words // Physica A, – 390 – pp. 2421-2428 (2011).

7. Ferrer-i-Cancho R., Sole R. V. The small world of human language // Proc. R. Soc. Lond. – B 268, 2261 (2001).
8. Caldeira S. M. G., Petit Lobao T. C., Andrade R. F. S., Neme A., Miranda J. G. V. The network of concepts in written texts // Preprint physics/0508066 (2005). [6] Ferrer-i-Cancho R., Sole R.V., Kohler R. Patterns in syntactic dependency networks // Phys. Rev. E 69, 051915 (2004).
9. D. V. Lande, A. A. Snarskii, E. V. Yagunova, E. V. Pronoza The Use of Horizontal Visibility Graphs to Identify the Words that Define the Informational Structure of a Text // 12th Mexican International Conference on Artificial Intelligence, 2013. – pp. 209-215.
10. *Colizza, V. and Flammini, A. and Serrano, M. A. and Vespignani, A. Detecting rich-club ordering in complex networks // Nature Physics. 2, 2006. – **2**: 110–115.*
11. Julian J. McAuleya, Luciano da Fontoura, Tibério S. Caetan. Program Rich-club phenomenon across complex network hierarchies // Applied physics letters, 2007, -V 91 (8), 084103
12. Heuvel M.P., Sporns O. Rich-Club Organization of the Human Connectome // Journal of Neuroscience, 2011. – 31 (44). – pp. 15775-15786.
13. M.E.J. Newman. Assortative mixing in networks // Phys. Rev. Lett. **89**, 208701 (2002).
14. Piraveenan, M., Prokopenko M., and A. Y. Zomaya. Local assortativeness in scale-free networks // EPL (Europhysics Letters) 84.2, 28002 (2008).

ОЦІНЮВАННЯ ЗАХИЩЕНОСТІ ІНФОРМАЦІЇ В КОМП'ЮТЕРНИХ СИСТЕМАХ ЗА СОЦІОІНЖЕНЕРНИМ ПІДХОДОМ

[1]Володимир Мохор, v.mokhor@gmail.com, [1]Оксана Цуркан, otsurkan24@gmail.com, [2]Василь Цуркан, v.v.tsurkan@gmail.com, [1]Ростислав Герасимов, gerasimov.rostislav@gmail.com

[1]ІПМЕ ім. Г.Є Пухова НАН України, Київ, Україна,

[2]ІСЗЗІ КПІ ім. Ігоря Сікорського, Київ, Україна

Захист інформації в комп'ютерних системах орієнтований на збереження її властивостей конфіденціності, цілісності та доступності від різноманітних за своєю сутністю несприятливих впливів. Потенційно можливий несприятливий вплив тлумачиться загрозою. Для запобігання або ускладнення можливості реалізацій загроз, зменшення потенційних збитків створюється та підтримується у дієздатному стані система заходів захисту інформації в комп'ютерних системах. Так.а система включає обчислювальну систему, фізичне середовище, персонал та інформацію. На збереження її властивостей у комп'ютерних системах суттєво впливає врахування нетехнічного аспекту, зокрема, персоналу (наприклад, керівника, адміністратора, користувача). З огляду на це, для оцінювання технічної захищеності інформації пропонується соціоінженерний підхід. У рамках такого підходу вразливості персоналу тлумачяться як його слабкості, потреби, манії (пристрасті), захоплення. Маніпулювання ними дозволяє отримати несанкціонований доступ до інформації без руйнування та перекручування головних для нього системоутворюючих якостей (цілісність, розвиток). Як наслідок, це призводить до нової моделі поведінки персоналу, створення сприятливих умов реалізації загроз безпеці інформації і, як наслідок, зменшенню здатності системи захисту інформації протидіяти їх впливові. Це відображається в таких формах як шахрайство, обман, афера, інтрига, містифікація, провокація. Використанню кожної з означених форм маніпулювання передує визначення її змісту шляхом ретельного планування, організування та контролювання. Означені дії є основою методів соціальної інженерії. З одного боку, вони реалізуються засобами сучасних

телекомунікацій. Тоді як з іншого, передбачається встановлення особистого контакту з персоналом. Таким чином, шляхом використання методів соціальної інженерії можливе виявлення, нейтралізування, запобігання появі уразливостей інформації в комп'ютерних системах. Цим підвищується її захищеність з урахуванням нетехнічного аспекту.

Постановка проблеми

Інформаційні ресурси окремих організацій і фізичних осіб являють собою певну цінність, мають відповідне матеріальне вираження і вимагають захисту від різноманітних за своєю сутністю несприятливих впливів. Потенційно можливий несприятливий вплив тлумачиться загрозою. Тому захист інформації в комп'ютерних системах полягає в створенні та підтриманні в дієздатному стані системи заходів для запобігання або ускладнення можливості реалізацій загроз, а також зменшення потенційних збитків. Оскільки комп'ютерна система включає обчислювальну систему, фізичне середовище, персонал і оброблювану інформацію, то на оцінювання технічної захищеності інформації суттєво впливає врахування нетехнічного аспекту, зокрема, персоналу (наприклад, див. рис. 1, керівництва, адміністратора операційної системи, користувачів). Тому для цього пропонується соціоінженерний підхід [1-8].

Сутність соціоінжереного підходу

У рамках соціоінженерного підходу вразливості персоналу тлумачяться як його слабкості, потреби, манії (пристрасті), захоплення. Маніпулювання ними дозволяє отримати несанкціонований доступ до інформації без руйнування та перекручування головних для нього системоутворюючих якостей (цілісність, розвиток). Як наслідок, це призводить до нової моделі поведінки персоналу, створення сприятливих умов реалізації загроз безпеці інформації і, як наслідок, зменшенню здатності системи захисту інформації протидіяти їх впливові (див. рис. 2).

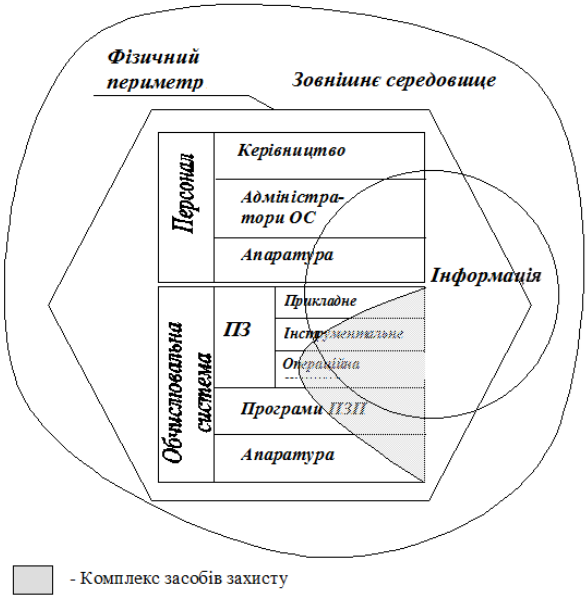


Рисунок 1 – Елементи комп'ютерної системи [3]

Це відображається в таких формах як, наприклад [9-13], шахрайство, обман, афера, інтрига, містифікація, провокація. Використанню кожної з означених форм маніпулювання передує визначення її змісту шляхом ретельного планування, організування та контролювання.

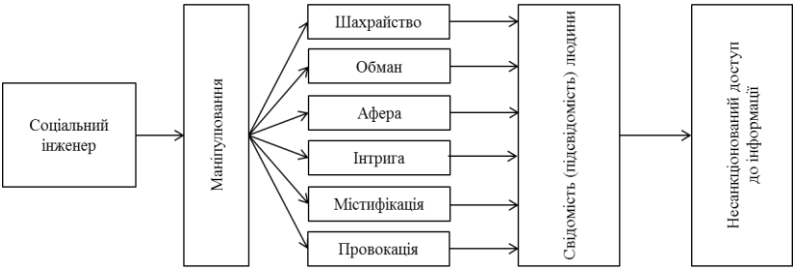


Рисунок 2 – Використання соціоінженерного підходу

З огляду на рис. 2, використання соціоінженерного підходу до оцінювання захищеності інформації в комп'ютерних системах передбачає цілеспрямований вплив на свідомість (підсвідомість) персоналу проти волі, але за його згодою. Такий вплив дозволяє управляти поведінкою керівництва, адміністратора, користувачів через слабкості, інтереси, потреби, схильності, переконання, звички, психічний та емоційний стан. Тому маніпулювання цими уразливостями і виражається в таких формах як шахрайство, обман, афера, інтрига, містифікація, провокація. Разом з тим, використанню кожної з означених форм маніпулювання передує визначення їх сутності шляхом ретельних планування, організації та контролювання.

У рамках соціоінженерного підходу використання атак соціальної інженерії орієнтоване на отримання "несанкціонового" доступу до інформації при оцінюванні її захищеності шляхом "негативного" інформаційно-психологічного впливу на свідомість або підсвідомість персоналу (див., наприклад [14], рис. 3).

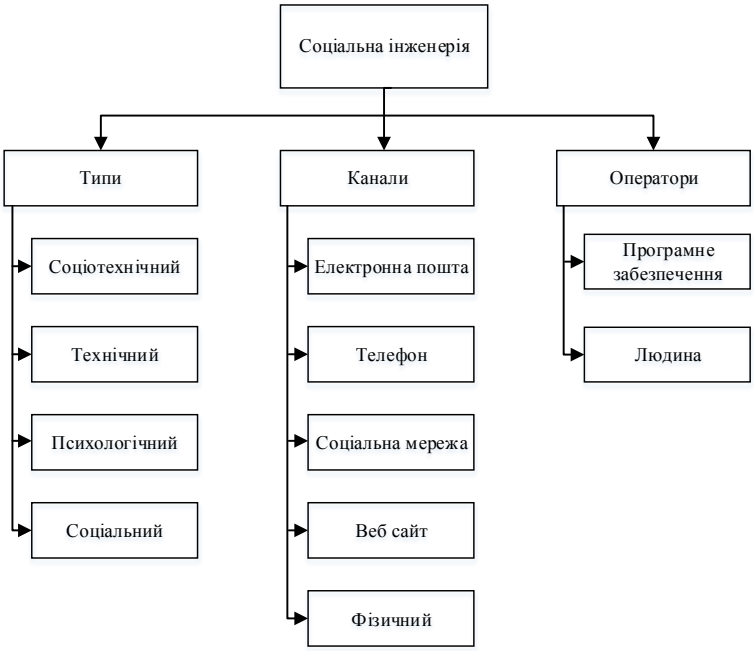


Рисунок 3 – Класифікування ознак реалізування атак соціальної інженерії [14]

Форми маніпулювання персоналом

Форми маніпулювання персоналом при оцінюванні захищеності інформації в комп'ютерних системах змінюються залежно від різновиду атак соціальної інженерії, а саме [15-20]:

1. Фішінг (Phishing) – масове розсилання електронної пошти великій групі адресатів. Ознайомлення з електронними листами спонукає їх до, наприклад, відкриття вкладення до листа, переходу за посиланням на веб-сторінку. Його метою є виманювання у довірливого або неуважного персоналу комп'ютерної системи персональних даних.

2. Фармінг (Pharming) – перенаправлення користувачів на шахрайські сайти для отримання їх логіну та паролю. Це досягається завдяки розповсюдженню електронної пошти серед користувачів, наприклад, соціальних мереж, онлайн-банкінгу, поштових веб-сервісів.

3. Прітекстінг (Pretextй) – отримання інформації або спонукання до вчинення певних дій обманом на основі заздалегідь складеного сценарію або створення фіктивної ситуації. Застосовується через телефон та потребує проведення попередніх досліджень для входження в довіру.

4. Смішінг (Smishing) **–** отримання інформації шляхом масового розсилання SMS повідомлень з посиланням на веб-ресурси або з реквізитами організацій (наприклад, фінансових). Внаслідок цього здійснюються відповідні дії, наприклад, дзвінок до банку для перевірки стану рахунку з зазначенням конфіденційних даних: номеру картки, терміну дії.

5. Вішінг (Vishing) **–** отримання інформації шляхом входження в довіру під час розмови через ір-телефон. При цьому в порушення конфіденційності здійснюється завдяки викладенню прохання у повідомленні зателефонувати на певний міський номер. Наприклад, вести номер карти, паролі, PIN-коди, коди доступу або іншу інформацію.

6. Спір фішінг (Spear Phishing) **–** надсилання листа електронної пошти конкретному адресату (наприклад, керівнику, адміністраторові, користувачеві), що спонукає його до обов'язкового перегляду та відповіді на отриманий лист.

7. Вейлінг (Whaling) **–** надсилання листа електронної пошти представнику керівництва організації, що спонукає його до обов'язкового перегляду та відповіді на отриманий лист.

Так (див. табл.), отримання несанкціонованого доступу до інформації за допомогою фішингу, фармінгу, смішінгу, вішінгу, спір

фішінгу, вейлінгу здійснюється шляхом використання таких форм маніпулятивного впливу як шахрайство та обман. Тоді як основою для створення фіктивних ситуацій при прітекстінгу є афера, інтрига, містифікація та провокація.

Таблиця

Форми маніпулятивного впливу при соціоінженерних атаках

№ п/п	Різновид соціоінженерної атаки	Форми маніпулятивного впливу					
		Шахрайство	Обман	Афера	Інтрига	Містифікація	Провокація
1.	Фішінг	+	+	–	–	–	–
2.	Фармінг	+	+	–	–	–	–
3.	Прітекстінг	+	+	+	+	+	+
4.	Смішінг	+	+	–	–	–	–
5.	Вішінг	+	+	–	–	–	–
6.	Спір фішінг	+	+	–	–	–	–
7.	Вейлінг	+	+	–	–	–	–

Тому при оцінюванні захищеності інформації в комп'ютерних системах за соціоінженерним підходом доцільно враховувати форми маніпулятивного впливу.

Методи соціальної інженерії

Оцінювання захищеності інформації орієнтоване на отримання відомостей про комп'ютерні системи – етап соціальної інженерії. Цей етап включено в аудит отримання і аналізування інформації зі зовнішнього середовища. Враховуючи високу ймовірність впливу людського фактору на захищеність інформації, на даному етапі вдало використовуються методи соціальної інженерії. Для успішного виконання запланованих дій соціальний інженер проводить роботу узгоджено з мережевим адміністратором. У його конструктивні дії входять: вивчення і аналізування змістовного боку комп'ютерної системи, пошук уразливостей, систематизування отриманих відомостей, розроблення схеми дій.

Використання методів соціальної інженерії для імітування дій порушника, що направлені на користувачів комп'ютерних систем організації, дозволяє оцінити рівень кваліфікації користувачів в

області забезпечення безпеки інформації і ймовірність реалізування атак соціальної інженерії.

Вхідною інформацією для соціального інженера при спробі отримання інформації про комп'ютерні системи може бути контактна інформація, що отримується з публічних джерел. Наприклад: прізвища, імена, посади користувачів. За отриманою вхідною інформацію вибираються, виокремлюються вірогідні уразливості в комп'ютерних системах, через які можливе реалізування загроз соціальної інженерії.

Основою використання методів соціальної інженерії є [9-11]:

- особливості, що керують людською свідомістю;
- аудиторія або поле діяльності;
- некомпетентність аудиторії у визначених термінах і предметних областях у сфері інформаційної безпеки;
- нестійкість психологічних властивостей особистості, що характеризуються поведінковими стереотипами. Їх можна використовувати для маніпулювання через основні потреби, слабкості, бажання, ідеали.

Більшість соціальних інженерів діє за ідентичними або близькими шаблонами. Тому вивчення прийомів їх «роботи» дозволяє виокремити такі рівні взаємодії з об'єктом впливу як домінування, маніпулювання, суперництво, партнерство.

Всі методи соціальної інженерії можна поділити на дві групи [9-11, 14-20]:

1. **Віддалена соціальна інженерія** реалізується засобами сучасних телекомунікацій шляхом використання:

1.1. «Телефону»

Завдяки телефонії, соціальний інженер може залишатися анонімним і в той же час мати прямий зв'язок з об'єктом впливу. Останнє важливо тому, що безпосередній контакт не дає співрозмовнику часу обміркувати поведінку у вірогідних ситуаціях, зважати на всі за та проти. Вирішувати необхідно швидко, до того ж під тиском соціального інженера. Оскільки під час телефонної розмови відбувається обмін тільки звуковою інформацією, то велику роль у прийнятті рішень відіграє анотація і голос співрозмовника. Дані характеристики підбираються у відповідності з моделлю поведінки соціального інженера для отримання інформації про об'єкт впливу, наприклад:

а) **начальник** – людина, яка звикла віддавати команди, цінує свій час, досягає поставленої мети. Манера розмови жорстка, нетерпляча. Повна впевненість у собі і легка (або повна) зверхність

до рядового персоналу. Своїм тоном показує, що проблема, з якою звернувся – дрібниця, яку необхідно вирішити якомога швидше. Ніяких прохань – тільки вимоги і вказівки. У відповідь на недовірливі або перевіряючі репліки – допустиме незадоволення і залякування співрозмовника;

b) **секретар** – дівчина (здебільшого) з приємним голосом. Завдання – виконати конкретне доручення начальника, не відволікаючись на умовності. Вона володіє інформацією про начальника, його справи, у своїй мові користується достовірними або недостовірними фактами, які складно перевірити. Характер розмови – м'який, з легким фліртуванням (якщо співрозмовник – чоловік). Реакція на небажання співпрацювати – бурхливе розчарування, скарга, що скаже начальство;

c) **технічний співробітник** – працівник організації, який характеризується поблажливим, але дружелюбним відношенням до клієнтів. Мета – усунути несправність. Супроводжується використанням специфічних термінів для відображення своєї компетентності. На відмову співпрацювати – реакція здивування, оскільки співпраця у першу чергу вигідна для клієнта. Жодних вмовлять –йому дається зрозуміти, що без його участі проблема тільки ускладнюється. Допустиме залякування тяжкими наслідками.

d) **користувач** – працівник, що виконує свої обов'язки і наляканий виникненням неочікуваної проблеми. Чітко виражений мотив швидкого вирішення усіх проблем і повернення до своєї рутинної роботи. Відсутність уявлення про характер проблеми, зацікавленість тільки в її вирішенні. Характер спілкування – показати безнадійність свого положення і готовність віддатися у руки спеціалісту.

1.2. Глобальної мережі Інтернет

Найбільш розповсюдженими способами реалізації методів соціальної інженерії за допомогою глобальної мережі Інтернет є:

a) проведення соціальної інженерії шляхом електронного листування;

b) проведення соціальної інженерії через системи обміну повідомленнями (Skype, Viber);

c) соціальна інженерія на форумах, чатах, блогах.

У даних випадках вдале реалізування соціальної інженерії обумовлене правильністю розроблення сценарію спілкування.

2. Особистий контакт. Найбільш складний і небезпечний метод соціальної інженерії. Крім перерахованих вимог до сценарію спілкування і моделі поведінки, соціальний інженер повинен

приділяти увагу своїй зовнішності і манерам «живого» спілкування. Для правильного візуального сприйняття, необхідно правильно підібрати:

a) колір одягу та взуття;
b) манери та жести при спілкуванні;
c) положення в просторі відносно співрозмовника.

Також при використанні методів соціальної інженерії необхідно характеризувати співрозмовника. За голосом або за зовнішністю, доцільно визначити яку його слабкість доцільно використовувати для досягнення поставленої мети. До основних слабкостей людини, використання яких разом з правильно підібраною поведінкою і сценарієм розмови дозволяє досягнути очікуваного результату, належать, наприклад:

a) довірливість;
b) страх;
c) жадібність;
d) відкритість;
e) зверхність;
f) милосердя.

Основними причинами впливу на об'єкт соціальної інженерії є, наприклад:

a) відчуття достоїнства;
b) прагнення до успіху;
c) матеріальна вигода.

Використання прийомів прихованого і прямого маніпулювання персоналом дають можливість соціальному інженеру дізнаватися і в подальшому використовувати інформацію для оцінювання її захищеності в комп'ютерній системі.

Висновок

Таким чином, оцінювання захищеності інформації в комп'ютерних системах за соціоінженерних підходом дозволяє запобігти, виявити, врахувати або усунути уразливості, що пов'язані або обумовлені діяльністю персоналу. У рамках соціоінженерного підходу його вразливості тлумачяться як слабкості, потреби, манії, захоплення. Маніпулювання ними призводить до нової моделі поведінки персоналу. Це відображається в таких формах як, наприклад, шахрайство, обман, афера, інтрига, містифікація, провокація. Використанню кожної з означених форм маніпулювання передує визначення її змісту шляхом ретельного планування, організування та контролювання. При цьому форми маніпулювання персоналом при оцінюванні захищеності інформації в комп'ютерних системах

змінюються залежно від різновиду атак соціальної інженерії. Так, отримання несанкціонованого доступу до інформації за допомогою фішингу, фармінгу, смішінгу, вішінгу, спір фішінгу, вейлінгу здійснюється шляхом використання таких форм маніпулятивного впливу як шахрайство та обман. Тоді як основою для створення фіктивних ситуацій при прітекстінгу є афера, інтрига, містифікація та провокація.

1. Мохор В. В. Наставления по кибербезопасности (ISO/IEC 27032:2013) / В. В. Мохор, А. М. Богданов, А. С. Килевой. К. : ООО "Три-К", 2013. – 129 с.

2. Information technology. Security techniques. Information security management systems. Requirements : ISO/IEC 27001:2013. – Second edition 2013-10-01. – Geneva, 2013. – P. 23.

3. Загальні положення щодо захисту інформації в комп'ютерних системах від несанкціонованого доступу : НД ТЗІ 1.1-002-99. – Чинний від 1999-04-28. – К. : ДСТСЗІ СБ України, 1999. – 15 с. – (Нормативний документ системи технічного захисту інформації).

4. Про захист інформації в інформаційно-телекомунікаційних системах : Закон України від 05.07.1994 № 80/94-ВР [Електронний ресурс] / Закони // Верховна Рада України. – Режим доступу : http://zakon2.rada.gov.ua/laws/show/80/94-%D0%B2%D1%80. – Дата доступу : верес. 2017. – Назва з екрану.

5. Захист інформації. Технічний захист інформації. Основні положення : ДСТУ 3396.0-96. – Чинний від 1997-01-01. – К. : ДСТСЗІ СБ України, 1996. – 11 с. – (Нормативний документ системи технічного захисту інформації).

6. Захист інформації. Технічний захист інформації. Порядок проведення робіт : ДСТУ 3396.1-96. – Чинний від 1997-01-01. – К. : ДСТСЗІ СБ України, 1996. – 15 с. – (Нормативний документ системи технічного захисту інформації).

7. Захист інформації. Технічний захист інформації. Терміни та визначення : ДСТУ 3396.2-96. – Чинний від 1997-04-11. – К. : ДСТСЗІ СБ України, 1996. – 19 с. – (Нормативний документ системи технічного захисту інформації).

8. Про затвердження Концепції технічного захисту інформації в Україні : Постанова Кабінету міністрів України від 08.10.1997 № 1126 [Електронний ресурс] / Постанови // Кабінет міністрів України. – Режим доступу :

http://zakon2.rada.gov.ua/laws/show/1126-97-%D0%BF. – Дата доступу : верес. 2012. – Назва з екрану.

9. Цуркан О.В. Використання соціальної інженерії для негативного інформаційно-психологічного впливу на людину в кібернетичному просторі / О.В. Цуркан, В.В. Мохор, В.В. Цуркан // Актуальні проблеми управління інформаційною безпекою держави : збірник матеріалів науково-практичної конференції, 20 березня 2014 року, м. Київ. – К. : Наук.-вид. центр НА СБ України, 2014. – Ч. 1. – С. 214-215.

10. Цуркан О.В. Соціоінженерний аспект негативного інформаційно-психологічного впливу на людину в кіберпросторі / О.В. Цуркан, В.В. Мохор // ITSEC: IV міжн. наук.-практ. конф., 20 – 23 травня 2014 р. : тези доп. – К. : НАУ, 2014. – С. 46.

11. Цуркан О.В, Маніпулятивна форма соціоінженерного впливу на особистість в кіберпросторі / В.В. Мохор, О.В. Цуркан, Р.П. Герасимов // Актуальні проблеми управління інформаційною безпекою держави: зб. матер. наук.-практ. конф. (Київ, 19 березня 2015 року). – К. : Центр навч., наук. та період. видань НА СБ України, 2015. – С. 303-304.

12. Остроухов В. В. Інформаційно-психологічна безпека особи: соціально-правові аспекти / В. В. Остроухов // Інформаційна безпека людини, суспільства, держави. – 2010. – № 1(3). – С. 38-41.

13. Жарков Я. М. Інформаційно-психологічне протиборство (еволюція та сучасність): Монографія / Я. М. Жарков, В. М. Петрик, М. М. Присяжнюк та ін. – К.: ПАТ "Віпол", 2013. – 248 с.

14. Krombholz K. Advanced social engineering attacks / K. Krombholz, H. Hobel, M. Huber, E. Weippl // Journal of information security and applications, (2014), pp. 1-10, http://dx.doi.org/ 10.1016/j.jisa.2014.09.005.

15. Цуркан О.В. Класифікація атак соціального інжинірингу / О.В. Цуркан, А.В. Жилін // Моделювання (Київ, 15-16 січ. 2009 р.) : XXVIII наук.-техн. конф.: тези доп. – К. : ПП "Системи, технології, інформаційні послуги", 2009. – С. 36 – 37.

16. Цуркан О.В. Аналіз соціоінженерних атак на людину в кіберпросторі / В.В. Мохор, О.В. Цуркан // Інформаційні технології та безпека: засади забезпечення інформаційної безпеки, 28 травня 2014 р.: матеріали міжнародної конференції ІТБ-2014. – К. : ІПРІ НАН України, 2014. – С. 100-102.

17. Winterfeld S. The Basics of Cyber Warfare: Understanding the Fundamentals of Cyber Warfare in Theory and Practice / S. Winterfeld, J. Andress. – Waltham : Elsevier, 2013. – 152 p.

18. Mouton F. Necessity for ethics in social engineering research / Francois Mouton, Mercia M. Malan c, Kai K. Kimppa d, H.S. Venter // ScienceDirect (2015), http://dx.doi.org/10.1016/j.cose.2015.09.001.

19. Mouton F. Social engineering attack examples, templates and scenarios / F. Mouton, L. Leenen, H. Venter // Computers & Security (2016), http://dx.doi.org/doi: 10.1016/j.cose.2016.03.004.

20. Junger M. Priming and warnings are not effective to prevent social engineering attacks / M. Junger, L. Montoya, F.-J. Overink // Computers in Human Behavior (2017), http://dx.doi.org/ 10.1016/j.chb.2016.09.012.

ВИКОРИСТАННЯ ОНТОЛОГІЧНОГО АНАЛІЗУ ДЛЯ СЕМАНТИЧНОЇ РОЗМІТКИ СТАНДАРТІВ З ІНФОРМАЦІЙНОЇ БЕЗПЕКИ

Рогушина Ю.В.[1], ladamandraka2010@gmail.com
Гладун А.Я.[2,3] glanat@yahoo.com
[1]Інститут програмних систем НАН України, Київ, Україна
[2]Інститут спеціального зв'язку та захисту інформації, НТУУ «КПІ ім. Сікорського», , Київ, Україна
[3]Міжнародний науковий центр інформаційних технологій і систем НАНУ, Київ, Україна

Анотація

Аналіз публікацій вказує на велику кількість національних стандартів, безпосередньо пов'язаних з інформаційною безпекою, які вже розроблені або знаходяться на стадії узгодження. Між цими стандартами та їх елементами існує складна та слабо структурована система ієрархічних відношень, взаємозв'язків, посилань та відповідностей, що потребує відповідної формалізації. Тому розробка нормативно-правової бази України для підтримки інформаційної безпеки потребує створення методів та засобів обробки і аналізу цієї інформації на семантичному рівні.

Через недостатньо глибоко стандартизовану та усталену україномовну термінологію в стандартах використовуються різні варіанти перекладу основних понять та назв. Це призводить до проблем із пошуком потрібної користувачам інформації у неструктурованих природномовних документах. Тому пропонується використовувати Wiki-технології для подання знань, пов'язаних з проблемами інформаційної безпеки. Wiki забезпечує динамічне поновлення інформації, пошук та навігацію у мультилінгвістичних ресурсах, а семантизація Wiki-ресурсу (приміром, за допомогою Semantic MediaWiki) дозволяє виконувати розмітки та здійснювати запити на рівні знань.

На жаль, вбудовані механізми менеджменту знань у Semantic MediaWiki не достатньо функціональні, і тому ми пропонуємо інтегрувати їх із засобами онтологічного аналізу: онтологія предметної області "Інформаційна безпека" використовується як базис для семантичної розмітки відповідного Wiki-ресурсу.

В роботі запропоновано приклад використання такого підходу, що демонструє семантичну розмітку Wiki-сторінок, побудованих на

основі аналізу стандартів з інформаційної безпеки, та виконання запитів до них.

Вступ

Останнім часом посилюються вимоги до використання нормативно-правового забезпечення України щодо оцінки відповідності систем управління інформаційною безпекою правилам та процедурам, які відповідають міжнародним стандартам та кращим міжнародним практикам. Разом з цим відбуваються якісні зміни у процесах керування знаннями для підтримки пошуку знань в системах нормативно-правової документації та забезпечення інтероперабельності термінологічної бази у сфері інформаційної безпеки, зумовлені інтенсивним впровадженням сучасних інтелектуальних інформаційних технологій.

У зв'язку з важливістю розробки національного нормативно-правового забезпечення інформаційної безпеки, великим обсягом міжнародних стандартів в цій сфері та термінологічними проблемами гармонізації національних стандартів України стає актуальною організація ефективного пошуку та навігації у різноманітних документах, пов'язаних з цією предметною областю. Важливість цієї проблеми обумовлена тим, що стандарти мають великий обсяг та складну структуру і семантичні методи дозволяють підвищення рівня пошуку інформації в галузі інформаційної безпеки. Це викликає потребу в створенні таких методів та засобів забезпечення доступу до необхідних відомостей, які легко використовувати, але які мають достатню виразну потужність для вирішення такої задачі.

Захист мереж. Основні концепції

Швидкий розвиток загальнодоступних мережних технологій (дротових, бездротових, Інтернету) пропонує значні можливості для діяльності різних організацій, як для транспортування інформації, так для створення і надання онлайнових сервісів. Однак, хоча це середовище забезпечує значні переваги, з'являються нові ризики безпеки, які потребують ефективного керування. Організаціям, які довіряють використанню великого об'єму інформації та відповідних мереж для своєї діяльності, втрата конфіденційності, цілісності і доступності інформації та сервісів може створити істотний несприятливий вплив на робочий процес. Отже, на перший план виступає належна захищеність мереж і пов'язаних з ними інформаційних систем і інформації. Іншими словами, реалізація і

підтримка адекватної мережної безпеки є абсолютно критичними для успіху діяльності будь-якої організації.

Сьогодні актуальною є проблема інформаційного убезпечення. Досвід експлуатації інформаційних систем і ресурсів у різних сферах життєдіяльності показує, що існують різні й досить реальні загрози втрати інформації, що приводять до значних матеріальних та інших збитків [1-4]. Розробка міжнародної та національної нормативно-правової бази документів для підтримки інформаційної безпеки, потребує застосування інтелектуальних методів та засобів обробки і аналізу релевантної інформації на семантичному рівні. Це створює передумови для усунення технічних бар'єрів між країнами в сфері технічного регулювання, в тому числі з питань безпеки.

Як відомо, процес розроблення стандартів у сфері інформаційних технологій (IT) і, зокрема, в напрямі інформаційної безпеки відбувається безупинно. Розробляють стандарти для відкритих систем, зокрема у сфері безпеки IT, спеціалізовані міжнародні організації і консорціуми. Критерії оцінювання інформаційної безпеки є методологічною базою для визначення вимог захисту комп'ютерних систем від несанкціонованого доступу, створення захисних систем та оцінювання ступеня захищеності. Послідовно публікуються проекти та версії стандартів на різних стадіях узгодження та затвердження. Значну кількість стандартів поетапно поглиблюють і деталізують у вигляді сукупності взаємозалежних груп щодо концепцій та визначених структур [5,6].

В процесі використання результатів гармонізованих міжнародних та національних стандартів у користувачів дуже часто виникає проблема зіставлення стандартів, яка на нашу думку може бути ефективно вирішена на основі семантичного підходу. Проблема зіставлення стандартів виникає при наявності в нормативно-правовій бази документів, розроблених, гармонізованих, модернізованих різними міжнародними та національними органами зі стандартизації, наприклад, ISO/IEC, ITU-T, IEEE, CEN, CENELEC, IAB, WOS, ANSI, W3C, ECMA, X/Open, OSF, OMG тощо. В якості узагальнених показників, що характеризують стандарти інформаційної безпеки, пропонується використовувати універсальність, гнучкість, гарантованість, реалізованість і актуальність.

Постановка задачі

Для забезпечення користувачам ефективного використання наявного національного та міжнародного нормативно-правового забезпечення з інформаційної безпеки, що включає динамічне поновлення інформації, пошук та навігацію у взаємопов'язаних

мультилінгвістичних ресурсах, пропонується використовувати семантичні Wiki-технології для подання таких знань та підвищення ефективності розв'язання поставлених задач. Щоб інтегрувати знання з різних джерел, потрібно розробити онтологію предметної області "Інформаційна безпека", елементи якої будуть використовуватися для семантичної розмітки Wiki-сторінок. Використання онтологічного підходу забезпечить інтероперабельність цих знань та можливість їх застосування в різних інформаційних системах. Це вимагає створення методів співставлення елементів Wiki-сторінок з елементами онтології та засобів автоматизації семантичної розмітки природномовних текстів.

Онтології як засіб подання розподілених знань

Для представлення та спільного використання розподілених знань різних предметних областей (ПрО) в сучасних інформаційних системах зараз широко застосовуються онтології [7]. Онтології базуються на фундаментальному теоретичному базисі дескриптивних логік, для них вже існують загальноприйняті стандарти опису, мови і програмні засоби.

Онтологічний аналіз дозволяє перетворювати представлення певної особи або групи осіб про зовнішній світ у формалізований набір термінів і правил їхнього використання, придатний для автоматизованої обробки. Онтологію можна розглядати як базу знань (БЗ) спеціального виду із семантичною інформацією про визначену ПрО [8]. Компоненти, які використовують в різних формалізаціях онтологічних описів ПрО, залежать від як парадигми представлення, так і від цілей побудови такої онтології. Більшість формальних моделей онтологій містять класи та їх екземпляри, властивості класів, відношення між класами та екземплярами класів і обмеження їх використання. В рамках проекту Semantic Web створено формальну мову подання онтологій OWL [9] та мову опису метаданих RDF, які фактично стали стандартом для подання знань у розподілених Web-застосунках. Для них створено велику кількість інструментальних засобів обробки, приміром, Protégé.

Semantic MediaWiki

Одним за найбільш поширених сьогодні Wiki-двигунів є MediaWiki– вільне програмне забезпечення з відкритим кодом для гіпертекстового Wiki-середовища, що є платформою для створення довідників, енциклопедій і каталогів. Ця програма дозволяє створювати та редагувати різноманітні Wiki-ресурси, надає

користувачам зручний та інтуїтивно зрозумілий інтерфейс та підтримує більшість функцій, необхідних для зручної колективної роботи.

Semantic MediaWiki (SMW) [10] забезпечує семантизацію інформації у Wiki-ресурсі. Це дозволяє користувачам додавати семантичні анотації до Wiki-сторінок, перетворюючи MediaWiki у семантичний ресурс. SMW перетворює MediaWiki на семантичний ресурс, дозволяючи автоматично інтегрувати інформацію, що знаходиться у різних Wiki-сторінках, генерувати відповіді на складні семантичні запити, будувати бази знань та візуалізувати їх результати, що відображають знання щодо семантичних властивостей та категорій інформаційних об'єктів, виконувати логічне виведення тощо, тобто обробляти інформацію на рівні знань.

У Semantic MediaWiki використовуються такі додаткові елементи розмітки, як *семантичні властивості* (для створення даних) та *семантичні запити* (для використання даних). Semantic MediaWiki має досить високу виразну потужність, надійну реалізацію і зручний користувальницький інтерфейс, сформовані на її основі IP часто обновляються і швидко збільшуються. Використання таких джерел обумовлюється тим, що Wiki-ресурси динамічно обновляються всім співтовариством користувачів, мають чітко визначену і просту для розуміння структуру і забезпечують обробку інформації на семантичному рівні, і, таким чином, забезпечують технологічну платформу для групового керування знаннями. Слід відмітити, що, на Semantic MediaWiki базується все більша кількість розробок сайтів, порталів та енциклопедичних видань, що викликає розвиток засобів для автоматизованої обробки таких баз знань. Тому використання саме цього програмного забезпечення значно збільшує можливості використання розроблених на його базі довідників та енциклопедій.

Онтологія "Інформаційна безпека"

Ця онтологія описує відповідну ПрО. Вона базується на наборі національних та міжнародних стандартів, пов'язаних з питаннями інформаційної безпеки. На сьогодні побудовано наступні базові класи – "Предметна область", "Стандарт" та "Термін стандарту", для яких побудовано відповідні екземпляри та підкласи. Екземпляри цих основних класів пов'язані об'єктними властивостями "Описано в стандарті", "Посилається на стандарт", "Відноситься до ПрО" тощо. Щоб формалізувати семантику відношень між елементами

онтології, що згенеровані за термінологічним базисом стандартів, побудовано специфічні для ПрО об’єктні властивості.

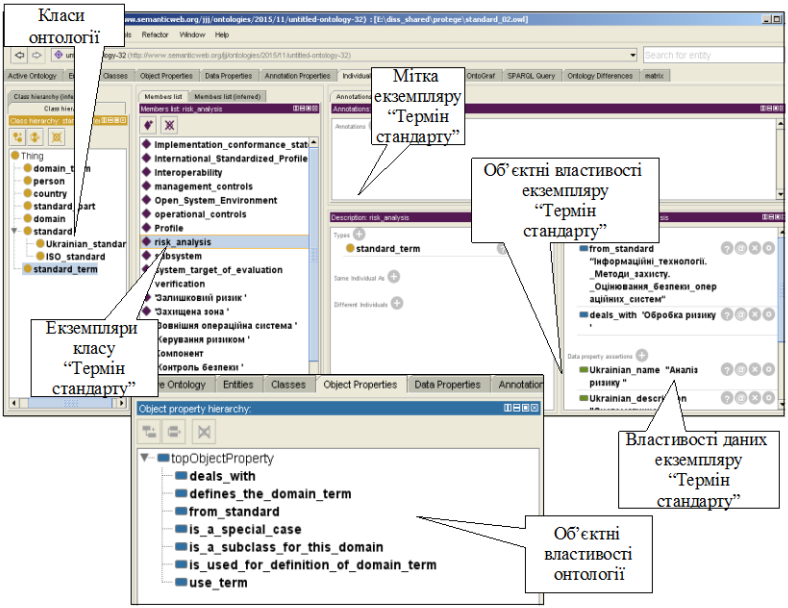


Рис.1. Об’єктні властивості та екземпляри класів в онтології “Інформаційна безпека”

Семантична розмітка Wiki-ресурсу термінами онтології “Інформаційна безпека”

У семантичній розмітці Wiki-ресурсу класам онтології відповідають категорії, елементам – сторінки, а об’єктним відношенням – семантичні властивості. Більш детально правила побудови цих відповідностей розглянуто в [11,12,13].

Окремі Wiki-сторінки доцільно створити як для кожного стандарту в цілому (перелік його підрозділів), так і для окремих підрозділів та визначень. Крім того, доцільно використовувати гіперпосилання на інші Wiki-ресурси, приміром, на Вікіпедію або на електронну версію Великої української енциклопедії (е-ВУЕ).

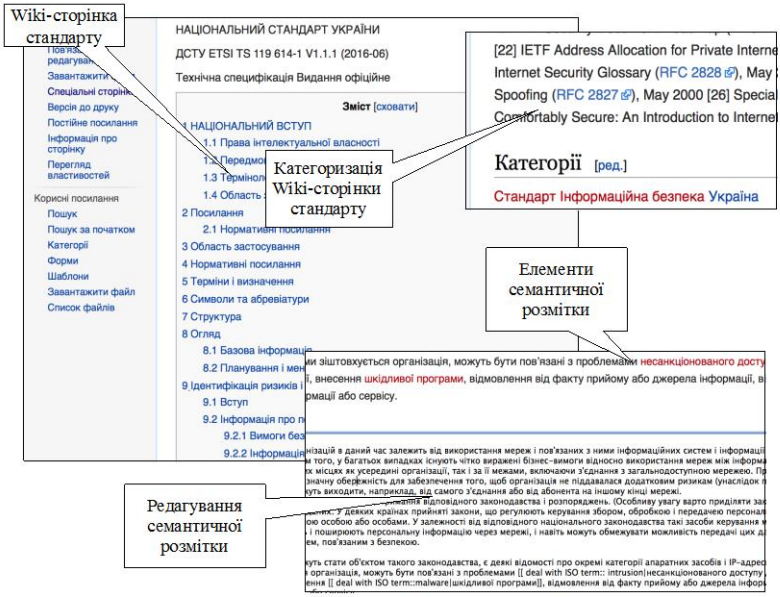


Рис.2. Семантична розмітка Wiki-ресурсу за допомогою онтології “Інформаційна безпека”

Таким чином, якщо вже побудована онтологія мовою OWL, тоді її досить просто використовувати в Semantic MediaWiki. Але зворотний процес не може бути цілком автоматизований. Більш того, при автоматичній генерації онтології за Semantic MediaWiki будуть загублені відомості, що містяться в OWL-онтології та стосуються характеристик класів і властивостей, що не мають аналогів у Wiki (зокрема, про еквівалентність класів і властивостей, їхній неперетин, про їхню область значення і визначення). У той же час, частина контента Semantic MediaWiki не може бути безпосередньо трансформована в онтології [14]. Наприклад, той факт, що в сторінках використаний одні і той же шаблон, говорить про те, що на цих сторінках описані інформаційні об'єкти одного типу, але для відображення цього в онтології треба створити специфічний клас і зв'язати його з елементом сторінки. Але потім за такою онтології неможливо зрозуміти, що треба створити в Wiki – шаблон чи категорію: вибір залежить від користувача, тому що для ІО зі специфічної, але формалізованою структурою доцільно

створювати шаблони, а у всіх інших випадках – сторінки, що будуть віднесені до якої-небудь категорії. Крім того, не можливо зв'язати з класом онтології не всю сторінку, а її конкретний фрагмент (крім тих випадків, коли в нього є підзаголовок).

Висновки

Використання технології Wiki та її семантизації для подання предметної області інформаційної безпеки є актуальним завданням у сучасну епоху зростання великих обсягів нормативно-правової мультилінгвістичної документації. Онтологічний підхід для вирішення поставленого завдання забезпечує інтероперабельне представлення знань та застосування відповідних технологій і форматів і дозволяє інтегрувати ці елементи знань із знаннями, представленими в Semantic Web. Онтологія національних та міжнародних стандартів (та інших документів) з інформаційної безпеки використовує семантичну розмітку природномовних текстів та пов’язує описи реальних або розроблюваних систем в яких реалізовані ці стандарти. Такий підхід значно спрощує пошук та аналіз таких документів та забезпечує можливість їх автоматичної обробки. Прикладом застосування такого підходу може бути пошук рекомендації щодо ідентифікації та аналізу ризиків мережної безпеки для визначення вимог до безпеки мереж. При цьому користувачеві не потрібно буде самостійно відслідковувати зміни в останній редакції обраного стандарту або передивлятися опис кожної потенційно придатної системи – співставлення має виконуватися автоматизовано. Технологічною основою для такої семантичної розмітки та пошуку може стати середовище Wiki із семантичним розширенням.

Перспективи розвитку у майбутньому результатів нашої роботи передбачають створення *глобальної семантичної мережі стандартів*, яка пов’яже окремі національні та міжнародні стандарти; об’єкти, що використовують ці стандарти та посилаються на них (як матеріальні, так і інформаційні об’єкти); фахівців, що є експертами в сфері розробки стандартів, та організації різного рівня, що підтримують різні види діяльності, пов’язаної із розробкою та використанням стандартів.

Список використаних джерел

1. Інформаційні технології. Методи захисту. Захист мережі. Частина 1. Огляд і поняття (ISO/IEC 27033-1:2015, IDT): ДСТУ ISO/IEC 27033-1:2018. — Вид. офіц. — Вперше введ. 2016-10- 01. —

К. : Держспоживстандарт України, 2017. — 89 с. (Національний стандарт України).

2. ISO/IEC 27033-2:2015 «Information technology – Security techniques – Network security – Part 2: Guidelines for the design and implementation of network security (Інформаційні технології. Методи захисту. Захист мережі. Частина 2. Настанови щодо розроблення та впровадження безпеки мережі).- www.iso.org.- 28p.

3. ISO/IEC 27033-3:2015 «Information technology – Security techniques – Network security – Part 3: Reference networking scenarios - Threats, design techniques and control issues (Інформаційні технології. Методи захисту. Захист мережі. Частина 3. Еталонні сценарії побудови мереж. Загрози, методи розроблення та проблеми керування) .- www.iso.org.- 30p.

4. ISO/IEC 27033-3:2015 «Information technology – Security techniques – Network security – Part 4: Securing communications between networks using security gateways (Інформаційні технології. Методи захисту. Захист мережі. Частина 4. Захист обміну даними між мережами за допомогою шлюзів безпеки) .- www.iso.org.- 22p.

5. Гладун А.Я., Хала К.О. Таксономія стандартів з інформаційної безпеки // Наука, технології, інновації, сер. Інформаційні технології, 2017, №2 .-С.53-67.

6. ДСТУ ISO/IEC 15408-1:2017 Інформаційні технології. Методи захисту. Критерії оцінки. Частина 1. Вступ та загальна модель (ISO/IEC 15408-1:2009, IDT).-223с.

7. Gruber, T.R. The role of common ontology in achieving sharable, reusable knowledge bases. – http://www.cin.ufpe.br/~mtcfa/files/ 10.1.1.35.1743.pdf.

8. Guarino N. Formal Ontology in Information Systems. Formal Ontology in Information Systems. Proceedings of FOIS'98, 3-15, 1998.

9. OWL 2 Web Ontology Language Document Overview. W3C. 2009 (Електронний ресурс). – http://www.w3.org/TR/owl2-overview/.

10. Krotzsch M., Vrandecic D., Volkel M.: Semantic MediaWiki. – http://c.unik.no/images/6/6d/SemanticMW.pdf

11. Рогушина Ю.В. Сопоставление семантических информационных ресурсов web на основе онтологического анализа // International Journal "Information Technologies & Knowledge" Volume 11, Number 1, 2017. – С.49-71. – http://www.foibg.com/ijitk/ijitk-vol11/ijitk11-01-p03.pdf.

12. Гладун А.Я., Рогушина Ю.В. Онтологічній підхід до проблем підвищення якості розроблення національних стандартів

України // Стандартизація, сертифікація, якість, № 2, 2016. – С. 19–28.

13. Гладун А.Я., Рогушина Ю.В. Семантичні технології: принципи та практики: монографія; ред. С. Кузнецов. — К. : ТОВ “ВД “АДЕФ- Україна”, 2016. — 387 с.

14. Рогушина Ю.В., Гладун А.Я., Снігир Г.В. Онтологічний підхід до розробки національних стандартів України з оцінювання безпеки інформаційних технологій// Информационные технологи и безопасность. ИПРИ НАНУ. Киев, 2017.-С.58-72.

ФУНКЦИОНАЛЬНАЯ СТОЙКОСТЬ УНИВЕРСАЛЬНОЙ МОБИЛЬНОЙ РЕКОНФИГУРИРУЕМОЙ СИСТЕМЫ ПРИ ВОЗДЕЙСТВИИ ЭЛЕКТРОМАГНИТНОГО ИЗЛУЧЕНИЯ ВЫСОКОЙ МОЩНОСТИ

Рубан И.В., Чурюмов Г.И., Токарев В.В., Ткачев В.Н.
ХНУРЭ, г. Харьков, Украина

This paper presents topical scientific and applied problems of increasing the survivability of a general-purpose reconfigurable system under conditions of external influence of powerful microwave pulses of electromagnetic radiation during registration, temporary storage and data transmission. As the general-purpose reconfigurable system one consider the following components: main drone and the micro-multi-drones, a system for receiving, storing and processing data. A brief description of the functionality of individual components of the system is provided in the framework of solving common problems: registration, transmission and processing of data. The current status of research devoted to the influence of microwave pulses of electromagnetic radiation to the hardware part of the general-purpose systems is analyzed. Theoretical studies were carried out to determine the regularities of the influence of microwave pulses of different power for possible software reconfiguration of the general-purpose reconfigurable system for solving the problem of increasing survivability. Conclusions are drawn regarding the search for ways to increase the survivability of both components and the system as a whole.

1. ВВЕДЕНИЕ

Особую актуальность и практический интерес вызывают исследования в области создания радиоэлектронных систем и средств, генерирующих мощное электромагнитное излучение (ЭМИ) для деструктивного воздействия на полупроводниковую элементную базу. Речь идет о подходах, основанных на новых физических принципах. Так как электромагнитное излучение высокой мощности может привести к выводу из строя радиоэлектронных компонентов универсальной реконфигурируемой системы без её физического

уничтожения, то актуальной является научно-прикладная задача повышения функциональной стойкости компонентов системы путем реконфигурирования программной архитектуры. Это позволяет производить перенастройку различных узлов, блоков и изменять функционирование универсальной системы в целом [1-3].

Анализируя сущность электромагнитного излучения высокой мощности в качестве фактора влияния на универсальную реконфигурируемую систему, важно отметить следующие особенности:

1. Динамика изменения уровня выходной мощности (как средней до десяток – сотен кВт, так и пиковой (импульсной) до единиц МВт).

2. Укорочение длины волны ЭМИ, в частности, за счет новых наработок в освоении миллиметрового и терагерцового диапазонов, а также расширение полосы генерируемых частот и уменьшение длительности импульса до единиц наносекунд – сотен пикосекунд.

3. Увеличение уровня мощности и укорочение длительности импульса повышает эффективность применения запасенной в импульсе энергии на больших расстояниях.

Целью настоящего доклада является рассмотрение возможных путей защиты универсальных мобильных систем от воздействия ЭМИ высокой мощности путем программной реконфигурации для повышения их функциональной стойкости.

2. ОСОБЕННОСТИ ФУНКЦИОНИРОВАНИЯ УНИВЕРСАЛЬНОЙ МОБИЛЬНОЙ СИСТЕМЫ

Универсальная мобильная реконфигурируемой система – это множество компонентов, которые обеспечивают регистрацию, передачу, хранение и обработку различных типов данных, а также находятся в отношениях и связях друг с другом посредством защищенных каналов связи, образуя определённую целостность и единство при выполнении поставленных задач. Такая система характеризуется наличием механизмов реконфигурации, реализуя автоматическую перестройку структуры сети обмена данными внутри системы для достижения наивысшей точки эффективности достижения цели функционирования системы при наличии работоспособных компонентов [4,5].

В докладе рассматривается совокупность таких компонентов: рой микромультикоптеров, мультикоптер-матка, компонент приема и хранения данных с целью дальнейшей ее обработки (рис. 1).

Предполагается, что данные между компонентами рассматриваемой системы (микромультикоптерами – микромультикоптерами, микромультикоптерами – мультикоптером-маткой, мультикоптером-маткой – компонентом хранения и обработки данных) передаются через зашифрованные каналы передачи данных.

Тогда, универсальную мобильную реконфигурируемую систему можно описать в виде целевой функции:

$$S = \{\Psi, X, F\}, \quad (1)$$

где Ψ – совокупность компонентов рассматриваемой системы; X – каналы передачи данных; F – цель системы.

Рассмотрим функции каждого из компонентов системы.

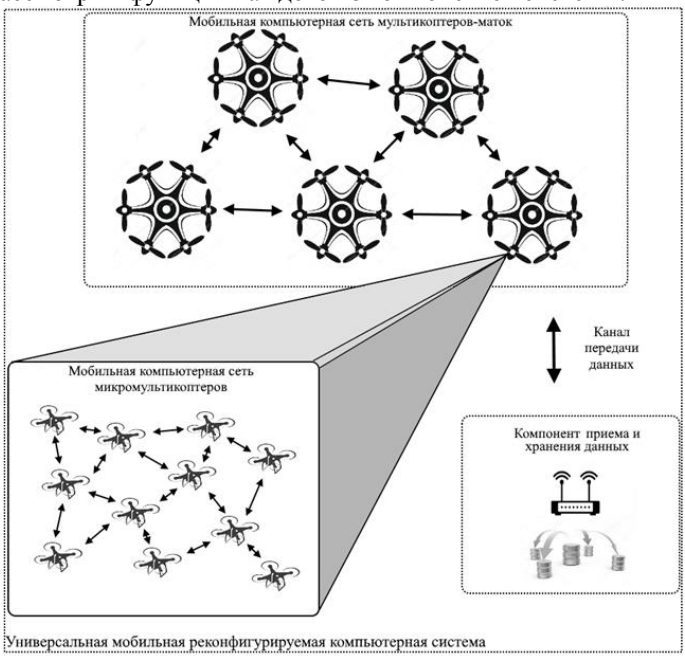


Рисунок 1 – Универсальная мобильная реконфигурируемая система

1. Рой микромультикоптеров. Данный компонент представляет собой часть подсистемы «микромультикоптер – иультикоптер-матка». Это обосновывается тем, что мультикоптер-матка выполняет

одну из функций – физический транспорт для микромультикоптеров. Функциями роя микромультикоптера является:

а) регистрация первичных данных согласно поставленной задачи к системе в целом;

б) регистрация электромагнитного излучения высокой мощности;

в) квазимаршрутизация с промежуточным хранением данных при передачи их к мультикоптеру-матке [6];

г) передача тривожного бродкаст-сообщения о электромагнитном излучении высокой мощности с целью обеспечения функциональной стойкости роя в целом.

2. Мультикоптер-матку можно рассматривать как концентратор данных в отношении роя микромультикоптеров и как систему промежуточного хранения данных в отношении универсальной системы. Характеризуется следующими выполняемыми функциями:

а) транспортировка роя микромультикоптеров в заданную точку согласно цели системы;

б) прием и хранение данных от роя микромультикоптеров [6];

в) ретранслятор команд управления рою микромультикоптеров (если есть возможность);

г) забор роя микромультикоптеров после выполнения их задач.

3. Компонент приема и хранения данных рассматривается как стационарный объект системы.

3. ПРИМЕР ОБЕСПЕЧЕНИЯ ФУНЦИОНАЛЬНОЙ СТОЙКОСТИ УНИВЕРСАЛЬНОЙ МОБИЛЬНОЙ РЕКОНФИГУРИРУЕМОЙ СИСТЕМЫ

Рассмотрим поведение компонентов универсальной мобильной реконфигурируемой системы при воздействии на них электромагнитного излучения высокой мощности.

При выполнении основной задачи универсальной системой в случае идентификации ЭМИ высокой мощности микромультикоптерами проводится оценка необходимости реконфигурации программной архитектуры для обеспечения функциональной стойкости универсальной мобильной реконфигурируемой системы.

Как показано на рис. 2., в случае возникновения вышеописанной ситуации применительно к микромультикоптеру, данные о электромагнитном излучении посредством сети роя микромультикоптеров передаются к мультикоптеру-матке (штрихпунктирный маршрут на рис. 2). Мультикоптер-матка после

приема и обработки тривожного бродкаст-сообщения принимает решение о реконфигурации программной архитектуры модели поведения роя. Например, при загрузке конфигурационного файла с моделью поведения о сборе роя, данная команда передается посредством функций квазимаршрутизации в сети роя.

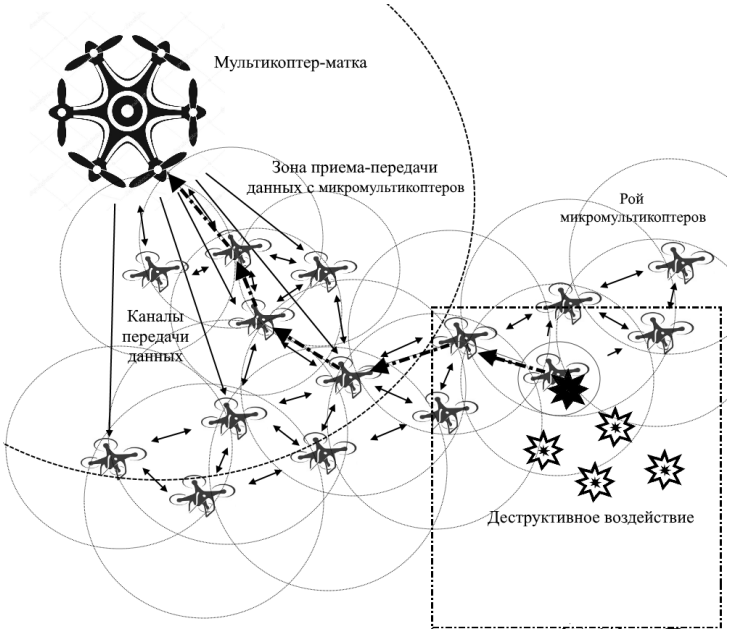


Рисунок 2. – Пример обеспечения функциональной стойкости универсальной мобильной реконфигурируемой системы

ВЫВОДЫ

В докладе заложены научно-методические основы обеспечения функциональной стойкости универсальной мобильной реконфигурируемой системы при воздействии электромагнитного излучения высокой мощности. В дальнейшем предполагается проведение исследований способов обеспечения живучести подобного класса систем. Спектр возможных сфер применения универсальных систем чрезвычайно широкий, за счет гетерогенности факторов воздействия.

Исследования проводятся в рамках выполнения фундаментальной научно-исследовательской темы «Создание научно-методических основ обеспечения живучести сетевых систем обмена информацией в условиях внешнего воздействия мощного СВЧ-излучения» на базе научно-учебной лаборатории Моделирования систем кафедры Электронных вычислительных машин ХНУРЭ.

СПИСОК ЛИТЕРАТУРЫ

1. Giri D.V. High-Power Electromagnetics (HPEM). From the 1960s into the 21st century. Book of Abstracts. EUROEM-2012, 2-6 July 2012, Toulouse, France. – 16 p.

2. Benford J., Swegle J.A. and Schamiloglu E. High Power Microwaves. Second Edition. CRC Press. 2007. – 556 p.

3. Starostenko V.V., Taran Ye.P., Churyumov G.I. and other. Near Field Zone of a Integrated Circuit Exposed to an Electromagnetic Wave in a Waveguide. Technical Physics Letters, Vol. 29, No. 1, 2003, pp. 29 – 31.

4. Радченко В.А. Мобильная подсистема «Мультикоптер-сенсорная сеть» в компьютерной системе хранения BIG DATA / В.А. Радченко, Д.А. Руденко, В.Н. Ткачев, В.В. Токарев // Системи управління, навігації та зв'язку. – Полтава. – 2017. - №4(44). – С. 102-105.

5. Радченко В.А. Проблема передачі даних типу BIG DATA у мобільній системі «Мультикоптер–сенсорна мережа» / В.А. Радченко, В.О. Лебедев, В.Н. Ткачев, В.В. Токарев// – Системи управління, навігації та зв'язку. – Полтава – 2017. – №2 (42). – С.154-157 .

6. Tkachov V. Method for transfer of data with intermediate storage / V. Tkachov, V. Savanevych // Problems of Infocommunications Science and Technology, 2014 First International Scientific-Practical Conference. – 14-17 Oct. 2014, Kharkiv. – 2 p. DOI: 10.1109/INFOCOMMST. 2014.6992315.

THE TECHNOLOGY OF SEMANTIC MODELING FOR KNOWLEDGE MANAGEMENT SYSTEMS IN ENVIRONMENT PROTÉGÉ 5

Senchenko V.R.[1], Koval A.V.[2,1]
[1]Institute for Information Recording of NAS of Ukraine, Kyiv, Ukraine
[2]National Technical University of Ukraine
"Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine

The main feature of semantic technology is to store and maintain the integrity of semantics (meaning of knowledge) separately from the contents of data files and from the code of the programs that implement them [1].Semantic simulation technology differs from traditional methods that combine the meaning of data and the processing procedures directly in the program code. This often leads to the need for a radical manual redevelopment of data structures and total revision of programs during their development or migrate to another platform.

Semantic technologies allow obtaining logical conclusions based on the rules of conceptual models and perform automatic re-design of data structures. Experience the semantic modeling of intelligent systems using ontologies, indicates that any subject domain (SD) can be described considerable number of ontologies. From the methodological point of view, it is quite understandable – each ontology reflects a person's perception of the developer of the functioning model of ontology (the main entities, classes, subclasses and their relationships within the general idea about of the subject domain). Therefore, it is advisable for the developer to apply such methodology and tools, which allow not only to create a model of ontology, but also to correct it in the process of mastering it, and understanding the features of its functioning, been aimed at constructing the most correct model.

Under the term ontology, we understood a system of concepts domain, which represented by a set of entities and their properties, interconnected relations, in order to create knowledge bases on their basis. Consequently, the main purpose of ontological modeling is to create a formalized knowledge model of the domain, which is stored electronically and may further improved through a more in-depth understanding of the features of the subject domain.

For the implementation of knowledge-based systems, it is expedient to use common language for describing ontologies such as OWL Lite, OWL DL, and OWL 2 [2] that use discriminatory logic for knowledge work. In this case, the semantic modeling process has performed using the Open

Knowledge Base Connectivity (OKBC) [3]. This model is based on the theory of frames and uses concepts such as "conceptualization of classes", "objects", "slots", "facets", and "inheritance" for representing knowledge about the subject area, which allows create various knowledge-based applications with a high level of interoperability (interoperability).

The formal semantics of the subject domain on the OWL describes how to obtain logical investigations, having the ontology of SD, which is to obtain facts that are not represented literally in the ontology, but logically follow from its semantics.

1. Methodological aspects of semantic modeling using Web Ontology Language

An applied ontology should describe concepts that depend both on the ontology of tasks and on the ontology of the subject domain. The purpose of the applied ontology is to create an electronic model of knowledge that allows:

- creation of general terminology of a subject domain, for common use and understanding by all users – system developers;
- give an exact and consistent definition of the meaning of each term;
- provide semantic tasks using axioms that automatically allow you to answer the main questions about the subject domain.

One of the most common languages for representing ontology is the Web Ontology Language (OWL). The OWL language contains elements such as Classes, Properties, and Individuals. [9]. All concepts of the subject domain are divided into classes, subclasses, instances (copies). The tag describes classes:

<owl:Class rdf:about="http://untitled-ontology-9#APM_MK">

Thus, the process of semantic modeling using the unified OKBC model consists of the following steps:

- Definition of the concepts of the subject domain, that is, the basic concepts – classes, entities, categories (Active Ontology, Entities, and Classes) that describe the SD;
- Determination of the set of properties, which describe the properties of the concepts of SD and allow, in the final sense, to create a knowledge base of the domain. Formation of concept properties in AKVS is performed using the mechanisms of definition, attributes and roles (Data Properties);

- Establishing relations between concepts of the subject area and their properties using the mechanisms of forming predicates (Object Properties), which should take into account the functional orientation of relationships, for example, "solves problems", or "is part of", etc.;
- Setting numerical or logical constraints, which used to describe the properties of instances (Individuals) of the knowledge base. This is realized through the mechanisms of axiom definition (Axiom) or facet (Valiu). For example, the maximum speed for terrestrial objects is limited to the value;
- Formation of knowledge base using the mechanism of description of instances of the knowledge base and its filling (Individual by class);
- development of typical query patterns for the knowledge base using the query language DL Query and the output machine - Reasoner;
- Checking the correctness of the functioning of the ontological model of SD from the point of view of its correspondence to the initial goals and the task and finding gaps in the ontology using the Ontograf research mechanism. The evaluation is based on the analysis of the results of testing by various output machines (Reasoner) and the compilation of various types of requests;
- Development of a strategy for improving the ontological model of SD and carrying out the relevant work.

Taking into account the complexity and ambiguity that arises in the process of describing the subject area, modern science offers several approaches to the creation of ontology [19]:

- Top-down. The use of this method requires the definition of the most general concepts of the domain, with further detail of objects in the class hierarchy and the concepts of the subject domain.
- Down - to the top. This approach begins with the definition of detailed and specific classes (the end of the tree of the hierarchy), followed by grouping into more general concepts.
- Combination of the first two methods. First, a description of the concepts that are fully understood, then association them into groups and create more complex concepts of subject domain.

2. Definition of the depth and scope of the subject domain

It is advisable to start developing an ontology with the purpose of determining its scope and scale. That is, the answer to a few basic questions:

- Which area will cover the ontology?
- What types of questions should answer information in ontology be?
- Who will use and maintain an ontology?

Of course, the answers to these questions change during the process of designing ontology software, but at any time, they allow you to limit the scale of the ontology if it becomes too complicated.

Consider the methodology for creating an ontology on an example of An Intelligent System For Studying Hydroacoustical Processes.

The main quality of such systems is the accumulation of two types of information:

- real data - hydroacoustic as a result the marine area scanning,
- data modeling - obtained as a result of mathematical modeling the behavior of the object.

Modeling of hydroacoustical processes requires the development of component models, including the information model of the water area, the database of the parameters of the marine environment (sea noise, ground, coastline, water temperature, deep, salinity), and the parameters of hydroacoustic devices and their interaction with the modeling medium.

Hydroacoustic processes allows to take into account both problems of direct modeling (for the purpose of obtaining objective data-knowledge about the marine facilities under study), and combine the obtained data with expert knowledge (represented as a characteristic set of parameters of real objects and their assessments based on the expert's experience).

Generalized structure of the knowledge-based modeling system for the identification, classification and definition of parameters of movement of marine facilities shown in Figure 1. The conceptual model to consist from such structural components [8]:

- **Simulation of marine environment** - performs the functions of creating a simulation scene (parameters of depths, temperatures and salinity, type of bottom, coastline, etc.) and location and specification of the parameters of marine objects (type, size, direction, speed, etc.);
- **Modeling of hydroacoustic device** - sets the scene of the location of fixing devices and their parameters (type, dimensions, sensitivity);
- **Hydroacoustic signals analysis** - is a set of tools for creating models for generating sonar signals and working out methods for their analysis (fast Fourier transforms, digital filtering, spectral, frequency, correlation analysis, etc.);
- **Knowledge Base Maintenance** - contains a set of tools for testing models of object recognition, their identification, classification and

definition of the parameters of the movement of objects, including methods of fuzzy logic;

- **Knowledge base maintenance** – is intended for the organization of tools for forming a knowledge base for solving problems of identification and classification of marine objects;
- **Inference engine** – designed to organize logical inference based on accumulated knowledge, including means of composition of product rules, self-learning and adaptation;
- **Information storage** - is the core of the system and provides information to all the structural components of the modeling system, and also contains information with expert assessments of the experiments conducted;
- **Administration and management tools –** provide settings for services and applications manage user access rights to information resources, manage security and performance of the modeling system.

The proposed system provides opportunities for end-to-end documentation of the processes of hydroacoustic experiments, which gives additional advantages in the formation and accumulation of knowledge about the studied processes, including the formation of scenes and modeling scenarios. Simulation involves the following steps:

1. Creation and maintenance of library of hydroacoustic models including: signal generation models, signal extraction models, signal analysis models, implementation of algorithms for classification and identification of objects.
2. Creation and maintenance tools for Entering both type data – model and real data (scanning water areas) into the knowledge base;
3. Development subject domain ontology;
4. Development Algorithms and software for recognition and Identification types of marine objects
5. Organization of logical inference on ontology
6. Creation learning system (scripts and learning algorithms considering SD).
7. Maintenance tools for rules setting based on fuzzy inference;
8. Development Algorithms identification based on logical inference on the ontology and production rules.
9. Development algorithms for knowledge classification and clustering;
10. Formation of classifiers of the system (noise-emitting objects, hydroacoustic systems, water areas).

11. Scene formation and Creating an experiment scenario
12. Fixing the experiment results in DB
13. Evaluation of simulation results

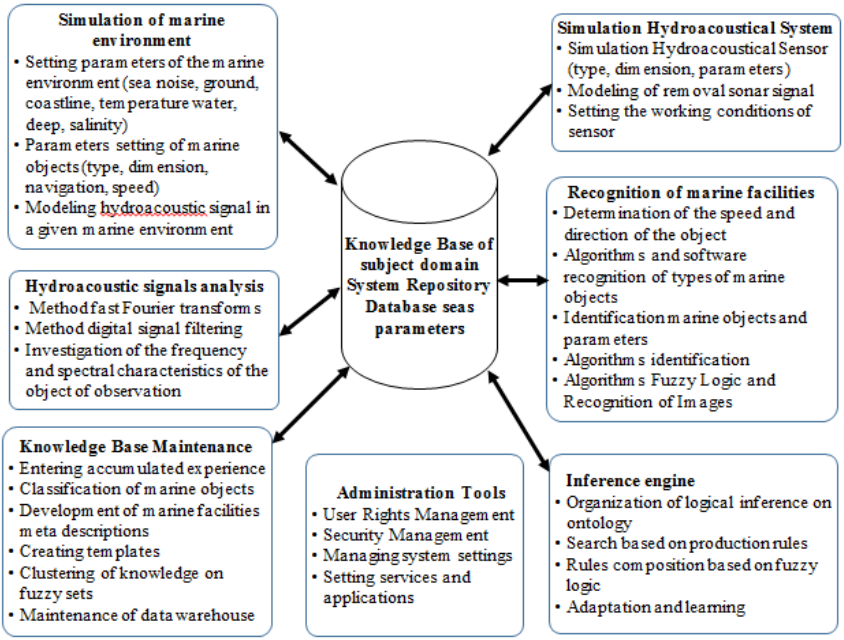


Fig. 1. Modeling system structure

3. Construction of the conceptual model of the subject domain ontology

Ontology model Ont(SD) of the intellectual system contains the basic concepts (entities - basic concepts of the subject domain), their attributes and describes the relations between them and can be represented as:

$$Ont(SD)_{IS} = < C^{(Ax)}, Ex^{(C)}, Rel^{(H)}, T^{(Q)}, Ax^{(s)}, Rul^{(S)} >$$

Where: $C^{(Ax)}$ - classes (Classes) - a finite set of basic concepts of hydroacoustic processes;

$Ex^{(C)}$- a finite set of Set Exemplars classes of ontology;

$Rel^{(H)}$- a finite Set of Relationships between classes and their types:

$T^{(Q)}$ - a finite Set of Attributes of each class, their data types and value fields;

$Ax^{(s)}$ – a finite Set of Axioms – define the basic concepts of SD, which are always true for it:

$Rul^{(s)}$ - a finite set of rules of logical conclusion.

Detailed consideration of the concepts of the subject domain and functional problems leads to the allocation of the following main classes:

1. Experiment – class $C^{(Exp)}$, defining characteristics of the conducted experiment – identification number, date of conduct, researcher identifier, service data.
2. Models_Hydroacoustic_processes – class $C^{(Mod)}$, describing various models, methods and algorithms for generation, selection and analysis of signals, as well as methods and algorithms for classification and identification of marine objects.
3. Hydroacoustic_objects – class $C^{(Obj)}$, consist from some subclass (marine objects, underwater objects and air objects) and describing properties objects.
4. Aquatorium is a class $C^{(Sea)}$ to describe the maritime area chosen for the experiment: the type of aquatorium, the name and refinement parameters for the type modeling algorithms (coordinates, depths, temperatures, salinity).
5. Hydroacoustic_System – class $C^{(Div)}$ that describes the coordinates, composition and types of hydroacoustic devices.
6. Experiment_scene – class $C^{(Sn)}$ to describe the simulation scene, taking into account the characteristics of the marine region, as well as the characteristics of the objects that participate in the experiment.
7. Modeling_Scenario – class $C^{(SnM)}$ to specify the sequence of individual stages of the simulation. Each step is a set of procedures "start-run-fix the result".
8. Model_estimation – class $C^{(Val)}$ for fixing the results of modeling and estimating the correctness of models.
9. Acoustic signal – class $C^{(Sig)}$ that defines the signal characteristics, including the date of the signal detection and the parameters of the locking device.
10. Waveguide - Class $C^{(Noise)}$ to describe the hydroacoustic interference affecting the propagation and distortion of the hydroacoustic signal, as well as the means of neutralizing interference.

In the ontology model, in addition to the main classes, subclasses representing instances of the corresponding classes are also included.

In accordance with the basic rules of the OWL language, a triplet called the RDF graph describes the class. In this graph, vertices are objects and objects, and as arcs are predicates [6]. From a mathematical point of view, the triplet is an instance of an element of a certain binary relation. The expression of the triplet asserts that a certain relationship indicated by the predicate connects objects marked as the subject and object in a particular in the triplet [7].

One of the tools for semantic modeling is the well-known ontology visual editor Protégé 5. - Designed by the University of Stanford [4]. Visual methods of designing ontologies help quickly and fully understand the structure of knowledge of the subject area, which is especially valuable for researchers working in the new subject area. Protégé 5 supports all phases of the ontology life cycle in accordance with ISO / IEC 15288: 2002 [9] requirements – from the development of a semantic network and the creation of a knowledge base on its basis, to the formation of user requests to these bases in order to obtain knowledge.

The main window of the Protégé 5 editor consists of tabs that represent the various tools for create model of knowledge such as: <Active Ontology>, <Entities>, <Classes>, <Data properties>, <Object properties>, <Individuals by class>, <DL Query>, <SPARDL Query>, <OntoGraf> and others (

4. Formation of the hierarchy of the main classes - taxonomy of the subject domain

Defining classes and creating their hierarchy (taxonomy) is key in the development of ontology SD. The taxonomy of classes is a tree of descriptive terms that have a hierarchical structure.

In Protege 5 editor creating classes $C^{(Ax)}$ occurs in the bookmark <Classes>. In the OWL classes are interpreted as a subset of individuals that are part of a defined class.

The peculiarity of designing in the environment Protege is that classes are considered as subclasses of the general ontology THING. According to the CamelCase notation for OWL [10] - all class names must begin with a capitalization and should not contain spaces. For securing the classifications, simply press the <Add subclass> button, In the window that opens, you must enter the name of the class. (Fig. 2.).

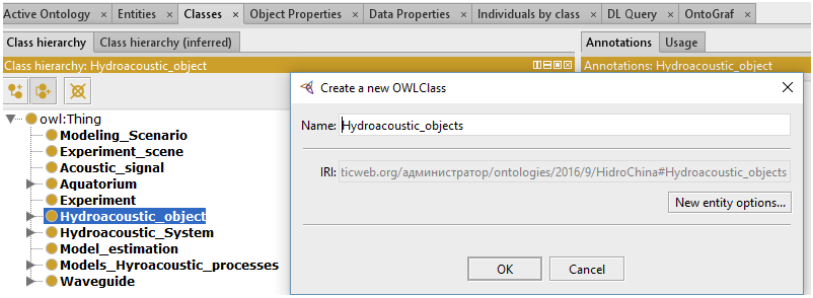


Fig. 2. An example of creating classes of the modeling system

By default, classes in OWL can intersect. In order to divide the classes, they need been disjointed. This ensures that an individual cannot be an instance of more than one class. To do this, in the <Classes> tab you need to define a class that should not intersect, then in the <Description> field, you need to click on the + side of the Disjoint With function and in the <Class hierarchy> window open the class that should not overlap with the specified class (Fig. 3.).

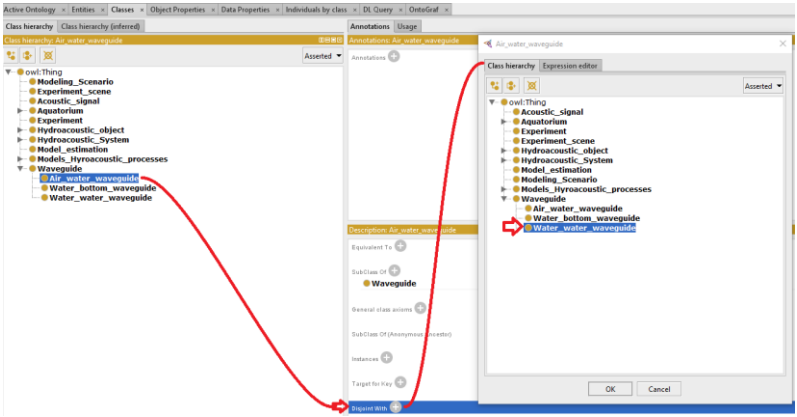


Fig. 3. Separation of classes into disjointed ones

OWL allows you to create annotations with various information (comments, creation date, author, links to resources, etc.) and metadata classes, properties, individuals and ontologies.

Tab <Classes>, <Object Properties>, and <Data Properties> are used to create annotations. Next, in the <Annotations> tab, click on the "+" and in the opened window, enter the desired comment (Fig. 4.).

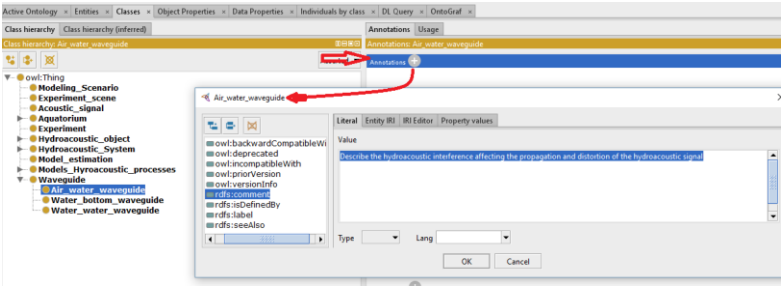


Fig. 4. Adding comments to the selected class

In Protégé editor each class defined by properties that describe the relationship between classes and divided into two types [12]:

- <ObjectProperty> - describes relations $Rel^{(H)}$ (types of relationships) that are established between particular classes of ontology.
- <DatatypeProperty> - describes the specific attributes $T^{(Q)}$ (characteristics) that define the class. For example, the speed of the marine object, its dimensions, technical parameters, etc.

5. Formation of a subset attributes of the domain subject

A subset of attributes $T^{(A)}$ describes the properties of classes $C^{(Ax)}$ and is used to enter specific values of instances classes - $Ex^{(C)}$.

The creation of attributes subset by Protege editor has performed using the <Datatype Properties> tab on toolbar of the Protégé editor 5. <DatatypeProperty> - associates an individual with a data type value of the RDF schema (defines the relationships between the individual and the data values) and describes the specific attributes $T^{(Q)}$ (characteristics) that define the class.

Next, the window containing the <OWL: DataPropertty> tab is displayed. When you click a button  toolbar, opens window <Create a new OWLDataPropertty>, in which you can enter a name property, such as <Cruising Speed> and click OK (Fig. 5.). After performing the corresponding actions provided by the process, a new type of property appears in the left frame.

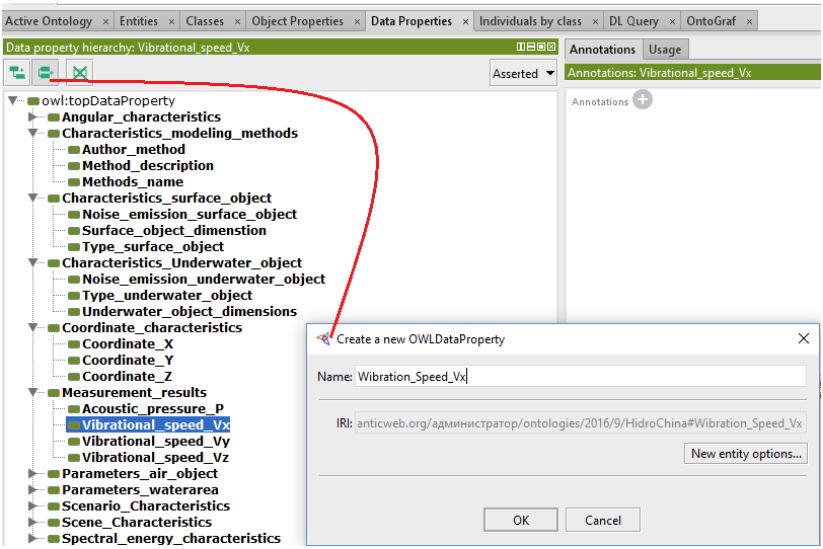


Fig. 5. Creation a new attribute in Protégé editor

Repeating this procedure can form a whole set of attributes for a SD. For a complex domain <OWL: DataPropertty> can be represent as a hierarchy structure. A set of properties for different instances of classes can be completely individual.

The predefined attributes has specified in the XML schema dictionary and can be represent in a variety of data formats: integers, floating point, lines, logical values, etc.

An example of the owl tag: DatatypeProperty <URI> in XML, which describes the class <Underwater_object>:

<owl: DatatypeProperty rdf: about = "http://www.semanticweb.org/ svr/ontologies/ -9# ="#Underwater_object_dimensions"/>

6. Formation of instances of classes for filling the knowledge base

Examples of classes – are specific objects SD (instance) that belong to a certain class - $C_i^{(Ax)}$. Every classes $C_i^{(Ax)}$ has describes subset of attributes $T_i^{(A)}$.

Instance creating process is executed through the tab <Individuals by class>, which is written by the type tag

<owl: NamedIndividual rdf: about = "....">

Process creating a new instance consists of the following steps:

1) Select tab <Individuals by class> in a toolbar Protégé editor. This opens a new window for create instance.
2) To creation a new instance of the selected class (Southern seas), you should click on the button  , as shown in Figure 6. This opens a new window <Creat a new OWLNamedIndividual>, which is intended to input the name of an instance. After performing the corresponding actions provided by the process, a new instance of the selected class appears in the left frame, for example, <Taiwan Strait>.
3)

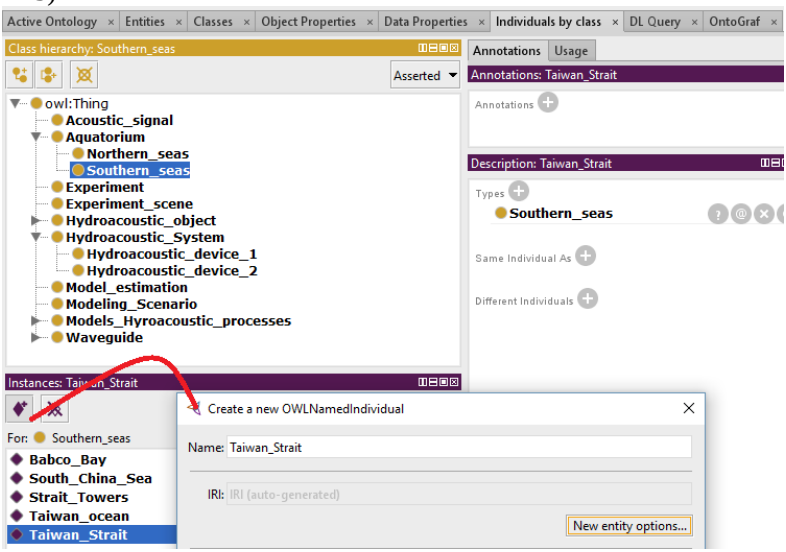


Fig. 6. Example of adding new instance

The procedure <Create a new OWL Named Individual> can also be used to fill the knowledge base with test samples. To do this, select the desired instance of the specified class (Southern_Seas), as shown in the figure 8, and in the field <Property assertions> with the instance name (Babco Bay), click on the icon <+> next to the function <Data Property assertions>. When clicked icon, a new <Data Property> window appears where is required in entity <Owl:TopDataProperty> has select the appropriate attribute, for example, <Maximum_depth> and enter its value

in the field <Value> right window (Fig. 7.). After clicking the OK button, its value is written in the ontology database

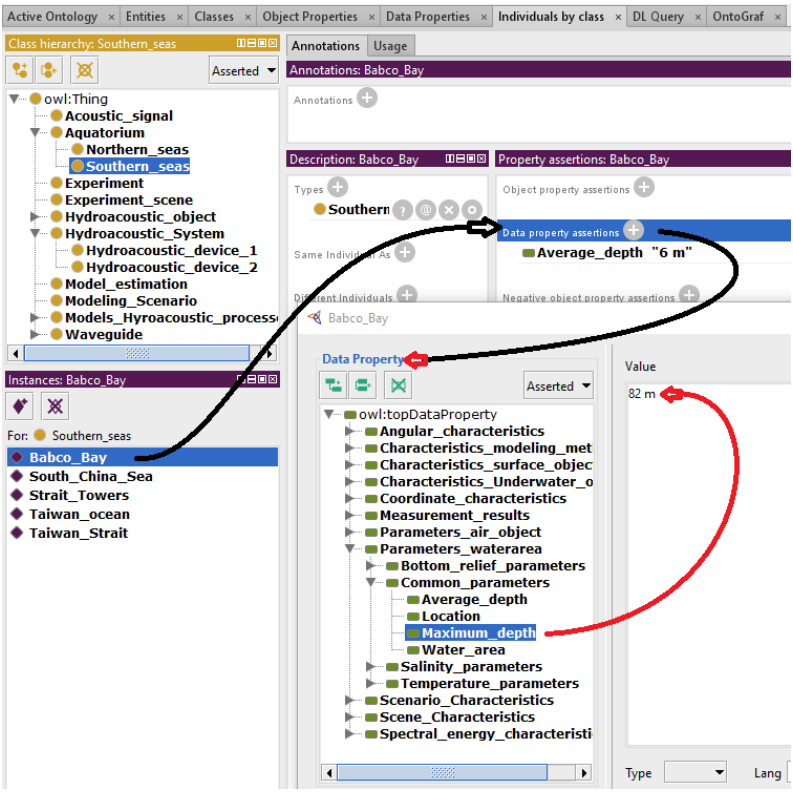


Fig. 7. Example of entering the value of a particular attribute

Because an instance of the class describes by some subset of attributes, this value-adding procedure must repeated for each attribute (Fig. 7.).

A set of properties for different instances of classes can be completely individual.

7. Defining and forming a subset of relationships and linking classes

In Protégé editor each class defined by properties that describe the relationship between classes. In Protégé, the Object properties describing

[12] by tab <ObjectProperty> - are relations between two classes or individuals.

With the help of the <ObjectProperty> tab, are performed the following actions:

1. Formation of a subset of the relationship $Rel^{(H)}$ between classes $C^{(Ax)}$ in a defined subject domain adding types of relationships to the list <owl: topObjectProperty>.
2. Formation of a subset Axioms $Ax^{(s)}$ in a defined subject domain - by binding certain class ontologies relations using the predicates from <owl: topObjectProperty>.
3. Formation of relationships between individuals using the predicates contained in <owl: topObjectProperty>

Forming a subset of relationships in a defined subject domain

To create a subset of relationships and adding it to owl: topObjectProperty, you must select the <ObjectProperty> tab in Protégé 5. and go to the owl: topObjectProperty line. The toolbar opens the tools to describe relationships. To adding a new type of relationship, you need to click a button  . In the window named <Create a new OWLObjectPropertty>, you must enter a name for the type of relation, for example, <has_communication_> (Fig. 8.) and press OK.

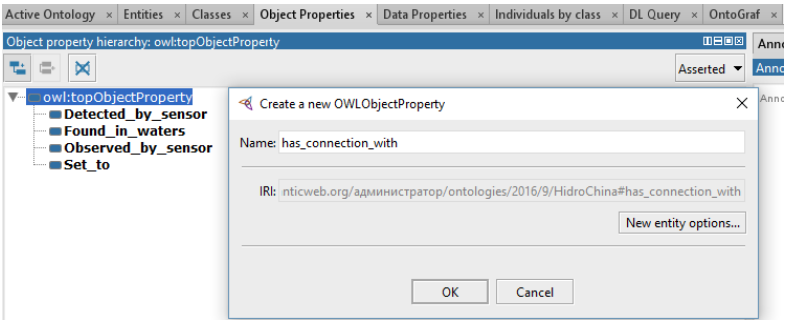


Fig. 8. An example of formation a new type of relationship

After performing the appropriate actions provided by the process, a new type of connection appears in the <owl: topObjectProperty> subset. Repeating this process for SD it is possible to form the whole subset of the types of zips for the selected SD.

Formation of a subset Axioms

Under the axiom are assertions that introduced into the ontology in the finished form, from which other statements can be deduced. Axioms bind two classes (the concept) with certain relations

The link between classes is described by the RDF graph

The created relationship between classes (Axioms) are described by the RDF-graph

<subject-predicate-object>

$Subject\ C_{si}^{(Ax)}$	$Rel_{ij}^{(H)}$	$Object\ C_{oj}^{(Ax)}$
($Class_i$)	Predicate	($Class_j$)

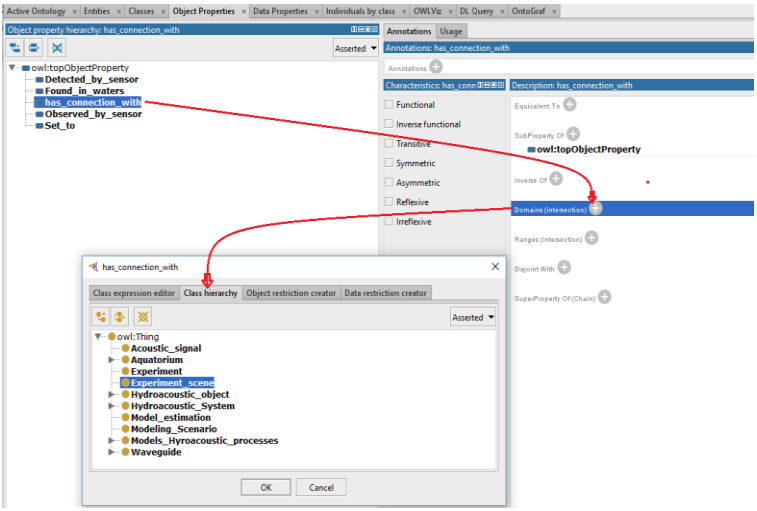
In the editor Protégé 5, two functions used to define the axiom of subject domain:

- Domains (intersection) – asserts that the subjects of such property statements must belong to the class extension of the indicated class description [].
- Ranges (intersection) – asserts that the values of this property must belong to the class extension of the class description or to data values in the specified data range.

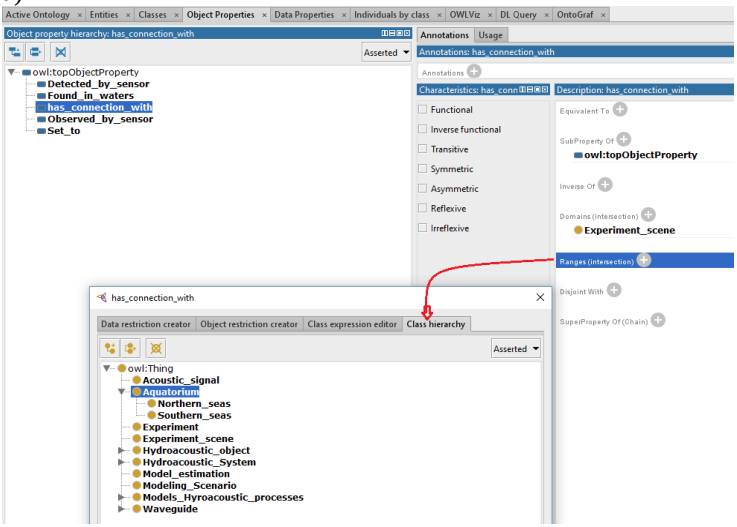
To create a subset of **Axioms** you must select the <ObjectProperty> tab in toolbars Protégé 5 and go to the owl: topObjectProperty line. In the <owl: topObjectProperty> subset select the desired relationship type. which binds two classes, for example class <Experiment_scene> connected with class <Aquatorium> relationship <has_connection_with>.

To do this, on the <ObjectProperty> tab in the <Description> window, you need to click on the "+" next to the function <Domains (intersection)>. When you click "+", a separate window opens where you want to select the <Class hierarchy> tab (Fig. 9 a) and define the class (<Experiment_scene>) to which the selected property is given.

When to selects second class you need to click on the "+" next to the function <Ranges (intersection)> in the <Description> window. Further, in the opens window you must select the class (<Aquatorium>) and click OK (Fig. 9 b). After completing the corresponding actions, a new relationship between the classes has created and stored in the model ontology of Subject Domain.



a)



b)

Fig. 9. Example created relationship between classes

Formation of relationships between individuals

For formation of relationships between two individuals in Protégé 5 you are need to go in <Entities> tab of toolbar. In the open window <Individuals by type> there is class hierarchy that has instances. Next, in

the class hierarchy (for example, <Helicopter>) you must select an instance (subject) and open it (<BoeigCH-47>). After that, two windows open <Description> and <Property Assertions >. For formation of relationships between selected individual you need to click on the "+" next to the function <Object Property Assertions>. After that opens a new form to enter two fields <Object property name> and <Individual name> (Fig. 10.). For example, individual <BoeigCH-47> characterized by a high level of acoustic noise the limiting values of which are recorded in the individual <Big_acoustic_noice>. In the field <Object property name> you need enter the relationship <Has_acoustic_noice> and in the field <Individual name> register individual <Big_acoustic_noice> and them to click OK.

After completing the corresponding actions, a new relationship between the individuals has created and stored in the model ontology of Subject Domain.

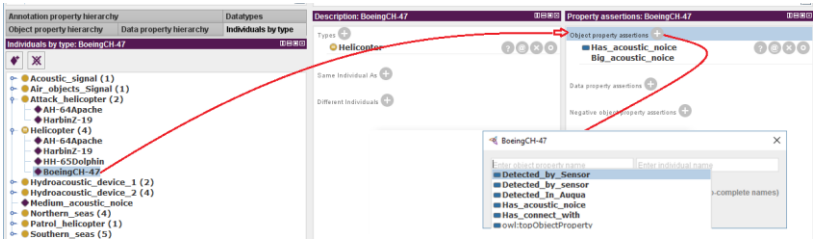


Fig. 10. Process creation of relationship between individuals

The class hierarchy built by means of Protégé editor has shown in Figure 11.

Subject domain ontology includes 34 classes, each of them has from 5 to 15 attributes; nearly 30 axioms: 23 associative relations, 15 “is-are” relations, nearly 70 heredity relations and 50 “class-data” relations are described; standard types of attributes values and limits for values of attributes are highlighted.

For example, Acoustic signal – class $C^{(Sig)}$ has the following attributes:

- Frequency of the signal
- Phase
- Amplitude
- Interference level
- Spectral-energy characteristics
- Vector-phase characteristics to identify the object

- Coefficient of total attenuation (pressure signals, Speed along the axes - V_x, V_y, V_z)
- Location coordinates
- Distance to the object,
- Speed of propagation,
- Position of the object relative to the receiver,
- Pressure at the depth of the receiver,
- The results of modulation of the noise spectra of objects
- Coefficient of noise emission.

Thus, Ontology represents the description of subject domain in the terms of knowledge processing theory. It allows describe the subject domain entities in the terms of ontology classes, gives classes objects – called "individuals". After all classes and their individuals are described and all relations on them are set it is possible to use algorithms of descriptive logic for ontology information processing as well as build the ontology graph. Ontology ben used for more adequate description of entities with which hydroacoustics system operates.

8. A comparative analysis of the expressive capabilities of SPARQL and DL-Query

The concept of discrete logic in the Protégé environment is realized through the use of a logical output machine - Reasoner. One of the most famous logical machines for finding hidden knowledge in ontology is Reasoner HermiT 1.3.8.413, which is based on the syntax of Manchester OWL. This syntax is a subset of the OWL DL and is based on gathering all information about a particular class, property, or individuality into a single frame construct construct. Reasoned Hermit 1.3.8.413 comes as a plug-in with the standard distribution of Protégé Desktop 5.

DL, implemented in Reasoner HermiT 1.3.8.413, operates with the Trim model primitives in the formation of DL Query:

- Individuals - which define individual entities of the subject area (Class name, Data Propertty name, Object Property name, individuals), with the participation of which a request is made to the BR;
- Concepts - defines the action (Conjunction, Disjunction, Class Addition, etc.) that must be performed on selected entities (Class name, Data Propertty name, Object Property name) when manipulating data;
- Roles - defines the type of relationship between the entities of the knowledge base (some, min, max, only, self, exactly, value).

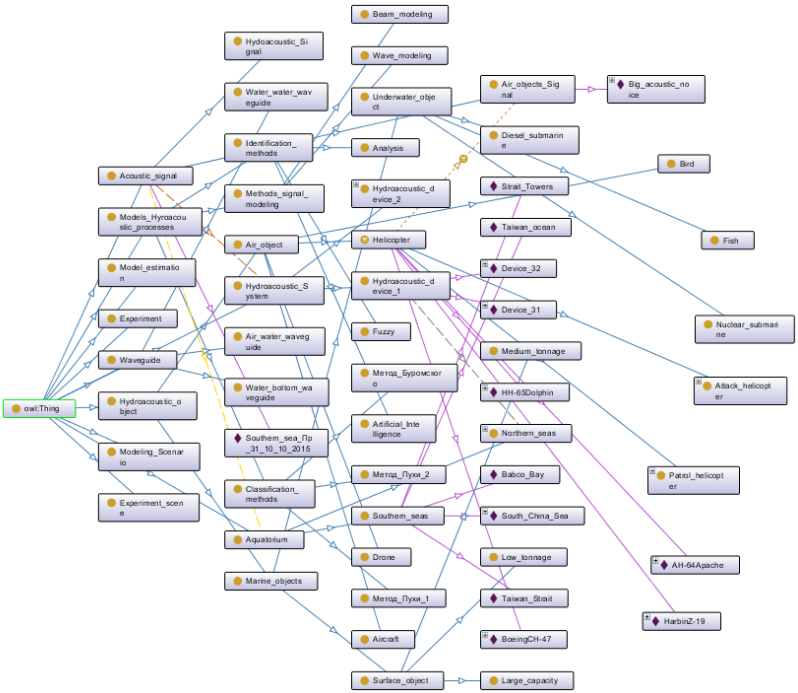


Fig. 11. Graph of ontology model using Protégé

Table shows the main logical functions of the Manchester OWL syntax in the formation of the DL Query [14].

Table

DL function	Linguistic designation	OWL Constructor	Explanation
C ⋂ D	***C* AND *D***	intersectionOf	Conjunction – individuals in *C* and *D*
C ⋃ D	***C* OR *D***	unionOf or *D*	Disjunction – individuals in *C*
¬ C	**NOT *C***	complementOf	Class supplements – individuals not in *C*
{a} ⋃ {b}..	**{A B ...}**	oneOf	One instance of the class
∃ R C	***R* SOME *C***	someValuesFrom	Choosing a Separate

			Relationship with C
∀ R.C	**R ONLY C**	allValuesFrom	Select all
≥ N R	**R MIN 3**	minCardinality	
≤ N R	**R MAX 3**	maxCardinality	
= N R	R EXACTLY 3	cardinality	Select an instance that matches the exact value of the property Data Propertty name
∃ R {a}	R VALUE a	hasValue	Select an instance with a specified value of the property Data Propertty name

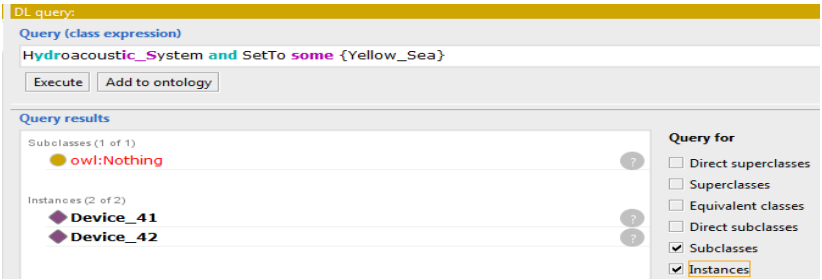
The response to requests for built ontology is extremely important as it used as a tool by which datas users can interact with ontologies and data. In this context, it is extremely important to use the language of the request to create an ontology - **expressive capabilities**.

A comparative analysis of the expressive capabilities of SPARQL and DL-Query was performed on the example of searching for knowledge in the ontology of "Modeling complex for working out algorithms and software for detection, classification and determination of parameters of motion of marine objects" request type:

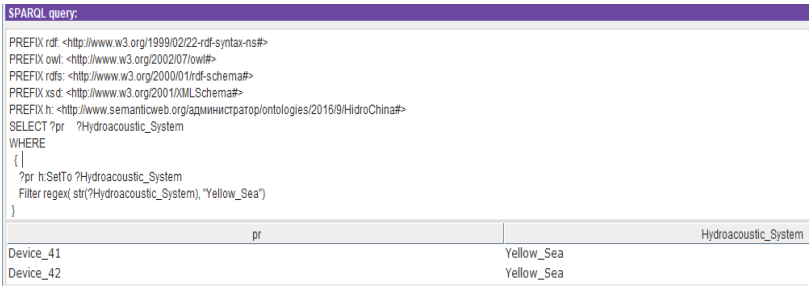
What are the Hydroacoustic_System installed in Yellow_Sea?

DL -Query allowed to detect two objects that are installed in Yellow_Sea (Figure 12a) DL query).

A similar SPARQL query also detected objects that were installed in Yellow_Sea (Figure 7 in) SPARQL query).



a) DL query



b) SPARQL query

Fig. 12. Comparison of results of the logical conclusion of the languages of DL and SPARQL queries

The SPARQL is the query language for the RDF data that uses the communication protocol to connect to RDF repositories. SPARQL is an analogue of the SQL language for relational databases. The SPARQL query is a collection of hobby templates that compare RDF data. The results of requests for SPARQL requests can be returned or reproduced in different formats:

• XML SPARQL defines an XML dictionary for returning result tables.

• JSON JSON "port" XML dictionary, especially useful for web applications.

• RDF Some SPARQL results points cause RDF replies, which, in turn, can be batchwise in different ways (RDF / XML, N-Triples, Turtle, etc.)

• HTML. When using an interactive form for working with SPARQL queries. Often implemented through the application of converting XSL to XML results

One of the significant benefits of the SPARQL query language is the ability to use the filtering method. For example, you need to find a Hydroacoustic_object Level Noise_emission greater than 60 db:

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
SELECT ?Hydroacoustic_object ?Noise_emission
 WHERE {
   ? Hydroacoustic_object: Helicopter ;
         rdfs:label ? Hydroacoustic_object;
         prop: Noise_emissionEstimate ? Noise_emission.
   FILTER (?Noise_emission  > 60 db) .
}
```

Another advantage of the SPARQL query language is the ability to locate analogues by looking at them in the WWW space.

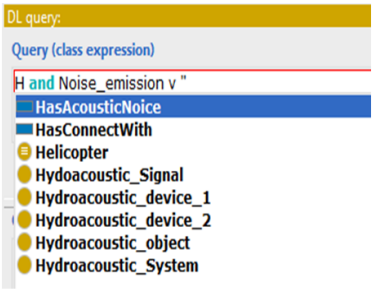
The disadvantage of using the SPARQL query language is the need for the end-user to know the SPARQL syntax and carefully write the names of entities and predicates. If the individual knowledge of the SPARQL language of the user is not proper, it is very difficult to correctly formulate requests for ontology. Because syntax errors lead to non-execution of the request and, in fact, to the complete impossibility of using ontology.

DL-Query in the protégé tool environment 5.2 implemented within the HermiT reasoner. DL-Query is a full version of OWL 2 Direct Semantics as the World Wide Web Consortium (W3C) [11]. HermiT also supports rules for DL-safe SWRL and description graphs, which allows for the precise description of arbitrarily-related ontology structures (many domain ontologies).

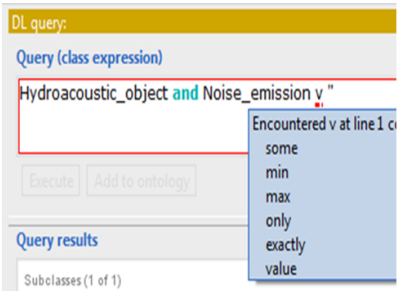
Reasoner HermiT perceives an ontology as a TBox set - a description of the hierarchy of the classes <class hierarchy> (concepts of the domain and their instances), and ABox - sets of properties that describe the classes and their instances (using the <owl: topDataProperty> and <owl: topDataProperty> built ontology).

Moreover, Reasoner HermiT supports Hypertableau-based algorithm, implemented on Java.

Consequently, when forming a specific DL request, the user can use the popup menu when writing the query object (Figure 13). That significantly facilitates the formation of a correct query. In addition, when choosing a DL function over triplets, it is also possible to use the built-in DL function menu.



Built-in DL menu to forming entities



Built-in DL function menu to forming predicates

Figure 13. Examples of using the popup menu when creating queries

DL query, SWRL, and SPARQL allow you to perform very complex queries to the ontology knowledge base that use different filtration methods to generate queries.

However, the DL query characterized by a higher level of expression because it, besides traditional SELECT query capabilities, also uses the Manchester OWL syntax logic device in the formation of DL Query [14]. This allows you to use a popup menu that reads the ontology of the domain and simplifies the formation of the DL request when describing the entities in the triplet. In addition, when writing a DL function, you can also use the popup menu.

These languages in the ontology analysis use n (dimensional) - triplets in the form of a query, which allows you to formulate very complex queries and to identify implicit tie between classes and instances of classes.

References

1. Semantic technology, available at: https://en.wikipedia.org/wiki/Semantic_technology
2. Knowledge representation languages: available at: http://www.cs.man.ac.uk/~stevensr/onto/node14.html
3. Open Knowledge Base Connectivity: :available at http://www.ai.sri.com/~okbc/
4. Protégé - Free, open-source ontology editor and framework for building intelligent systems: available at http://protege.stanford.edu

5. RDF Model Theory: available at: https://www.w3.org/TR/2002/WD-rdf-mt-20020429/
6. Semantic Web Technologies (2013): available at: https://open.hpi.de/course/semanticweb
7. Senchenko·V.R. Koval O.V. Globa·L.S. Novogrudska R.L. Intelligent modeling system based on cloud-technology, Prosedings of International Conference Radio Electronics & Info Communications" (UkrMiCo), DOI:10.1109/UkrMiC.2016.7739646, IEEE Digital Library, 2016
8. Web Ontology Language:available at: http://www.w3.org/TR/owl-features/
9. Camel case notation: available at: https://en.wikipedia.org/wiki/Camel_case
10. OWL 2 Web Ontology Language Document Overview (Second Edition) W3C Recommendation 11 December 2012: available at: http://www.w3.org/TR/owl2-overview/
11. HermiT OWL Reasoner: http://www.hermit-reasoner.com/
12. FaCT++ reasoned: http://owl.cs.manchester.ac.uk/tools/fact/
13. Platform for developing CBR applications: http://gaia.fdi.ucm.es/research/colibri/jcolibri
14. The Manchester OWL Syntax, available at: http://ceur-ws.org/Vol-216/submission_9.pdf

УДОСКОНАЛЕНА СТРУКТУРА СИСТЕМИ ВИЯВЛЕННЯ ВТОРГНЕНЬ В ІНФОРМАЦІЙНО-ТЕЛЕКОМУНІКАЦІЙНІ МЕРЕЖІ ТА СИСТЕМИ СПЕЦІАЛЬНОГО ПРИЗНАЧЕННЯ

І.Ю. Субач[1], В.В. Фесьоха[2],
[1]Інститут спеціального зв'язку та захисту інформації НТУУ "КПІ ім. Ігоря Сікорського"
[2]Військовий інститут телекомунікацій та інформатизації

На сьогоднішній день побудова ефективної інфраструктури безпеки інформаційно-телекомунікаційних систем та мереж (ІТСМ) не можлива без застосування механізмів захисту, які відповідають сучасним тенденціям ведення кібернетичної боротьби. Одним з основних оборонних засобів ІТСМ від інформаційно-руйнівних впливів (втручань) у вигляді кібернетичних вторгнень та/або атак традиційно виступають системи виявлення та/або запобігання вторгненням (СВВ), основна задача яких зводиться до оперативного їх виявлення та, в ідеальному випадку, ініціювання сценарію припинення факту компрометації об'єкта, що атакується.

Оскільки природа кібернетичних атак є різноманітною, то ефективність СВВ, з точки зору адаптації до них, в певній мірі залежить не тільки від покладених у них методів та моделей виявлення атак, але й від реалізованого у них підходу до функціональної та фізичної архітектурної побудови.

Аналіз структури існуючих СВВ, які є відкритими на основі загальнодоступних ліцензій (*Suricata, Bro, OSSEC, Prelude ma* ін.) показує, що основними елементами, які входять до їхньої структури є:

- підсистема збору інформації про об'єкт, який захищається;
- підсистема аналізу (виявлення) атак та вторгнень на об'єкт;
- підсистема спостереження за станом об'єкта.

Головним недоліком такої архітектури є відсутність підсистеми інформаційної підтримки оператора з безпеки, що в свою чергу, не дозволяє йому здійснювати оперативний (*Online Analytical Processing, OLAP*) та інтелектуальний (*Data Mining*) аналіз інформації про стан системи (мережі) з метою виявлення неочевидних, об'єктивних та корисних на практиці знань для подальшого прийняття рішення щодо ідентифікації несанкціонованого кібернетичного втручання в їхню роботу. До того ж розглянуті системи є недостатньо повними з точки зору охоплення

рівнів функціонування, що не дозволяє широко використовувати їх у складі ІТСМ.

У зв'язку з цим виникає актуальна задача удосконалення розглянутої структури СВВ шляхом введення до неї підсистеми підтримки прийняття рішень (ППРП) адміністратора з інформаційної безпеки та зміни способу її застосування в інформаційно-телекомунікаційних системах спеціального призначення.

У доповіді розглянуто удосконалену структуру СВВ у складі ІТСМ та реалізацію на її основі багатоешелонованої методики виявлення кібернетичних атак, яка відповідає новітній концепції *Next generation intrusion prevention systems (NGIPS)*, яка у теперішній час набуває значного поширення.

Удосконалення структури СВВ полягає у додаванні до неї ППРП, яка дозволяє знаходити неочевидні, об'єктивні та корисні на практиці знання про кібернетичні атаки, які не виявляються традиційними методами сигнатурного аналізу, а також здійснювати гнучкий багатовимірний оперативний та інтелектуальний аналіз інформації про досліджуваний об'єкт з метою автоматизованого донавчання системи (добування нових знань як про процес функціонування ІТСМ, так і про нові типи кібернетичних атак, з подальшим накопиченням їх у базі знань СВВ). Такий підхід дозволяє доповнювати (уточнювати) суб'єктивну точку зору експертів, яка має конкретне практичне застосування.

В основу багатоешелонованої методики застосування СВВ покладено інформаційну технологію двохетапного аналізу інформації про об'єкт, що захищається. Так, на першому етапі здійснюється дослідження параметрів стану ІТСМ спеціального призначення шляхом застосування методів виявлення зловживань (чітка синтаксична відповідність між правилом у базі сигнатур та кібернетичним втручанням, що відбувається), які, у свою чергу, характеризуються високою продуктивністю та майже повною відсутністю хибних спрацьовувань. Другий етап передбачає застосування методів виявлення аномалій в роботі ІТСМ (виявлення невідомих кібернетичних атак на основі знаходження набору ознак, який не відповідає очікуваній поведінці досліджуваного об'єкту).

У даній методиці, виявлення ознак кібернетичних атак на першому етапі аналізу інформації про стан ІТСМ не потребує ініціювання другого етапу, і навпаки, пошук ознак стану ІТСМ, які не відповідають очікуваній поведінці з метою ідентифікації аномалій, здійснюється у випадку позитивного рішення на першому етапі.

Таким чином, запропонована удосконалена структура СВВ є підґрунтям для підвищення ефективності діяльності оператора з безпеки ІТСМ спеціального призначення за показниками оперативності та обґрунтованості прийняття рішень щодо виявлення та запобігання кібернетичним атакам в режимі часу, близькому до реального, а багатоешеленована методика застосування СВВ дозволяє зробити ще один новий крок у бік здійснення їхньої адаптованості до невідомих кібернетичних атак.

EXPLORING DIGITAL FORENSICS TOOLS IN CYBORG HAWK LINUX

Tmienova Nataliia, Ilarionov Oleg, Ilarionova Nina
Faculty of Information Technology
Taras Shevchenko National University of Kyiv,
Kyiv, Ukraine
tmyenovox@gmail.com, oilarionov@gmail.com,
ilarionovanm@gmail.com

Computer forensics (software and technical expertise) belongs to the category of engineering and technical expertise. It is an important element in a number of computer expertises, because it allows to build a holistic system of evidence comprehensively. The importance of computer forensics is explained by the increased role of the computers in the modern world. A huge number of offenses and crimes is committed precisely with the help of computer technologies. The computer forensics and expertise of computer equipment is especially relevant in criminal and civil cases. Expertise of computers, hardware, software, databases due to the continuous improvement of computer equipment and software is one of the most complex types of research.

The community of free software developers is constantly creating assemblies of utilities designed for software and technical expertise. The most popular is the Kali Linux collection, whereas Cyborg Hawk Linux is undeservedly ignored.

The purpose of our research is to describe the capabilities of the Cyborg Hawk Linux tools.

1 Introduction

Development of information technologies, penetration of computer technology advancements into applied and scientific sphere and into everyday human life has its drawbacks, unfortunately. There are many intruders who use these achievements for mercenary, criminal purposes. In this regard, there is a need to transform special knowledge from the field of computer information into the field of forensic science to uncover and investigate crimes that relate to computer technologies. Computer forensic allows obtaining the most reliable information concerning computer crimes. This type of research is widely used in consideration of cases in civil and criminal legal proceedings and is one of the most relevant and demanded.

Computer forensics covers a broad range of activities associated with identifying, extracting, and considering evidences from digital media. It can be defined as the use of scientifically derived and proven methods toward the preservation, collection, validation, identification, analysis, interpretation, documentation, and presentation of digital evidence [1] derived from volatile and non-volatile media storage [2]. Various hardware, software, and information objects are objects of computer forensics.

Computer forensics can be divided into the following types: hardware, software, network, and information forensics. The following methods are used in the process of computer forensics:

- ✓ method of software research;
- ✓ method of hardware research;
- ✓ method of information research.

Carrying out of software and technical expertise is necessary in cases when a crime or an offense was implemented with using computer facilities or information data, and when special knowledge in the field of computer technology is required to establish traces of crime and other forensically significant information. In particular, software and technical expertise provides the solution of the following expert tasks:

- ✓ identification of properties, qualities, status and features of the using of technical computer systems;
- ✓ establishing the development and using features of software products;
- ✓ establishing the facts of the equipment use during the documents creating or committing other actions related to the crime;
- ✓ access to information on attached devices;
- ✓ research information created by a user or a program for the implementation of information processes;
- ✓ establishing the features of the functioning of the computer facilities which implement network information technology.

Computer forensics is used to find out the digital evidence using different tools. It is quite difficult and complex process. Digital investigations take place in three main phases. In first phase, the investigator takes images of digital device and copies these images from the target device to some other device for in-depth analysis. In second phase, which is called analysis, the investigator identifies the digital evidence using different types of techniques such as recovering the deleted files, obtaining information of user accounts, identifying information about the attached devices like USB, CD/DVD drives,

external hard disks and so on. The third phase is called reporting in which the investigator reconstructs the actual scenario based on the sequence of activities happened on the target system [3].

Digital forensic analysis is divided into two main categories. The first category is the static forensic analysis. During this analysis, all the target devices that are required in the analysis are shutdown. The second category is live analysis. During this type of analysis, the system stays in the boot mode [4] to acquire pertinent information from the physical memory content.

Live analysis aims at gathering evidence from systems using different operations and techniques related to primary memory content. Live forensic is the most challenging kind of digital forensic investigations. To perform the live forensics, it is vital to understand the basic techniques and tools used in digital forensics. The investigator needs to acquire the complete image of a computer usage history as well as the current state through live forensic analysis tools. Though static analysis is kind of a developed part of digital forensics, but other techniques related to live analysis need to be developed to mitigate its weaknesses [3].

Classifications of computer forensics tools include open source, proprietary, hardware, software, special purpose and general purpose [5]. Each tool has its own advantages and disadvantages. The choice of forensics tools depends on the nature of the studying, the obtained results, the requirements for safety and economic efficiency of the tool.

Brian Career [6] reports that open source tools are as effective and reliable as proprietary tools. Manson and his team [7] compared one open source tool and two commercial tools. They found that all three tools produced the same results with different degree of difficulty.

2 Description of the most popular tools designed for carrying out software and technical expertise

The community of free software developers is constantly creating assemblies of utilities designed for software and technical expertise. A comprehensive review of the top twenty open source free computer forensics investigation tools can be found in [8]. For a list of proprietary computer forensics tools see [9] and [10].

The most popular assemblies of utilities intended for carrying out software and technical expertise are:

- Kali Linux [11] – Kali Linux is an open source project that is maintained and funded by Offensive Security, a provider of

world-class information security training and penetration testing services.

- CAINE [12] (Computer Aided Investigative Environment) – CAINE is the Linux distro created for digital forensics. It offers an environment to integrate existing software tools as software modules in a user friendly manner. This tool is open source.
- DEFT [13] (Digital Evidence & Forensic Toolkit) – The Linux distribution DEFT is made up of a GNU / Linux and DART (Digital Advanced Response Toolkit), suite dedicated to digital forensics and intelligence activities.
- PHLAK [14] (Professional Hacker's Linux Assault Kit) – PHLAK is a modular LiveCD Linux distribution with a focus on pen-testing, forensics, and network analysis. It includes two lightweight GUIs (XFCE4 and Fluxbox) and loads of tools, including crackers, sniffers, MITM utilities, and data recovery and duplication utilities.
- Cyborg Hawk Linux [15] – Cyborg Hawk Linux is a Ubuntu based Linux Hacking Distro also known as a Pentesting Linux Distro it is developed and designed for ethical hackers and penetration testers. Cyborg Hawk Distro can be used for network security and assessment and also for digital forensics. It also has various tools suited to the testing of Mobile Security and Wireless infrastructure.
- BackTrack 5 R3 [16] – BackTrack is intended for all audiences from the most savvy security professionals to early newcomers to the information security field. BackTrack promotes a quick and easy way to find and update the largest database of security tools collection to-date.
- Parrot Security OS [17] – Parrot Security OS is a cloud friendly operating system designed for Pentesting, Computer Forensic, Reverse engineering, Hacking, Cloud pentesting, privacy/anonimity and cryptography. Based on Debian and developed by Frozenbox network.
- BackBox Linux [18] – BackBox is a Linux distribution based on Ubuntu. It has been developed to perform penetration tests and security assessments. Designed to be fast, easy to use and provide a minimal yet complete desktop environment, thanks to its own software repositories, always being updated to the latest stable version of the most used and best known ethical hacking tools.

The using of assembly, rather than individual software tools, can improve reliability, safety and performance.

The most popular compilation is the Kali Linux, which contains about 300 utilities, whereas Cyborg Hawk Linux [15], which contains more than 800 tools, is undeservedly ignored.

The purpose of our research is to describe the capabilities of the Cyborg Hawk Linux tools.

3 Our Virtual Machine Platform

Cyborg Hawk is a Linux based operating system that comes with a rich repository of security and forensics tools. The computer forensics tools are grouped into several categories. We use the forensics tools within the Cyborg Hawk.

VMware Workstation is a hypervisor that runs on 64-bit computers [19]. It enables us to set up multiple virtual machines and network them together. Each virtual machine can execute on different distribution of Linux operating system. VMware Workstation is proprietary software but we used the trail version for free. Below are the steps for setting up the platform for our experiment.

1. Install VMware Workstation on a machine;
2. Create a virtual machine on the VMware workstation;
3. Install Cyborg Hawk Linux on the virtual machine;
4. Launch Cyborg Hawk Linux from the virtual machine;
5. From the list, select forensics and then select a tool.

4 Description of the Cyborg Hawk Linux tools

The Cyborg Hawk Linux disk image was investigated on a VMware Workstation virtual machine running on a 64-bit computer.

There are 15 classes in the Cyborg Hawk Linux software analysis toolkit, each of which is divided into categories and subcategories that contain different number of utilities (table 1).

Class	Category	Number of tools
1. Collection of information	Network Research	68
	Proxy	3
	VPN analysis	3
	Cataloging the Web	63
1. Vulnerability	Network	14

Class	Category	Number of tools
assessment	Web applications	68
2. Exploits	Browser Capture Platforms	1
	Database	5
	Network	23
	Social Engineering	2
	Web Attack	19
4. Increasing of privileges	Password attacks	66
	Listening to channels (sniffing)	29
	Substitution (spoofing)	31
5. Access support		25
6. Reports	Work with evidence	7
	Radio capture of data from the screen	2
	Software documentation	2
7. Reverse engineering	Debuggers	4
	Disassembly	6
	Exploit development tools	4
	Tools for modeling (RE Tools)	15
8. Reliability testing (stress tests)	DOC	21
	Phasers	22
	Stress testing of wireless networks	2
9. Forensic	Acquisition	21
	Cryptography	4
	Data recovery	42
	Digital anti-forensic	1
	Digital forensic	14
	Forensic evaluation tools	40
	Forensic suite	5
	Network investigation	2
	Secure wipe	4
	Steganography	8
10. Tools for wireless attacks	Bluetooth	25
	Miscellaneous tools	8
	The radio channel monitoring	10

Class	Category	Number of tools
	WiFi	36
11. RFID / NFC tools		43
12. Hacking of hardware		4
13. Analysis of VOIP		24
14. Mobile Security	Tools for development	3
	Forensic of devices	9
	Penetration testing	9
	Reverse Engineering	20
	Wireless analyzers	6
15. Analysis of malicious software		5

Table 1. Classes and tools category of open source software and technical expertise in Cyborg Hawk Linux

Several utilities have been selected in each category. For each of them we investigated its purpose, sequence and results of work on our virtual machine. One of the conclusions of the studying is that many utilities perform several functions and thus they belong to different classes and categories in the collection. Therefore, the number of original programs is much smaller than were stated by the developers of the assembly. In addition, a significant limitation in using of the assembly is that it is only designed to work with 64-bit processors.

5. Forensics Tools Experiment

There are several categories of computer forensic tools in the disk image of Cyborg Hawk Linux v1. Some categories have several tools. In the following subsections we will study the tools for Forensics.

5.1 Acquisition

Twenty one tools of this category are divided into 10 groups.

Let's consider the basic packages of tools.

AFF Package (*affcat, affconvert*) or the Advanced Forensics Format (AFF). AFF was created as an open and extensible file format for storing disk images and associated metadata. The goal was to create a disk imaging format that would not block users into their proprietary format,

which can limit its analysis. The open standard allows researchers use their preferred tools for solving crimes, collecting information and resolving security incidents quickly and efficiently. The format was implemented in AFFLIB which was distributed with an open source license.

Img package (*img_cat, img_stat*) outputs the contents of an image file. Image files that are not raw will have embedded data and metadata. *Img_cat* will output only the data. This allows you to convert an embedded format to raw or to calculate the MD5 hash of the data by piping the output to the appropriate tool.

Img package displays the contents of the image file. Image files that are not raw will have built-in data and metadata. Img_cat will return only data. This allows converting the built-in format to raw or calculating the MD5 hash of the data by submitting the output to the appropriate tool.

TSK KIT (*tsk_comparedir, tsk_gettimes, tsk_loaddb, tsk_recover*) – compare the contents of a directory with the contents of an image or local device. Sleuth Kit (TSK) allows exploring the compromised file system of a computer. TSK is a collection of UNIX command-line tools that can analyze NTFS, FAT, FFS, EXT2FS, and EXT3FS file systems. TASK reads and processes the file system structures independently, so the file system of the operating system does not need support.

5.2 Cryptography

There are 4 tools in this category (*Luks-Ops, TrueCrack, TrueCrypt, Tcpcryptd*).

TrueCrypt is a program for installing and maintaining a drive immediately. Immediate encryption means that the data is automatically encrypted or decrypted immediately before downloading or saving it without user intervention. Any data stored on the encrypted volume can be read (decrypted) without using the correct password or the correct encryption key. The *TrueCrypt* volume before decryption is nothing more than a series of random numbers.

5.3 Data recovery

The tools of this category are divided into four groups (*Carving Tools, Password Forensics, PDF Forensics, Ram Forensics*).

Carving Tools contains 20 programs that specialize in recovering files, missing disk partitions, etc.

Password Forensics contains 3 programs (*chntpw, md5deep, rahash2*), which allow to delete passwords to Windows, calculate and compare MD5 hash-functions and checksums. The main set of tools for password security is in the category of the fourth class (table 1). It contains 66 tools.

PDF Forensics: This tool will parse a PDF document to identify the fundamental elements used in the analyzed file. It will not render a PDF document.

Ram Forensics: *volafox* and *volatility* are the tools for working with memory dumps of RAM. Supports memory dumps from all major operating systems.

5.4 Digital anti-forensics

Chkrootkit is the only tool in this category. *Chkrootkit* is a scanner that monitors the presence of rootkits on the local system by some search attributes. The program has several modules to search for rootkits and other unsafe objects. As expected, rootkits in our virtual machine were not detected.

5.5 Digital forensics

There are 14 tools in this category (*autopsy, binwalk, bulk_extractor, chkrootkit, dc3dd, dcfldd, extundelete, foremost, fsstat, galleta, tsk_comparedir, tsk_gettimes, tsk_loaddb, tsk_recover*).

5.6 Forensics evaluation tools

There are 40 tools in this category (*affcompare, affcopy, affcrypto, affdiskprint, affinfo, affsign, affstats, affuse, affverify, affxml, autopsy, binwalk, blkcalc, blkcat, blkstat, bulk_extractor, cuckoo, ffind, fls, foremost, galleta, hfind, icat-sleuthkit, ifind, ils-sleuthkit, istat, jcat, mactime-sleuthkit, missidentify, mmcat, pdgmail, readpst, reglookup, reglookup-timeline, reglookup-recover, SIGFIND, sorter, srch-strings, tsk_recover, vinetto*).

AFF Package was considered in 5.1.

Libewf package (*ewfacquire, ewfacquirestream, ewfexport, ewfinfo, ewfverify*) writes data of data carriers from devices and files into EWF files. *Ewfacquire* can be used to create disk images in the EWF format. It includes several message digests including MD5 and SHA1. To create an image of */dev/sdb1* and logging data to */root/Desktop/log.txt*, we obtained the image by issuing this command on CyborgLinux *ewfacquire -d sha1 -l /root/Desktop/log.txt /dev/sdb1*.

5.7 Forensics suite

There are 5 tools in this category (*autopsy, capstone, dff, dff-gui, dumpzilla*). DFF (Digital Forensics Framework) is used to collect, preserve and identify digital evidence. We need to load pre-prepared file with a forensic image into DFF and analyze the data file by one of the built-in modules (figure 1).

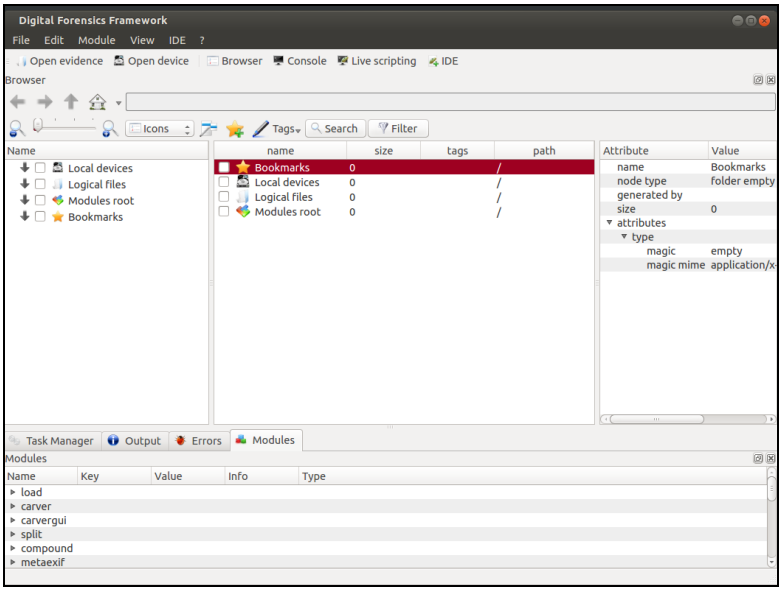


Figure 1. DFF analysis of evidence

5.8 Network investigation

There are 2 tools in this category (*p0f, xplico*).

P0f is a passive operating systems fingerprinting tool. All the host has to do is connect to the same network or be contacted by another host on the network. The packets generated through these transactions gives p0f enough data to guess the system. In our experiment, by issuing the command *p0f -f /etc/p0f -i eth0*, we were able to read fingerprints from */etc/p0f* and listen on eth0 via libpcap application.

Xplico is a Network Forensic Analysis Tool (NFAT) that is capable of extracting application data from packet capture files. It is best suited for offline analysis of PCAP files but it can also analyze live traffic. Xplico can extract email, HTTP, VoIP, FTP, and other data directly from the PCAP files. It is able to recognize the protocols with a technique named Port Independent Protocol Identification (PIPI).

5.9 Secure wipe

The tools in this category include 4 tools (*sdmem, sfill, srm, sswap*).

5.10 Steganography.

8 tools of this category are divided into 2 groups.

Steg Toolkit (*stegbreak, stegcompare, stegdetect, stegdeimage, steghide*) is used for the embedding system and the password when the attack succeeded for an image.

Snowdrop is intended to bring (relatively) invisible and modification-proof watermarking to a new realm of "source material" – written word and computer source codes. The information is not being embedded in the least significant portions of some binary output, as it would be with a traditional low-level steganography, but into the source itself.

Vinetto extracts the thumbnails and associated metadata from the Thumbs.db files.

Outguess is a universal steganographic tool that allows the insertion of hidden information into the redundant bits of data sources. The nature of the data source is irrelevant to the core of outguess. The program relies on data specific handlers that will extract redundant bits and write them back after modification. Currently only the PPM, PNM, and JPEG image formats are supported, although outguess could use any kind of data, as long as a handler were provided.

Stegdetect will look for signatures of several well-known steganography embedding programs in order to alert the user that text may be embedded in the image file, such as jpeg. To see if there is steganography embedded message in our *n.jpg* file in a USB drive, we launched *Stegdetect* by using this command: *stegdetect -t /media/cyborg/B4FE-5315/n.jpg.*

The result of executing the utility is shown below: *stegdetect -t /media/cyborg/B4FE-5315/n.jpg: negative,*
where *negative* indicates no message embedded in the *n.jpg* file.

5 Conclusions

In this work we have demonstrated the application of various computer forensics tools on Cyborg Hawk Linux. We showed the syntax for using the tools and the result of executing the tools on our virtual machine. As it was demonstrated the tools produce consistent results according to their specifications. However, similar results can be obtained by using physical machines. Our results will help the computer forensics investigators on selecting appropriate tool for a specific purpose. It also helps penetration testers to check for signs of vulnerabilities on their system. We showed that Cyborg Hawk Linux is a good choice for forensics investigators for several reasons. These include that the tools are free, easy to use, do not need configuration, and produce consistent results.

References

[1] O.L. Carroll, S.K. Brannon, and T. Song, "Computer forensics: Digital forensic analysis methodology", *Comp. Forensic*, vol. 56, no. 1, pp. 1-8, Jan.2008.

[2] M. Meyers and M. Rogers, "Computer forensics: The need for standardization and certification", *Int. J. Digit. Evidence*, vol. 3, no. 2, pp.1-11, Sep.2004.

[3] S. Rahman and M. N. A. Khan, "Review of live forensic analysis techniques", *Int. J. of Hybrid Inf. Technology*, vol.8, no.2, pp.379-388, 2015

[4] S. Yadav, "Analysis of digital forensic and investigation", *VSRD-IJCSIT*, vol. 1, no. 3, pp. 171-178, 2011

[5] A. Ghafarian, H. Seno, and S. Amin. "Exploring digital forensics tools in Backtrack 5.0 r3". Proceedings of International Conference on Security and Management - SAM'14, 2014.

[6] B. Carrier "Open source digital forensics tools: The legal argument". AtStake. Oct. 2002. [Online]. Available: http://dl.packetstormsecurity.net/papers/IDS/atstake_opensource_forensics.pdf [Accessed Oct. 27, 2017]

[7] D. Manson, A. Carlin, S. Ramos, A. Gyger, M. Kaufman, and J. Treichelt. "Is the open way a better way? Digital forensics using open source tools". System Sciences. HICSS 2007. 40th Annual Hawaii International Conference on Science, pp 266-270. [Online]. Available: https://www.computer.org/csdl/proceedings/hicss/2007/2755/00/27550266b.pdf [Accessed Oct. 27, 2017]

[8] A.Z. Tabona "Top 20 free digital forensics investigation tools for sysadmins". [Online]. 2002. Available : http://www.gfi.com/blog/top-20-free-digital-forensic-investigation-tools-for-sysadmins/ [Accessed Oct. 27, 2017]

[9] Wikipedia, "List of digital forensics tools". [Online]. Available : http://en.wikipedia.org/wiki/List_of_digital_forensics_tools [Accessed Oct. 27, 2017]

[10] Mares and Company, "Alphabetical list of links to manufacturers, suppliers, and products". [Online]. Available http://www.dmares.com/maresware/linksto_forensic_tools.htm [Accessed Oct. 27, 2017]

[11] KaliLinux [Online]. Available: https://www.kali.org/ [Accessed Oct. 27, 2017]

[12] CAINE (Computer Aided INvestigative Environment) [Online]. Available: http://www.caine-live.net/ [Accessed Oct. 27, 2017]

[13] DEFT (Digital Evidence & Forensics Toolkit) [Online]. Available: http://www.deftlinux.net/ [Accessed Oct. 27, 2017]
[14] PHLAK [Online]. Available: https://sourceforge.net/projects/phlakproject/ [Accessed Oct. 27, 2017]
[15] Cyborg Hawk Linux [Online]. Available: http://cyborg.ztrela.com/ [Accessed Oct. 27, 2017]
[16] BackTrack [Online]. Available: http://www.backtrack-linux.org/ [Accessed Oct. 27, 2017]
[17] Parrot Security OS [Online]. Available: https://www.parrotsec.org/ [Accessed Oct. 27, 2017]
[18] BackBox Linux [Online]. Available: https://backbox.org/ [Accessed Oct. 27, 2017]
[19] VMware [Online]. Available: http://www.vmware.com [Accessed Oct. 27, 2017]

PARAMETRIC MONITORING OF COMPUTING PROCESSES IN INFORMATION AND COMPUTING SYSTEMS

KhlaponinYurii [1] (y.khlaponin@gmail.com),
Khoroshko Volodimir [2] (professor_va@ukr.net),
Khokhlacheva Yuliia [2] (hohlachova@gmail.com),
Gavrilko Evgen [3] (gev.1964@ukr.net)
[1] Kyiv National University of Construction and Architecture, Kyiv, Ukraine
[2] National Aviation University, Kyiv, Ukraine
[3] State University of Telecommunications, Kyiv, Ukraine

An engineering approach based on modeling and measuring the parameters of information streams of ICS at the level of their hardware components is proposed, which allows to estimate the distribution of these streams when analyzing computational processes and studying the properties of ICS.

1. Introduction

The effectiveness of the application of information-computing systems (ICS) in specific conditions depends on the organization of the computing process (CP). With optimal VP organization in regards of some criterion the performance of the ICS - and, consequently, the efficiency - can be increased. Obviously, this explains the heightened interest in the organization of the CP, requiring the conduct of ICS studies operating in a particular mode (single-program, multiprogram with fixed or variable number of tasks, time-sharing and real time).

2. Problem formulation

There are a large number of approaches and methods for analyzing the organization of the CP, which is based on the measurement, modeling and analytical study of the values of certain indicators characterizing the work of the ICS. In recent years, methods based on modeling have been widely used, as well as experimental methods based on measurements of the parameters of real-time ICS systems on solvable problems [1-3].

A feature of most known approaches and methods of analyzing the CP is the formal consideration of certain processes and events occurring in the system and characterizing it at various stages of functioning at a logical level, i.e. from the point of view of an analyst or a system programmer, when individual structural links and the distribution of information streams (IS) between the hardware components of the ICS

are not taken into account. Usually in this case, assumptions of various kinds are accepted, which are caused not by the logic of the ICS, but by the specific nature of the applied mathematical apparatus. For example, by using methods based on the use of queuing networks as ICS models (queue system models), certain assumptions are made regarding the distribution of the streams entering the system of tasks and the duration of their servicing (decision time), which is caused by the requirement of analytical solvability of the model. In other words, the requirement of analytical solvability of the model imposes a number of additional restrictions [1], which is one of the reasons restraining the use of ICS models with queues.

Another reason for the limited use of these models is that the formal consideration of the CP often leads to undesirable practical consequences - for example, the adoption of incorrect decisions about changing the individual parameters of the CP based on the results of modeling, which made an unrealistic assumption about the distributions of streams entering the system of tasks and the time of their solution. Obviously, a different approach is required, which is devoid of this shortcoming. Such an approach can be, for example, operational analysis of computer performance [3]. The assumptions of this approach are directly related to the logic of the computer, and its conclusions are based on calculations of their performance through the values of operational variables (parameters) measured on a finite time interval, which is allowing to establish simple relationships between parameters and indicators of computer operation. The method of operational analysis is mainly based on the use of ICS models with queues and requires from an analyst the knowledge of the principles of ICS functioning at a logical level. It is allowing to evaluate the integral performance of the system, which is dependent on a large number of parameters and approximately calculated from the results of measurements without the use of ICS means. Based on the results of approximate calculations, the values of the basic indicators of the ICS are determined and then decisions are taken on the organization of the CP [1,3].

In a number of cases, when analyzing CP and studying the properties of ICS for specific applications, researchers and developers of universal and specialized ICS are interested in the differential IS system, which allows to assess the distribution of processes among individual structural components of ICS and, based on this, to make a decision either to improve the organization of the CP or to modernize technical resources. To obtain the differential statistics of IS, it is possible to apply an approach based on modeling and measuring the parameters of ICS

information streams at the level of their hardware components (HC), i.e. at the physical level. The peculiarity of this approach (i.e. engineering approach) is to consider from the engineer's point of view the structure of the ICS, the processes and events occurring in it, when it is required to decompose the system; to highlight in its complex structure the backbones and directions for which the IS is distributed; to identify the sets of IS's and parameters that characterize these streams; to establish relationships between threads and their parameters; to develop requirements for the measurement of these parameters; to conduct (by means of measurements) the collection and analysis of statistical data of IS. Based on the results of this analysis, an assessment is made of the distribution of IS among the HC of the system, which allows to identify "bottlenecks" in the organization of the CP or in the ICS structure and outline ways of eliminating them.

3. Solutions

Consider the structure of the generalized information stream (GIS), which conditionally includes all information streams in the modern ICS. By GIS we mean the set of Φ, which are the time-varying sequences of address codes $A(r)$, instructions $K(r)$, microinstructions $M(r)$, data $D(r)$, conditions $C(r)$, control information $V(r)$, interrupt requests $P(r)$, transmitted from sources h to consumers offline formation over the backbones $M_y(y = \overline{1,Y})$. The symbol r is used to distinguish elements of a certain sequence of codes (for example, for $r = Y_{ij}$, the symbol $A(Y_{ij})$ denotes the address of the device Y_{ij}); $r \in R$, where $R = \{r\}$ is the set of values. Suppose that if a sequence of codes contains elements of the same type, then the symbol r is omitted. It is obvious that the code sequences $\Phi^{A(r)}, \Phi^{K(r)}, \Phi^{M(r)}, \Phi^{D(r)}, \Phi^{C(r)}, \Phi^{U(r)}, \Phi^{P(r)}$ are homogeneous information streams and components of the generalized information stream Φ:

$$\Phi = \left\{\Phi^{A(r)}, \Phi^{K(r)}, \Phi^{M(r)}, \Phi^{D(r)}, \Phi^{C(r)}, \Phi^{U(r)}, \Phi^{P(r)}\right\}; \quad (1)$$

$$\Phi = \bigcup_{x \in X} \Phi^{x(r)}; \quad (2)$$

Where x is the symbol used to distinguish the components of the stream $\Phi^{D(r)}$; $X = \{A, K, M, D, C, U, P\}$ is the set of values from A to P. Expression (1) and (2) characterize the structure of the stream Φ,

showing its heterogeneity due to the presence of various types of components $\Phi^{x(r)}$.

Any information stream in the ICS can be characterized by the information capacity N, the information transfer speed V, the intensity λ, (the number of determined receipts per unit time), the probability $\Theta(n)$ of the arrival of certain sequences (or probability distribution functions Θ of two random sequences), the time τ between the arrival of consecutive codes (information units) and the time interval T sec of their arrival at certain nodes of the system, the priority π of the selection of sequences for processing, the reliability of transmission ε and magnitude information loss ξ.

Denote by z the value of the parameter $\Phi; z \in Z$, where $Z = \{N, V, \lambda, \Theta(n), \tau, T, n, \varepsilon, \xi\}$ is the set of values from N to ξ. The homogeneous stream $\Phi^{x(r)}$ will be characterized by the parameters $z^{x(r)}; z^{x(r)} \in Z^{x(r)};$ where $Z^{x(r)} = \{N^{x(r)}, V^{x(r)}, \lambda^{x(r)}, \Theta^{x(r)}(n), \tau^{x(r)}, T^{x(r)}, n^{x(r)}, \varepsilon^{x(r)}, \xi^{x(r)}\}$ is the set of values of the parameter $Z^{x(r)}$. Since the set Z_Φ of parameters of the generalized information stream Φ includes subsets $Z^{x(r)}$ of parameters of homogeneous information streams $\Phi^{x(r)}$, then for Z_Φ next expression is viable:

$$Z_\Phi = \bigcup_{x \in X} Z^{x(r)}; \quad (3)$$

for the subset $Z^{x(r)}$ viable expression is:

$$Z^{x(r)} = \{z^{x(r)}\}_{z \in Z, x \in X}. \quad (4)$$

Expressions (3) and (4) respectively set the sets of parameters of the generalized Φ and homogeneous $\Phi^{x(r)}$ information streams and reflect their quantitative-qualitative characteristics. In general, for Φ, expressions (2) and (3) can be regarded respectively as its structural and parametric models. Expression (4) is a parametric model of a homogeneous stream $\Phi^{x(r)}$.

Represent the set Z_{Φ} of parameters of the stream Φ in the form of a matrix:

$$W_{z^{x(r)}} = \begin{bmatrix} N^{A(r)} & V^{A(r)} & \lambda^{A(r)} & \Theta^{A(r)} & \tau^{A(r)} & T^{A(r)} & \pi^{A(r)} & \varepsilon^{A(r)} & \xi^{A(r)} \\ N^{K(r)} & V^{K(r)} & \lambda^{K(r)} & \Theta^{K(r)} & \tau^{K(r)} & T^{K(r)} & \pi^{K(r)} & \varepsilon^{K(r)} & \xi^{K(r)} \\ N^{M(r)} & V^{M(r)} & \lambda^{M(r)} & \Theta^{M(r)} & \tau^{M(r)} & T^{M(r)} & \pi^{M(r)} & \varepsilon^{M(r)} & \xi^{M(r)} \\ N^{D(r)} & V^{D(r)} & \lambda^{D(r)} & \Theta^{D(r)} & \tau^{D(r)} & T^{D(r)} & \pi^{D(r)} & \varepsilon^{D(r)} & \xi^{D(r)} \\ N^{C(r)} & V^{C(r)} & \lambda^{C(r)} & \Theta^{C(r)} & \tau^{C(r)} & T^{C(r)} & \pi^{C(r)} & \varepsilon^{C(r)} & \xi^{C(r)} \\ N^{U(r)} & V^{U(r)} & \lambda^{U(r)} & \Theta^{U(r)} & \tau^{U(r)} & T^{U(r)} & \pi^{U(r)} & \varepsilon^{U(r)} & \xi^{U(r)} \\ N^{P(r)} & V^{P(r)} & \lambda^{P(r)} & \Theta^{P(r)} & \tau^{P(r)} & T^{P(r)} & \pi^{P(r)} & \varepsilon^{P(r)} & \xi^{P(r)} \end{bmatrix}.$$

The rows of the matrix are subsets $Z^{x(r)}$ of parameters characterizing the corresponding streams $\Phi^{x(r)}$, and the columns are homogeneous subsets of parameters characterizing the stream Φ. The elements $Z^{x(r)}$ of the matrix $M_{r^{x(r)}}$ are parameters of the stream Φ. The elements pertaining to the row are different types of parameters of the homogeneous stream, and the elements related to the column are the same type parameters of heterogeneous streams. The fixation of the matrix $W_{z^{x(r)}}$ (or its elements)for a given time interval or when performing a certain work on the TDF allows to obtain the necessary data that can be used for statistical analysis of the operation of various HC systems. That includes channels and input-output devices, i.e. to analyze the work of the ICS and organize a CP on it.

The structures considered and the set of parameters of the GIF contain a set of elements by means of which any IS taking place in the ICS can be assigned or presented, i.e. the structural components $\Theta^{x(r)}$ of the stream Θ can be included into the structure of a specific stream $\Theta_{h,l}$ in a certain backboneM_y, and from the set $Z_{\Phi}$of the parameters of the GIF it is possible to select any subset $Z^{x(r)}$ of parameters characterizing the corresponding backbone information stream $\Phi_{h,l}$. Thus, the expressions (1), (4) and the matrix $W_{z^{x(r)}}$ contain all the necessary components by means of which it is possible to formally describe any IS in the ICS.

Let's consider some dependencies between the indicated parameters of the IS. First of all, note that the information capacity $N^{x(r)}$ of the stream $\Phi_{h,l}^{x(r)}$ should be understood as the number of transmitted

information codes $x(r)$ from the source h to the consumer l defined by the expression

$$\mathrm{N}^{x(r)} = \mathrm{V}^{x(r)}\mathrm{T}^{x(r)}. \quad (5)$$

Usually, in the process of the ICS functioning, the transfer $\Phi_{h,l}^{x(r)}$ between its HC is not carrying out continuously, but at arbitrary time intervals $t_e^{x(r)}(e = \overline{1,E})$:

$$\mathrm{T}^{x(r)} = \sum_{e=1}^{E} t_e^{x(r)}, \quad (6)$$

Where E is the number of time intervals; $t_e^{x(r)}$ is the time difference between the end $t_{ek}^{x(r)}$ and the beginning $t_{e\mathrm{H}}^{x(r)}$) of the transfer $x(r)$

$$t_e^{x(r)} = t_{ek}^{x(r)} - t_{e\mathrm{H}}^{x(r)}. \quad (7)$$

Taking into account (6) and (7), expression (5) takes the form

$$\mathrm{N}^{x(r)} = \mathrm{V}^{x(r)} \sum_{e=1}^{E} t_{ek}^{x(r)} - t_{e\mathrm{H}}^{x(r)}. \quad (8)$$

Denote by $n_e^{x(r)}$ the number of transmitted information codes $x(r)$ in the time intervals $t_e^{x(r)}$, then

$$\mathrm{N}^{x(r)} = \sum_{e=1}^{E} n_e^{x(r)}. \quad (9)$$

In the case where the IS is represented as a transmitted blocks of information containing a different number of bytes, the value $\mathrm{N}^{x(r)}$ for the observation period T can be represented as

$$\mathrm{N}^{x(r)}(T) = \sum_{\gamma=1}^{\Gamma} \delta_{\gamma}^{x(r)}. \quad (10)$$

Where $\delta_{\gamma}^{x(r)}$ is the number of bytes of information $x(r)$ in the block γ, and Γ is the number of transferred blocks.

The intensity value $\lambda^{x(r)}$ for the stream $\Phi_{h,l}^{x(r)}$ is determined from expression

$$\lambda^{x(r)} = \mathrm{N}^{x(r)}/\mathrm{T}^{x(r)}. \quad (11)$$

Typically, in the process of functioning of multiprogramm ICS's which have a non-stationary and random input stream of tasks, there is a situation when IS is transferred between HC at arbitrary moments of time during the considered period T. In other words, the probability $\Theta^{x(r)}(n)$ that the number n of information codes $x(r)$ comes from the source to the consumer at a given time is equal to the probability of their arrival at any

other time. It follows that the probability $\Theta^{x(r)}(n)$ corresponds to a Poisson distribution [5]:

$$\Theta^{x(r)}(n) = \frac{\omega^{-\Omega(n)}\Omega^{n}(n)}{n!}, \quad (12)$$

where $\Omega(n)$ is the average value of n for a given T.

If it's assumed that the process of transfer of information codes $x(r)$ in streams $\Phi_{h,l}^{x(r)}$, which are distributed between the HC of the considered ICS, is stochastic in nature, then the probability $\Theta(\tau^{x(r)} \leq \tau_з)$ of the fact, that the time intervals $\tau^{x(r)}$ between the arrival of these codes will be less than some predetermined number $\tau_з$, corresponds to the exponential distribution for which [4,5] can be described as:

$$\Theta\left(\tau^{x(r)} \leq \tau_з\right) = 1 - \omega^{-\tau_з/\Omega(\tau^{x(r)})}, \quad (13)$$

where $\Omega(\tau^{x(r)})$ is the average value of the time $\tau^{x(r)}$.

It is quite possible that for some IS's in a certain ICS there is an equable distribution of the probability of transmission of information codes $x(r)$. Then the value of $\Theta^{x(r)}(n = N)$ will be constant and equal to the probability of transfer of Θ codes $x(r)$at a given moment.

The reliability of the transmission of information codes $x(r)$ in the streams $\Phi_{h,l}^{x(r)}$ is determined from expression

$$\varepsilon^{x(r)} = 1 - \Theta_0^{x(r)}; \quad (14)$$

where $\Theta_0^{x(r)}$ is the probability of an error in the ICS during the transmission of the stream $\Theta_{h,l}^{x(r)}$.

The required level $\varepsilon^{x(r)}$ in the ICS is provided by automatic control and correction of the detected errors during the functioning of the system.

The considered IS parameters and some of the dependencies between them, which are represented by expressions (5)-(14), can be used to calculate the integral and differential performances of individual HC and ICS in general. However, it should be noted that the dependencies between the IS parameters, which are necessary to define certain indicators, are determined in each specific case based on the objectives of the ICS study and the depth of the analysis. The latter also applies to integral and differential indicators, which are introduced if necessary. For example, consider some of these indicators.

For HC, which are simultaneously the source h and the consumer of information streams, an indicator is introduced that characterizes the time $T_s^{\text{апп}}$ of their occupation by receive-transmit operations equal to the sum of the time spent on transmitting $T_s^{x_h(r)}$ and receiving $T_s^{x_l(r)}$information codes $x(r)$.

$$T_s^{\text{зпп}} = \sum_{x\in X; r\in R} T_s^{x_h(r)} + T_s^{x_l(r)}, \quad (15)$$

where *s* is the HC index; $s \in S, S = \{s\}$ is the set of indices; $T_s^{x_h(r)}$and $T_s^{x_l(r)}$ consist of their time intervals $t_\alpha^{x_h(r)}(\alpha = \overline{1,A})$ and $t_\beta^{x_l(r)}(\beta = \overline{1,B})$.

$$T_s^{x_h(r)} = \sum_{\alpha=1}^{A} t_\alpha^{x_h(r)} \quad (16)$$

$$T_s^{x_l(r)} = \sum_{\beta=1}^{B} t_\beta^{x_l(r)} \quad (17)$$

In turn, $t_\alpha^{x_h(r)}$ and $t_\beta^{x_l(r)}$ can be defined as time differences respectively between the ends $t_{\alpha k}^{x_h(r)}$, $t_{\beta k}^{x_l(r)}$ and beginnings $t_{\alpha \text{н}}^{x_h(r)}$, $t_{\beta \text{н}}^{x_l(r)}$ of occupation:

$$t_\alpha^{x_h(r)} = t_{\alpha k}^{x_h(r)} - t_{\alpha \text{н}}^{x_h(r)}; \quad (18)$$

$$t_\beta^{x_l(r)} = t_{\beta k}^{x_l(r)} - t_{\beta \text{н}}^{x_l(r)}. \quad (19)$$

Taking into account (16)-(19), expression (15) can be described as:

$$T_s^{\text{зпп}} = \sum_{x\in X; r\in R} \left[\sum_{\alpha=1}^{A} (t_{\alpha k}^{x_h(r)} - t_{\alpha \text{н}}^{x_h(r)}) + \sum_{\beta=1}^{B} (t_{\beta k}^{x_l(r)} - t_{\beta \text{н}}^{x_l(r)}) \right].$$

To analyze the usage of RAM, I/O devices and buses, usually an index is of interest, which is characterizing the total N_s^{Σ}volume of information codes $x(r)$ transmitted by $N_s^{x_h(r)}(T)$ and received $N_s^{x_l(r)}(T)$ in the time interval T:

$$N_s^{\Sigma}(\text{T}) = \sum_{x\in X; r\in R} \left[N_s^{x_h(r)}(T) + N_s^{x_l(r)}(T) \right]. \quad (20)$$

Taking into account (5), expression (20) is taking form

$$N_s^{\Sigma}(\mathrm{T})=\sum_{x\in X;r\in R} V_s^{x(r)}\,(T_s^{x_h(r)}+T_s^{x_l(r)}).$$

When researching the organization of the CP on the existing ICS, as well as in the design of new systems, it is necessary to estimate the utilization coefficient for various types of their HC constituents. Usually, this coefficient is determined [] from the ratio of the equipment load to the maximum load that this equipment can withstand, or from the ratio of the time of equipment occupancy to the total time of its operation. Denote by $\rho_s(T)$ the utilization coefficient of HC, distinguished by the index s, on the interval T. For most HC

$$\rho_s(T) = T_s^{\text{зпп}}/T_s^{\Phi}, \tag{21}$$

where T_s^{Φ} is the total time of HC functioning on the considered interval T, the value of $\rho(T)$ has an upper $sup\rho_s$ and a lower $inf\rho_s$ boundary:

$$inf\rho_s = min\rho_s = 0 \le \rho_s(T) \le 1 = sup\rho_s = \max\rho_s.$$

Obviously, in the ideal case, it should be

$$\lim_{T_s^{\text{зпп}}\to T_s^{\Phi}} \rho_s(T) = 1.$$

The integral indicator of the work of HC can be determined with the help of expression (21), and the differential indicators, which may be indicators characterizing the degree of occupancy (usage) of HC by the transfer $\rho_s^{x_h(r)}(T)$ and the reception $\rho_s^{x_l(r)}(T)$ of information codes $x(r)$ on the interval T, are determined from the expressions

$$\rho_s^{x_h(r)}(T) = \frac{T_s^{x_h(r)}}{T_s^{\Phi}};$$

$$\rho_s^{x_l(r)}(T) = \frac{T_s^{x_l(r)}}{T_s^{\Phi}}.$$

To assess the performance of the I/O channels $K_j(j=\overline{1,J})$ performing the functions of the IS distributor in modern ICS and being the central link between different types of external devices (ED), RAM and the central processor, an indicator can be used that reflects their actual capacity $G_{K_j}(T)$ on the interval T:

$$G_{K_j}(T) = N_{K_j(T)}^{\Sigma}/T_{K_j}^{\text{зпп}},$$

Where $N^{\Sigma}_{K_j(T)}$ is the total volume of all information transmitted from K_j and received by it on the interval T; $T^{\text{зпп}}_{K_j}$ of the occupation of K_j by the receive-transmit operations on the given interval. For $G_{K_j}(T)$, as well as for $\rho_s(T)$, there are upper and lower boundaries; i.e. $inf G_{K_j}(T) = min G_{K_j}(T) = 0 \le G_{K_j}(T) \le V^{\text{x(r)}}_{K_j} = sup G_{K_j}(T) = max G_{K_j}(T)$

In the ideal case, the value of $G_{K_j}(T)$ can be equal to the receive-transmit speed $V^{\text{x(r)}}_{K_j}$ of the information codes $x(r)$ of the channel K_j, from the RAM and the ED or from the ED and RAM:

$$\lim_{T^{\text{зпп}}_{K_j} \to T^{\Phi}_{K_j}} K_j(T) = V^{\text{x(r)}}_{K_j}.$$

The considered examples clearly show only some possibilities of using the proposed engineering approach in the analysis of CP and studying the properties of ICS for specific applications. However, they do not exhaust all its capabilities, which can be disclosed when considering other dependencies and indicators, introduced in accordance with the objectives of the ICS study.

For the fullest use of the capabilities of the engineering approach, it is necessary to know the structure of the ICS being studied and the peculiarities of its functioning at the HC level, which will make it possible to distinguish the lines and directions of the IS transfer among separate HC in this structure, and also to determine the types of homogeneous IS's transmitted in these backbones.

Thus, it is possible to formally describe any IS in the ICS and obtain the required expressions for calculating the integral and differential performance of individual HC and ICS in general. The analysis of the expressions, obtained as a result of the formalization of the IS for the parameters, dependencies and indicators considered, allows to determine the basic requirements for the selection of measuring means (MM), by means of which it is possible to obtain differential and integral IS statistics in ICS. First of all, it should be noted that most of these expressions contain parameters as components, characterizing either the number of transmitted information codes, or the time taken to transmit them, or both, i.e. the parameters $N^{x(r)}_{s}$ and $T^{x(r)}_{s}$ are basic and are used to determine the values of other parameters, dependencies and indicators.

It follows that one of the main requirements for MM is the ability to measure various values of the parameters $N_s^{x(r)}$ and $T_s^{x(r)}$ for multiple HC systems when transmitting or receiving information codes $x(r)$. Since the transmission (reception) of $x(r)$ between the HC during the operation of the ICS is carried out at random intervals, the measuring device must record the start and end times of transmission (reception) of these codes; calculate the value $N_s^{x(r)}$ on each time interval and accumulate its total sum at all intervals; determine the duration of each interval and accumulate the values $T_s^{x(r)}$ as a sum of individual time intervals; provide the possibility of specifying different values of the measurement period T; differentiate the measured parameter values for various information codes $x(r)$. In addition, the MM should ensure, during the measurement of microscopic events (the parameters of the IS considered above) at the register level, high levels of accuracy, reliability and resolution, without introducing additional interference that could affect the results of the operation of the ICS. They should be convenient and simple in practical application, correctly interact with the ICS in the process of collecting, fixating and accumulating the values of the measured IS parameters.

4. Conclusion

The autonomous multichannel hardware measurers are the most suitable for the listed requirements, the use of which is most effective for obtaining differential and integral statistics of IS in ICS. However, modern ICS of domestic and foreign production are not equipped with such measurers, so it is necessary to solve a number of issues related to their development and manufacturing.

The proposed engineering approach makes it possible to establish simple relationships between the parameters of IS and the performance of individual HC's and ICS's as a whole, as well as to estimate the distribution of these streams between these components, necessary for the analysis of CP's and the study of ICS properties for specific applications.

References

1. Ferrari D. Evaluation of the performance of computing systems. –M: Mir, 1981. – 576 p.

2. Stoyer R. Multi criteria optimization, theory, calculations and applications. – M: Radio and communication, 1992. – 504 p.

3. Dmitriev Y.K., Horoshevsky V.G. Computing systems and micro-computers– M: Radio and communication, 1982. – 304 p.

4. Skorobogatko E.A., Timchenko N.P., Khoroshko V.A. The choice of optimal data traffic in information networks // Zahist Informatsii, special issue, 2014. – P. 50-59.

5. Ivanchenko E.V., Khoroshko V.A. Evaluation of the efficiency of diagnostics of the means of information exchange in security systems and their networks. // Information processing systems, Volume 2 (118), v.2, 2014. – P. 96-101.

ПРОБЛЕМНО-ОРИЕНТИРОВАННАЯ ПЛАТФОРМА ТРАНСФЕРА ЗНАНИЙ ДЛЯ ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ В СОЦИОТЕХНИЧЕСКИХ СИСТЕМАХ

Цыганок В.В. [1], **Борохвостов И.В.** [2], **Роик П.Д.** [1]
[1]***Институт проблем регистрации информации Национальной академии наук Украины, г. Киев, Украина***
[2]***Центральный научно-исследовательский институт вооружения и военной техники Вооруженных Сил Украины, г. Киев, Украина***

Вступление

Одной из актуальных проблем глобального характера на сегодня в мире остается проблема как можно более полного использования имеющихся знаний. Эффективность применения знаний в различных областях, несомненно, принадлежит множеству важнейших факторов, от которых напрямую зависит всеобъемлющее развитие человечества. Согласно исследованиям, проводимым американской компанией Delphi Group в начале нынешнего тысячелетия [1], довольно значительная доля знаний, используемых определенной организацией являются неформализованными и не зарегистрированы на носителях информации. Эти знания базируются только лишь на навыках, опыте и интуиции определенных специалистов-экспертов (см. рис. 1).

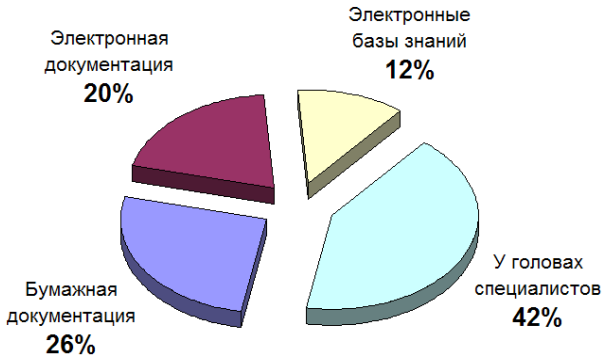


Рис. 1. Состав знаний, используемых определенной организацией в США (по результатам исследования Delphi Group)

Поэтому при принятии обоснованных решений в разных сферах деятельности нужно опираться на все имеющиеся знания, как

доступные на текущий момент, зарегистрированные на определенных носителях, так и экспертные. Это особенно касается слабо структурированных предметных областей, где уровень неопределенности и неполноты информации по обыкновению очень высокий. Кроме того, для полноты использования знаний нужно задействовать текстологические методы их получения и обработки, такие как Natural Language Processing, контент-мониторинг, Data Mining, анализ онтологий, и т.п..

Идеей, направленной на внесение значительного вклада у решение выше упомянутой проблемы, может быть интеграция знаний разной природы, полученных в результате их автоматизированного экстрагирования из разных источников, которые изменяются во времени, с экспертными и формализованными знаниями, их обработка и хранения в соответствующих базах с последующей передачей (трансфером) этих знаний для использования при поддержке принятия решений. Базируясь на этой идее предлагается создать соответствующую проблемно-ориентированную аппаратно-программную платформу трансфера знаний, схема которой предлагается на рис. 2.

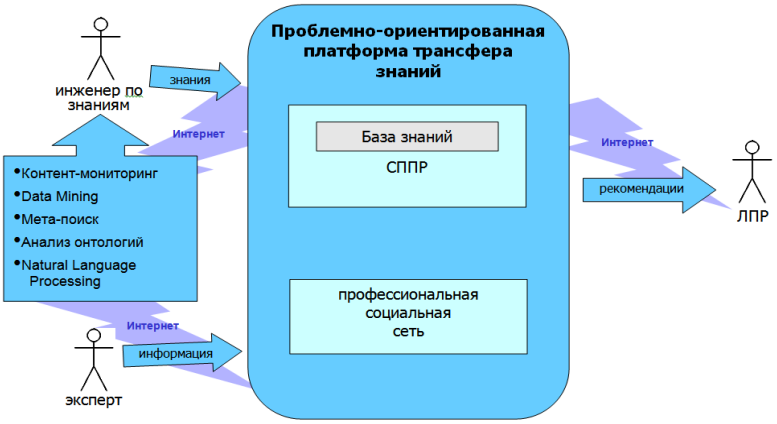


Рис. 2. Схема проблемно-ориентированной платформы трансфера знаний

Предполагается, что основным актером в системе является инженер по знаниям, который, владея определенными знаниями в предметной области, пользуясь инструментарием экстрагирования знаний и организовывая групповые экспертизы создает модели

предметных областей в виде соответствующих баз знаний. В дальнейшем на основе созданной базы знаний генерируются рекомендации для лиц, принимающих решения (ЛПР). Функционально в системе за генерацию рекомендаций отвечает система поддержки принятия решений (СППР), как составляющая проблемно-ориентированной платформы трансфера знаний [2]. Еще одной составляющей платформы должны быть профессиональная социальная сеть, объединяющая специалистов-экспертов разных направлений деятельности, как источников получения знаний. Относительная компетентность этих специалистов в текущем вопросе является неотъемлемым фактором при учете предоставленной ими информации [3].

Цели создания проблемно-ориентированной платформы трансфера знаний

Основной целью создания проблемно-ориентированной платформы есть повышение эффективности применения разных видов знаний при поддержке принятия решений в слабо структурированных предметных областях, к которым относятся социотехнические системы, включающие в себя человеческую и технологическую составляющую. Цель будет достигнута путем разработки методов, моделей и программных средств, позволяющих более эффективно получать знания из различных источников, одновременно обрабатывать знания разной природы и применять их для выработки управленческих рекомендаций. Актуальность научных исследований обусловлена тем, что значительная доля знаний в слабо структурированных областях является незарегистрированной и недоступной для использования на текущий момент, причем, существенным процентом неформализованных знаний владеют лишь определенные специалисты благодаря своему уникальному опыту и интуиции. Целесообразно при построении баз знаний использовать все доступные знания, как явные (формализованные), так и неявные (которые могут быть полученные в информационных сетях и при обращении к экспертам).

На основе баз знаний, построенных путем декомпозиции и структуризации неформализованных проблем, применяя разработанные методы поддержки принятия решений, планируется разработать удобные механизмы предоставления знаний для использования лицами, которые в них нуждаются. Для дальнейшего их эффективного применения возникает потребность в обработке знаний разной природы, полученных из разных источников: общей

их проверки на достоверность, полноту, сбалансированность, согласованность, возможность их систематизации и обобщения.

Следует отметить, что методы и средства, которые способствуют повышению уровня применения таких знаний, в полной мере не разработаны. Результатом исследований становятся новые методы, модели и созданые на их основе программные средства, позволяющие с достаточной полнотой получать знания об определенной слабо структурированной предметной области, одновременно обрабатывать знания разной природы, при необходимости – формализовать их, а также сформировать механизм использования этих знаний лицами, принимающими решения в разных областях.

Итак, существует необходимость создания ряда новых методов и моделей:

- методов систематизации неявных знаний;
- моделей и методов совместного использования явных и неявных знаний при создании проблемно-ориентированных баз знаний;
- моделей и методов эффективного получения знаний от экспертов и инженеров по знаниям с учетом их относительной компетентности;
- методов определения адекватности моделей слабо структурированных предметных областей (при условии отсутствия эталонов и уникальности решений);
- эффективных методов поддержки принятия решений на основе созданных баз знаний в слабо структурированных неформализованных предметных областях.

При разработке

- методов систематизации неявных знаний будут использованы методы системного анализа – метод декомпозиции при построении структурной модели предметной области, методы теории графов для анализа сетевой структуры базы знаний, линейной алгебры при обработке информации в матричном виде, и т.п.;
- моделей и методов совместного использования явных и неявных знаний при создании проблемно-ориентированных баз знаний целесообразно применять методы количественного и качественного анализа (для соответствующих видов знаний), методы Natural Language Processing для анализа текстовых данных информационного поиска, контент-мониторинга для выявления понятийных сущностей, методы математической статистики при оценивании статистических данных контент-

мониторинга, методы определения согласованности экспертных оценок с целью их возможного применения для определения согласованности знаний разной природы, и т.п.;

- моделей и методов эффективного получения знаний от экспертов и инженеров по знаниям с учетом их относительной компетентности могут применяться методы экспертного оценивания (метод парных сравнений, в том числе для группового оценивания [4], с использованием шкал разной подробности [5], с обратной связью с экспертами [6], методы группового ординального оценивания [7,8], и т.п.), методы определения согласованности экспертных знаний (предложенный авторский показатель согласованности с двойным применением формулы энтропии [9], спектральный коэффициент согласованности [10], consistency index [11] и др.), агрегации экспертных оценок (авторский комбинаторный метод – перебора покровных деревьев – в модификациях с вычислением среднего арифметического и среднего геометрического [12, 13], метод собственного вектора [14], метод логарифмических наименьших квадратов [15]), и т.п.;
- методов определения адекватности моделей слабо структурированных предметных областей и соответствующих критериев будут применены имеющиеся методы математической статистики для определения критериев адекватности моделей, таких как критерий Фишера и др;
- эффективных методов поддержки принятия решений (на основе баз знаний, созданных в слабо структурированных неформализованных предметных областях) применимы метод анализа иерархий [11, 14], метод целевого динамического оценивания альтернатив [16], морфологический анализ [17], SWOT-Анализ [18], и т.п. для синтеза более эффективных методов поддержки принятия решений или модификации имеющихся.

Состояние разработки проблемы и предполагаемые пути для ее решения

В настоящее время имеющиеся методы получения и обработки знаний при поддержке принятия решений разделяют на коммуникативные (мозговой штурм, круглый стол, анкетирование, интервью, групповые и индивидуальные экспертные парные сравнения, и т.п.) и текстологические (Natural Language Processing, контент-мониторинг, Data Mining, анализ онтологий, и т.п.) [19]. Эти

методы предназначены для оперирования знаниями, принадлежащими лишь к некоторым определенным категориям. Так, в условиях строгих временных ограничений нецелесообразно организовывать групповую экспертизу из-за долгой продолжительности и высокой стоимости процедур сбора экспертных знаний. В тот же время, для основательного, полного и всестороннего описания междисциплинарной проблемы в слабо структурированной предметной области недостаточно ограничиваться лишь результатами контент-мониторинга и Data Mining'a; для повышения достоверности такого описания целесообразно привлекать также экспертные знания. Междисциплинарный характер проблемы не позволяет ограничиться лишь онтологиями отдельных специализированных предметных областей, которые, при этом не одинаково детализированы. Величины взаимного влияния ключевых понятий онтологии в контексте разных по смыслу проблем могут не совпадать. Data Mining в слабо структурированных предметных областях характеризуется проблемами нерепрезентативности и недостаточности данных.

Попытки использования результатов Data Minig, NLP, контент-мониторинга в качестве входных данных для экспертных методов являются нецелесообразными вследствие невозможности определения количественных соотношений между объектами базы знаний и невозможности организовать процесс повышения согласованности данных путем обратной связи с источниками информации.

Кроме того, модели предметных областей в виде иерархической древовидной целевой структуры (например, метод анализа иерархий Саати [11]) не могут быть достаточно адекватны к слабо структурированным системам, моделируемых через использование в них лишь объектов качественного (не количественного) характера – ориентация лишь на экспертное оценивание, наличие исключительно положительных влияний (связей) между объектами, отсутствие обратных связей в иерархии, отсутствие учета динамики влияний, и т.п.

Относительно применения экспертного оценивания при получении знаний, то есть необходимость предоставлять возможность экспертам в ходе экспертизы использовать шкалы разной подробности [20] с целью по возможности полного получения знаний от каждого эксперта с непричинением давления на него. Это дополнительно предоставляет возможность учитывать

относительную компетентность каждого эксперта в определенном вопросе экспертизы.

Выводы

Итак, методы и модели получения, обработки и применения знаний при поддержке принятия решений, которые бы предоставляли возможность одновременно использовать в полной мере как имеющиеся знания об определенной слабо структурированной предметной области, так и неформализованных знаниях (полученные из источников информации разной природы и от экспертов) сейчас не разработаны, хотя необходимость их разработки уже приобрела актуальность. Требования по их разработке связаны с созданием проблемно-ориентированной платформы трансфера знаний, направленной на решение глобальной проблемы как можно более полного использования накопленных человечеством знаний.

Исследования выполнены в рамках проекта Ф73/23558 «Разработка методов и средств поддержки принятия решений при выявлении информационных операций». Проект является победителем конкурса Ф73 на грантовую поддержку научно-исследовательских проектов Государственного фонда фундаментальных исследований Украины и Белорусского республиканского фонда фундаментальных исследований.

Литература

1. Тузовский А.Ф., Чириков С.В., Ямпольский В.З. Системы управления знаниями (методы и технологии). Томск: Изд-во НТЛ, 2005. – 260 с.
2. Циганок В.В. Концепція створення систем підтримки прийняття рішень, що адаптивні до рівня компетентності експертів. *Реєстрація, зберігання і обробка даних*. – 2011. – т.13. №2. – С.106-114.
3. Каденко С.В., Циганок В.В. Визначення відносної компетентності експертів під час агрегації парних порівнянь. *Реєстрація, зберігання і обробка даних*. – 2017. – т.19. №2. – С.69-83.
4. Zgurovsky M.Z., Totsenko V.G., Tsyganok V.V. Group Incomplete Paired Comparisons with Account of Expert Competence. *Mathematical and Computer Modelling*. – February 2004 . – v.39, №4-5 .– P.349-361.
5. Tsyganok V.V., Kadenko S.V. & Andriichuk O.V. Using different pair-wise comparison scales for developing industrial strategies.

International Journal of Management and Decision Making. – 2015. – vol. 14, issue 3. – P. 224-250.
6. Totsenko V.G., Tsyganok V.V. Method of paired comparisons using feedback with expert. *Journal of Automation and Information Sciences.* – 1999. – Vol.31, No9. – P.86-97.
7. Tsyganok V.V., Kadenko S.V. On Sufficiency of the Consistency Level of Group Ordinal Estimates. *Journal of Automation and Information Sciences.* – 2010. – v.42, issue 8.– P.42 47.
8. Tsyganok V.V. Providing sufficient strict individual rankings' consistency level while group decision-making with feedback. *Journal of Modelling in Management.* – 2013. – vol. 8, issue 3. – P. 339-347.
9. Olenko Andriy & Tsyganok Vitaliy Double Entropy Inter-Rater Agreement Indices. *Applied Psychological Measurement.* – 2016. – v.40(1). – P.37-55.
10. Totsenko V.G. Spectral Method for Determination of Consistency of Expert Estimate Sets / V.G.Totsenko // Engineering Simulation. – 2000. – 17. – P.715-727.
11. Саати Т. Принятие решений. Метод анализа иерархий: Tomas Saaty. The Analytic Hierarchy Process. – Пер. с англ. Р.Г.Вачнадзе. – М.: Радио и связь, 1993. – 315 с.
12. Циганок В.В. Комбінаторний алгоритм парних порівнянь зі зворотним зв'язком з експертом. *Реєстрація, зберігання і обробка даних*. – 2000. – Т.2, №2. – С.92-102.
13. Tsyganok V.V. Investigation of the aggregation effectiveness of expert estimates obtained by the pairwise comparison method. *Mathematical and Computer Modelling.* – August 2010. – v.52, №3-4. – P.538-544.
14. Saaty T.L. The Analytic Hierarchy Process: planning, priority setting, resource allocation. N.Y.: McGraw Hill. – 1980. – 287 p.
15. Bozoki S., Fulop J., Ronyai L. On optimal completion of incomplete pairwise comparison matrices. *Mathematical and Computer Modelling*. – 2010: 52(1-2), P. 318-333.
16. Totsenko V.G. One Approach to the Decision Making Support in R&D Planning. Part 2. The Method of Goal Dynamic Estimating of Alternatives. *Journal of Automation and Information Sciences.* – 2001. – vol. 33, #4. – P. 82-90.
17. Панкратова Н.Д., Савченко І.О. Морфологічний аналіз. Теорія, проблеми, застосування. – Київ: Наук. думка,. 2015. – 245 с.
18. Майсак О.С. SWOT-анализ: объект, факторы, стратегии. Проблема поиска связей между факторамию *Прикаспийский*

журнал: управление и высокие технологии. – 2013. – № 1 (21). – С.151-157.
19. Таран Т.А., Зубов Д.А. Штучний інтелект. Теорія і застосування. Луганськ: Вид-во СНУ ім. В.Даля, 2006. – 240 с.
20. Циганок В.В. Вибір шкали оцінювання експертом у процесі виконання ним парних порівнянь в системах підтримки прийняття рішень. *Реєстрація, зберігання і обробка даних.* – 2011. – т.13. – №3.– С.92-105.

CONSIDERING IMPORTANCE OF INFORMATION SOURCES DURING AGGREGATION OF ALTERNATIVE RANKINGS

Tsyganok V.V. (vitaliy.tsyganok@gmail.com),
Kadenko S.V. (seriga2009@gmail.com),
Andriichuk O.V. (oleh.andriichuk@i.ua)
Institute for Information Recording of National Academy of Sciences of Ukraine
Kyiv, Ukraine

The paper outlines several approaches to aggregation of rankings of alternatives, provided by a set of individual information sources (IS). It is shown, that, depending on dimensionality of the set of alternatives, different aggregation methods can be applied. Particular attention is given to weighted aggregation of alternative rankings provided by different IS. Some intuitive assumptions concerning IS weight calculation are set forth. It is assumed that IS weight should reflect both quantity and quality of information it provides, as well as expert estimate of IS credibility based on previous experience of IS usage. Based on these assumptions, expressions for IS weight calculation are suggested. The approaches, suggested in the paper can be aplied to information, provided by sources of different nature, i.e., experts, search engines, paper and online documents, etc.

1 Introduction

In the process of informational and analytic research a common problem, often faced by analysts is compiling a ranking of a set of objects or alternatives (products, electoral candidates, political parties etc) according to some criterion. This criterion may be explicitly formulated by a decision-maker, or presented in the form of a particular topical query (such as “top comic actors”, “best online footwear stores”, “downtown restaurants”, etc). Even these randomly selected formulations demonstrate that it is problematic to select these alternatives, satisfying the respective queries based on solely numeric data, as there is no quantitative criterion, according to which the alternatives could be measured. A common approach, summarized in the most explicit way, perhaps, by Tom Saaty ([1]), is as follows: if you cannot measure the alternatives, the best way to numerically describe them, is to compare them among themselves. Conceptually, comparisons of alternatives according to some pre-defined criterion can be classified as either cardinal or ordinal estimates. Cardinal

pair-wise comparisons of alternatives bear information about numeric relation between them according to the given criterion, while ordinal comparisons indicate only the ordering of these alternatives. Cardinal pair-wise comparisons are good for relatively small numbers of alternatives, lying within the same order of magnitude. Ordinal pair-wise comparisons are easier to obtain and process and they can be used to compile rankings of large numbers of alternatives.

Amount of data, used in informational and analytical research is, usually, large, and that is why ordering of alternatives is more frequently used. For example, this is one of the possible reasons why web search engines operate mostly with rankings (orderings) of references and not with their numerically expressed ratings of some kind.

In order to compile a ranking of alternatives based on all available information, rankings, coming from several information sources, need to be taken into consideration and aggregated in some way. Authors of [2] show that information space is influenced by multiple factors of qualitative nature, which are difficult to formalize and describe in numeric terms. Based on these characteristics, it was shown in [3] that information space was a typical example of a weakly structured subject domain. In order to provide at least some analytical description of a weakly structured domain, data, coming from any available sources, needs to be taken into consideration. In this paper we will try to address the most general case, when information sources can include paper and online documentation, web search results, and expert judgments.

The problem with aggregation of data, coming from different information sources is that, in the general case, these sources have different weights, depending, again, on several factors. For example, it is reasonable to assume, that more credible information source, providing larger amount of information, should be assigned greater weight prior to data aggregation. While there is, roughly, a dozen of methods of ranking aggregation (proposed by Borda, Condorcet, Kemeny, and others), that are commonly known and described in a multitude of publications, the problem of defining the weights of data sources (voters, judges, experts, search engines etc.) during data aggregation is still relevant and open to discussion.

So, in further sections of our paper, we are going to set forth a method, allowing to aggregate data, presented in the form of alternative rankings, coming from different information sources, particularly focusing on different aspects of information source weight definition problem.

2 Literature review

The problem of rank aggregation became relevant back in the ancient times, when democratic societies and voting systems started to emerge. An extensive analysis of voting rules and their evolution is provided, for example, in [4].

The first rank aggregation methods, that emerged in the process of democratic voting evolution, which are still relevant today, were suggested in the late 18th century by Borda ([5]) and Condorcet ([6]). Since then multiple approaches to aggregation of individual preferences using different social welfare functions were devised. We find it most appropriate to mention the period of 1950-s and 1960-s – the time when Arrow's impossibility theorem was formulated (see [7]) and attempts were made to bypass its constraints (particularly, by Kemeny [8] and Copeland [9]). F.Aleskerov in [10] summarizes Arrovian aggregation rules in his book.

In the 2000s, with growing popularity of online information search engines, the problem of rank aggregation became even more relevant, as it became evident, that beside data obtained from voters, experts, and analysts, it was necessary to take online information sources into account (sometimes, with minimum human involvement). In this context, we should mention several papers from the 2000s, particularly [11-13]. In these publications the authors suggest and compare several approaches, targeted at aggregation of data from online information sources.

Definition of weights of information sources and weighted aggregation of rankings represents a separate matter. Methods of Borda and Condorcet can be easily extrapolated to the case when information sources have different weights, as shown by Totsenko in [14]. Kemeny's median is a more problematic case when information sources have different weights. Ordinal factorial analysis methods, described in [15, 16] allow analysts to calculate weights based on available sets of alternative rankings, previously provided by experts. This approach can be extrapolated to calculation of weights of online information sources as well, if evaluators provide some global alternative ranking a priori.

Dwork et al in [11] stress the importance of involvement of human evaluators in the process of preliminary information source weight definition. However, their paper focuses mostly on comparative analysis of rank aggregation methods and does not address the problem of IS weight calculation directly.

As for IS weights, we can, again, mention Tom Saaty ([1]) and his academic school (including Ramanathan & Ganesh [17], Forman & Peniwati [18], as well as Yang et al [19]). However, their methods are

primarily targeted at aggregation of expert estimates, represented in the form of pair-wise comparison matrices (PCM), provided in some ratio scales (and not rankings).

In our current paper we are going to suggest information source weight calculation methods based on both preliminary human evaluation of information source credibility and amount and quality of information, provided by a given information source.

3 The problem statement

What is given: n information sources ($IS_1 - IS_n$). Each of IS provides a ranking R_i, that includes m_i alternatives (i=1..n). Information sources can be experts, search engines, analysts who monitor online content, paper or online documents, etc. We should stress, that the number of alternatives, provided by different information sources is, in the general case, different ($m_i \neq m_j; i, j = 1..n$).

We should find an aggregated (global) ranking of alternatives.

4 Solution ideas for the case of equal weights of information sources

We suggest building a solution algorithm based on available methods of aggregation of individual rankings (ordinal estimates). The most common ranking aggregation methods were proposed by Borda, Condorcet and, perhaps, Kemeny. We should stress that if we are dealing with experts, the number of alternatives an expert can analyze "in one go" is no more than 7±2. However, search engines can return rankings of hundreds of links. That is why, usage of domination matrix – based methods is not recommended for such dimensionalities. For example, if we are dealing with 10 IS, providing rankings of 1000 alternatives each, then we have to analyze and process 10 matrices with 1000×1000 cells. In such cases Borda method seems to be the most appropriate option, as it operates with ranking vectors and not domination matrices. Besides, it is easier to extrapolate Borda method to the case when IS have different weights (while, for in stance, extrapolation of Markov chain-based methods looks more problematic). If all IS have the same weight, we can use the following algorithm.

1) Find the length of a ranking with all unique alternatives, featured in individual rankings $M = card(\bigcup_{i=1}^{n} \{A_j^{(i)}; j = 1..m_i\})$.

2) Form a unified set of alternatives.. For this purpose we should add to each ranking R_i of m_i alternatives all the remaining alternatives $\bigcup_{k=1}^{n}\{A_j^{(k)}; j=1..m_k\}/\{A_j^{(i)}; j=1..m_i\}$ with rank m_i+1. During unification of alternatives, provided by different IS, we should check whether alternatives are conceptually the same. If IS are experts, then for this purpose we can use the methods of semantic similarity determination, suggested in [20].

3) As a result, we will get M alternatives in each ranking. Each ranking R_i will include m_i alternatives with different ranks and (M – m_i) alternatives with the same rank (m_i+1). (A similar approach was mentioned by Dwork et al in [11]).

4) Add the ranks of each alternative across all IS.

5) Sort (order) the alternatives in the order of increment of individual rank sums. This sorting will result in the rank order that we should find.

$$\sum_{j=1}^{n} r_{kj} > \sum_{j=1}^{n} r_{lj} \Rightarrow r_k < r_l, k, l = 1..M \qquad (1),$$

where r_{kj} и r_{lj} – are the ranks of alternatives A_k and A_l in the ranking, provided by IS number j.

5 Usage of different methods under smaller cardinality of the set of alternatives

If the number of alternatives amounts to one or two dozens, or if the decision-maker (analyst) is interested only in the rank order of the first few alternatives, provided by IS, then it makes sense to use other ranking methods (beside or instead of Borda) in order to aggregate the individual ranking results.

For example, let us assume, we have 10-15 IS, and each of them provides a ranking of alternatives, from which we are particularly interested in the top 10-20 items. If all alternatives (items) are different, then we have 300 (15×20) different items at most. In fact (especially if we are talking about search engines), many items featured in individual rankings across different IS are the same. So, if we have 15 IS and each of them provides a ranking of 20 alternatives, then the cardinality of the set of unique items will amount to approximately 100 alternatives. Consequently, if we use aggregation methods operating with ordinal pair-wise comparison matrices (PCM) (such as Condorcet rule or Kemeny's median) and not with ranking vectors (like Borda or Markov Chain based

methods), we will have to analyze 15 square matrices containing 100×100 cells. As strict rank order relationship is reciprocally symmetrical (if alternative A_1 ranks higher than A_2, then A_2 ranks lower than A_1), during aggregation of PCM we will have to analyze just matrix elements above the principal diagonal $\{d_{ik},\ i<k\}$. That is why, even if the aggregate ranking features 100 alternatives, then we will have to analyze (100×99)/2 or 4950 cells in each PCM. These calculations demonstrate, that under smaller cardinality of alternative set, we can use Condorcet rule and Kemeny's median for aggregation of individual ranking results.

In our view, the disadvantage of Borda method is that during aggregation of individual rankings we are witnessing an explicit heuristic transition from ordinal preference scale to ratio scale. However, this transition is inevitable, because representation of alternative estimates in ratio scale is the necessary and sufficient condition of existence of linear global criterion, allowing us to aggregate these estimates (as proven by Litvak in [21]). For example, an alternative with rank 2 does not necessarily have to be exactly 2 times better than alternative with rank 4 (we should remind, that dominating alternatives are assigned smaller ranks).

Condorcet rule, Markov chains, and Kemeny's method allow us to avoid such an explicit transition from ranks to ratios.

6 Existing methods: a brief overview

Condorcet rule. The key idea of Condorcet method is as follows. If alternative A_i dominates over alternative A_k in the majority of individual rankings, then this preference relationship should be maintained in the global (aggregate) ranking. In order to get an aggregate ranking we should first build individual ordinal PCMs. If in some individual ranking R_j alternative A_i dominates over A_k, then the respective element of ordinal PCM equals 1, if A_i is dominated by A_k – (-1), if the alternatives are equal – 0:

$$d_{ik}^{(j)} = \begin{cases} -1, A_i \prec A_k \\ 0, A_i \equiv A_k \\ 1, A_i \succ A_k \end{cases} \qquad (2)$$

In order to aggregate rankings, provided by several information sources according to Condorcet's rule, we should limit the number of alternatives to be featured in the global ranking by a fixed number M and build ordinal PCM based on rankings, provided by all IS. $\left\{D^{(j)} = \{d_{ik}^{(j)}; i,k = 1..M\};\ j = 1..n\right\}$. After that we should calculate the sums

of all respective elements of individual PCM and fill ordinal PCM, corresponding to the global rank ordering relationship (tournament).

$$D=\{d_{ik};i,k=1..M\} \text{ where } d_{ik}=\begin{cases}0, \sum_{j=1}^{n} d_{ik}^{(j)}=0\\ 1, \sum_{j=1}^{n} d_{ik}^{(j)}>0\\ -1, \sum_{j=1}^{n} d_{ik}^{(j)}<0\end{cases} \quad (3).$$

After that we should add the elements in each line of the global rank ordering relationship matrix *D* and rank the obtained line sums.

$$\sum_{k=1}^{M} d_{ik} > \sum_{l=1}^{M} d_{lk} \Rightarrow r_i < r_l \quad (4),$$

where r_i and r_l are ranks of alternatives A_i and A_l in the global ranking *R*.

The disadvantage of Condorcet's rule (beside dimensionality limitation) is that equal alternative ranks may emerge and transitivity of the global preference relationship may be violated as a result of the so-called "Condorcet paradox". The paradox is one of the specific cases of violation of Arrow's requirements to social choice functions [7]. If feedback is acceptable, then transitivity of the global preference relationship can be achieved if we use algorithms, described in [22, 23].

In the context of aggregation of results of online data search, provided by several engines, the advantage of Condorcet's rule (mentioned by Dwork et al in [11]) is its ability to filter spam content.

A more "just" aggregation rule in view of Arrow's impossibility theorem is Kemeny's rule (sometimes called Kemeny's median).

Kemeny's median can be considered the analogue of average mean for ordinal (non-cardinal) estimates.

As ordinal estimates (or ranks) do not bear information on quantitative relation between alternatives, such concepts as Euclidian distance metric or center of mass (center of gravity) do not apply to rank order vectors.

Distance between two rankings of a given set of alternatives depends upon the number of elementary permutations that is needed to obtain one ranking from the other. Kemeny's distance between two rankings is based on Haming metric. If we have two rankings R_1, R_2 of a set of alternatives $A=\{A_1..A_m\}$, and respective domination matrices D_1 and D_2 are built based on these rankings according to formula (2), then Kemeny's distance *K* is calculated as follows.

$$K(R_1,R_2) = \sum_{i=1}^{m}\sum_{k=1}^{m}\left|d_{ik}^{(1)} - d_{ik}^{(2)}\right| \quad (5),$$

If we have a set of n rankings $\{R_1..R_n\}$ of alternatives $A=\{A_1..A_m\}$, then, by definition ([8]), Kemeny's median of this set of rankings is a ranking R:

$$R = \underset{R\in T}{arg\,min}\sum_{j=1}^{n} K(R,R_j) \quad (6),$$

where T is a set of all possible rankings of alternatives. Kemeny's median calculation procedure is very labor-intensive. Some estimates of complexity of this problem are provided in [11]. One of the "simplest" algorithms of Kemeny's median calculation is provided by Litvak in [21].

7 Generalization of suggested approaches to the case of different weights of information sources

Let us now assume that IS weights $\{w_1..w_n\}$ (reflecting their credibility) have been provided by experts, calculated based on previous estimation experience (as described in [15, 16]), or obtained in some other way. In this case, respective formulas for ranking aggregation will change, as all rankings, provided by individual IS, will be taken into consideration together with their respective weights. Formula (1) (corresponding to Borda aggregation method) will look as follows.

$$\sum_{j=1}^{n} w_j r_{kj} > \sum_{j=1}^{n} w_j r_{lj} \Rightarrow r_k < r_l, k,l = 1..M \quad (7)$$

Formula (3) (corresponding to Condorcet method) will look as follows.

$$D = \{ d_{ik}; i,k = 1..M \}$$

where

$$d_{ik} = \begin{cases} 0, \sum_{j=1}^{n} w_j d_{ik}^{(j)} = 0 \\ 1, \sum_{j=1}^{n} w_j d_{ik}^{(j)} > 0 \\ -1, \sum_{j=1}^{n} w_j d_{ik}^{(j)} < 0 \end{cases} \quad (8)$$

In formula (6) weight coefficients will become multipliers for respective Kemeny distance values.

$$R = \arg\min_{R \in T} \sum_{j=1}^{n} w_j K(R, R_j) \quad (9).$$

Extrapolation of Markov chain based methods, described in [11], represents a problem that should be addressed in a separate research.

Experience and numerous publications indicate that when a group of experts participates in decision-making process, individual expert competence should be taken into consideration (providing, the expert group is relatively small) (for details – see [24-26]). Similarly, if several IS are used to build a ranking of alternatives according to their relevance in terms of a particular information query or in the process of some analytical research, we should take their weights into account as well.

Now let us address the issue of calculation of IS weights in greater detail. So far we have come up with several conceptual approaches to IS weight calculation. These approaches are outlined in the next section.

8 Approaches to calculation of the relative weights of information sources

1) **Experience-based approach** was set forth in [15, 16]. In essence, experts or evaluators provide their ranking of alternatives and then, based on their ranking and rankings, provided by individual IS, the weights of IS are calculated (under assumption that rankings are aggregated using Borda or Condorcet rule).

2) **Definition of relative IS weights based on quantity and quality of provided information.** It is reasonable to assume, that the relative weight of an IS should depend on quantity and quality of information, provided by the source in terms of every information query (in the given subject domain). Criteria, representing quality and quantity of information search, following a given query can be formulated as follows.

- Number of alternatives (references or links in case of online information search), provided by the IS in response to a particular query;
- Relevance of these alternatives (references, links), i.e. their correspondence to the specific query.

Relevance of the results of functioning of an IS can be defined based on the estimates of users (evaluators, experts) from the given subject domain. These estimates of IS can be based on previous experience of information search queries.

Relevance of results of IS work can be considered in terms of search queries of some particular type (search for information in some given language, formulas, images etc.). If it is possible to outline some query type, expert estimate will depend upon IS capability of processing this particular type of queries. Otherwise (when it is problematic to outline specific query types) it makes sense to use an average value of expert estimate (across different query types).

In the context of this subsection we suggest calculating IS weight using the approach, similar to methods, used by Totsenko for calculation of expert competence in [27, 28]. In accordance to this approach, expert competence should depend on self-estimate, objective estimate, and mutual estimate.

$$k = s(x_1 b + x_2 v), \tag{10}$$

where k is the relative competence of the expert, s is self-estimate, b is an objective component, v is mutual estimate of expert group members, while x_1, x_2 represent the coefficients of relative importance of objective and mutual estimates' respectively.

When the weights of "generalized" IS (i.e. experts, search engines, documents, etc.) are calculated, *mutual estimate* can be replaced by the ration between the number of alternatives, provided by the *i*-th IS in relation to the total number of alternatives, provided by all IS V_i:

$$V_i = \frac{m_i}{\sum_{k=1}^{n} m_k}, \tag{11}$$

where m_i is the number of alternatives, provided by *i*-th IS, n is the total number of information sources.

Self-estimate in formula (10) can be replaced by the value of expert estimate E_i. This estimate depends on previous experience of IS usage.

For each particular query type (for instance, Ukrainian, English, image, formula, other) we should apply the respective average value of E_i. Definition of query types is not the subject of this paper and should be addressed separately.

Objective estimate in formula (10) can be replaced by the ratio between the number of alternatives, provided by *i*-th IS and the total number of unique alternatives, provided by all IS O_i (thus, it will reflect the ability of IS to provide unique information).

$$O_i = \frac{m_i}{P}, \tag{12}$$

where *P* is the number of unique alternatives, provided by all IS, that is $P = card(\bigcup_{i=1}^{n} \{A_j^{(i)}; j = 1..m_i\})$. If IS are experts, semantic similarity methods, suggested in [20] can be used to define whether the alternatives are relevant in the context of the given query.

Similarly to expert competence, weight of IS will be calculated as follows.

$$w_i^* = E_i(x_1 O_i + x_2 V_i), \tag{13}$$

where w_i^* is the non-normalized IS weight, , x_1, x_2 are the respective coefficients of components' relative importance, and E_i is the non-normalized expert estimate value. Normalized values are calculated as follows.

$$W_i = \frac{w_i^*}{\sum_{i=1}^{n} w_i^*}, \tag{14}$$

where W_i is the normalized relative IS weight.

We propose to follow the assumption that "objective" and "mutual" IS estimates should form a convex combination ($x_1 + x_2 = 1$), so x_1 and x_2 should lie within the [0;1] range and vary depending on particular information query.

As a result of each information search, every IS provides several alternatives (in the general case they are different, as mentioned above).

Figure 1 represents an example, where 5 IS provide 9 unique alternatives. 1st and 6th alternatives were provided by 2 IS, 2nd and 5th – by 3 IS, 3rd alternative – by 5 IS, while other alternatives were provided by 1 IS each.

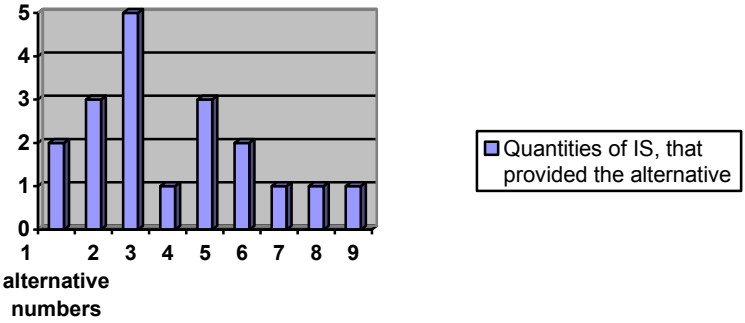


Fig. 1 Example of a unified set of alternatives, provided by several IS

In order to illustrate the behavior of dependence of x_1 and x_2 on information search results, let us consider several extreme cases.

Extreme case 1

Let us assume that all IS provided completely different alternative sets. Generalized information search results are displayed on Fig. 2: 5 IS provide 9 different alternatives and each of the alternatives is provided by a single IS (for instance, IS1: a1, a2; IS2: a3, a4; IS3: a5; IS4: a6, a7, a8; IS5: a9).

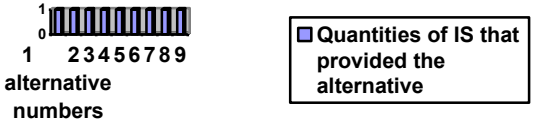


Fig. 2 Example of a unified set of alternatives, where each alternative is provided by a single IS

If all IS provide different alternatives in response to some query, the relevance of these alternatives seems questionable. In such cases we suggest using expert estimate (based on previous experience of IS usage)

as the key component of IS weight, as it reflects the experience of previous information search sessions and the ability of IS to find new information.

Extreme case 2

Let us assume, that all IS provided the same alternatives. An example of a hypothetical information search results are presented on fig. 3: 5 9 alternatives are provided by all 5 IS (although the order, in which alternatives are ranked by different sources, may be different).

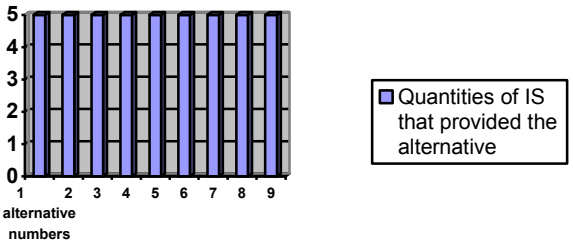


Fig. 3 Example of a unified set of alternatives, where each alternative is provided by all IS

In such a case, IS weights should be equal to E_i, because all IS are equally efficient in providing alternatives in response to information query. Their weights can be defined, again, only based on the previous history of information search sessions that is reflected by the expert estimate.

In order to devise a formal expression for x_1 and x_2, based on these intuitive considerations, let us introduce the indicator ρ, characterizing the alternative frequency function or density of representation of alternatives across all information sources.

$$\rho = \frac{\sum_{j=1}^{P} h_j}{nP}, \qquad (15)$$

where h_j is the quantity of IS, that provided *j*-th alternative, while *n* is the total quantity of IS.

In "extreme case 1" ρ assumes the minimum value $\rho = \frac{1}{n}$.

In "extreme case 2" ρ assumes maximum values $\rho = 1$.

In view of above-mentioned convexity requirements, we suggest calculating x_2 as

$$x_2 = \rho;\ x_1 = 1 - x_2. \quad (16)$$

In "extreme case 1" the normalized relative IS weight assumes the value of $w_i = \frac{E_i((1-\frac{1}{n})O_i + \frac{1}{n}V_i)}{\sum_{j=1}^{n} E_j((1-\frac{1}{n})O_j + \frac{1}{n}V_j)}$, and in "extreme case 2" it assumes the value of $w_i = E_i$.

3) <u>Statistical approach</u> represents a modification of the previous approach. In order to take the quantity of information, provided by the IS ("objective" weight component in formula (13)) into account, we can normalize the numbers of alternatives across all IS (see formula (11)) and then calculate the importance of the respective component (i.e. value of x_1) as dispersion if this indicator (normalized number of alternatives).

$$x_1 = D(V);\ x_2 = 1 - x_1\ . \quad (17)$$

In this case, if all IS provide the same number of alternatives, the respective component of their relative weights can be neglected as it does not vary across different IS. The respective component will only come into play when the numbers of alternatives across IS are different.

9 Conclusions

Several approaches to aggregation of alternative rankings provided by several individual IS have been described. Applicability of this or that approach depends on the specificity of IS (which can be represented by experts, analysts, search engines, online or paper documents etc.) and on the number of alternatives in the rankings. Several ways of IS weight definition have been suggested, based on heuristic expert decision support and statistical methods. IS weights calculation methods can be applied to aggregation of data, coming from IS of different nature.

Although the approaches, suggested in the paper are, to a large extent, heuristic, they are based on the fundamental assumption, that during aggregation of information, coming form different sources, relative weights of IS should depend on quality and quantity of information,

provided by these sources and, beside that, reflect expert estimate of their credibility, based on previous experience of their usage.

Further research envisions experimental study of the suggested methods.

This paper is prepared as part of project #F73/23558 “Development of Decision-making Support Methods and Means for Detection of Information Operations”. The project won the contest #F73 for grant support of scientific research projects held by The State Fund for Fundamental Research of Ukraine and Belarusian Republican Foundation for Fundamental Research.

References

1. Saaty, T.L., 1994. Fundamentals of Decision Making and Priority Theory with The Analytic Hierarchy Process. / T.L.Saaty; RWS Publications, Pittsburgh PA, 204220.
2. Горбулін В.П., Додонов О.Г., Ланде Д.В. Інформаційні операції та безпека суспільства: загрози, протидія, моделювання: монографія – К., Інтертехнологія, 2009 – 164 с.
3. Kadenko S.V. Prospects and Potential of Expert Decision-making Support Techniques Implementation in Information Security Area / S.V.Kadenko // CEUR Workshop Proceedings – 2016 – pp. 8-14.
4. Rzążewski, K. Każdy głos się liczy! Wędrówka przez krainę wyborów / K.Rzążewski, W.Słomczyński, K.Życzkowski - Wydawnictwo Sejmowe Warszawa 2014 – 420 p.
5. J. C. Borda. Memoire sur les elections au scrutin. Histoire de l'Academie Royale des Sciences, 1781.
6. M.-J. Condorcet. Essai sur l'application de l'analyse a la probabilite des decisions rendues a la pluralite des voix, 1785.
7. Arrow, K. J. Social Choice and Individual Values / K.J. Arrow – 2nd edition. Yale University Press – 1970 – 144 p.
8. Kemeny, J. G. Mathematics without Numbers / John G. Kemeny // Daedalus Vol. 88, No. 4, Quantity and Quality (Fall, 1959), pp. 577-591.
9. Copeland A. H. A reasonable social welfare function. / A. H. Copeland. Mimeo, University of Michigan, 1951.
10. Aleskerov, Fuad T. Arrovian Aggregation Models / Fuad T. Aleskerov. Springer, 1999 – 226 p.
11. Dwork C. Rank aggregation methods for the Web / C. Dwork, R. Kumar, M. Naor, D. Sivakumar // WWW '01 Proceedings of the 10th international conference on World Wide Web – 2001 – Pages 613-622.

12. Renda M.E. Web metasearch: rank vs. score based rank aggregation methods / M.E. Renda, U. Straccia // Proceedings of the 2003 ACM symposium on Applied computing – 2003 – Pages 841-846.
13. Аборок, Ш.К. Метапоисковая система с использованием показателя обобщенной релевантности документа для формирования результатов поиска / Ш.К. Аборок, В.И. Костин // Информатика и компьютерные технологии-2005: сб. тр. 1-ой Междунар. студ. науч.-техн. конф., 15 дек. 2005 г. - Донецк, 2006. - С. 268-269.
14. Totsenko V.G. Method of Determination of Group Multicriteria Ordinal Estimates with Account of Expert Competence / V.G. Totsenko // Journal of Automation and Information Sciences. – 2005. – vol. 37. -N10. - P.19-23.
15. Kadenko S.V Defining the Relative Weights of the Alternatives Estimation Criteria Based on Clear and Fuzzy Rankings / S.V. Kadenko // Journal of Information and Automation Sciences – 2013 – vol. 45, i. 2 – pp. 41-49.
16. Kadenko S.V. Determination of Parameters of Criteria of "Tree" Type Hierarchy on the Basis of Ordinal Estimates / S.V.Kadenko // Journal of Information and Automation Sciences – 2008 – vol. 40, i.8 – pp. 7-15.
17. Ramanathan, R. Group Preference Aggregation Methods Employed in AHP: An Evaluation and Intrinsic Process for Deriving Members' Weightages / R.Ramanathan, L.S.Ganesh// European Journal of Operational Research 79 – 1994, pp. 249-265.
18. Forman, E.Aggregating individual judgments and priorities with the analytic hierarchy process/ Forman, E. & Peniwati, K // European Jouranl of Operational Research, Vol. 108, 1998 – pp.131-145
19. 19)Yang Q, Du P-a, Wang Y, Liang B. A rough set approach for determining weights of decision makers in group decision making. PLoS ONE 12(2) – 2017: e0172679. doi:10.1371
20. Андрійчук О.В. Метод змістової ідентифікації об'єктів баз знань систем підтримки прийняття рішень / О.В. Андрійчук // Реєстрація, зберігання і обробка даних. — 2014. — Т. 16, № 1. — С. 65-78.
21. Литвак Б. Г. Экспертная информация. Методы получения и анализа / Б. Г. Литвак. – М.: Радио и связь, 1982. – 185 с.
22. Tsyganok V.V. On Sufficiency of the Consistency Level of Group Ordinal Estimates / V.V.Tsyganok, S.V.Kadenko // Journal of Automation and Information Sciences. – 2010. – v.42, issue 8.– p.42-47.
23. Kadenko S.V. A Method for Improving the Consistency of Individual Expert Rankings during Their Aggregation /

S.V.Kadenko, V.V.Tsyganok // Journal of Automation and Information Sciences. – 2012. – v.44, issue 4. – p.23-31.
24. Циганок В.В. Урахування компетентності експертів при визначенні групового ранжирування / В.В. Циганок, О.В. Андрійчук // Реєстрація, зберігання і обробка даних. – 2011. – т.13. – №1. С. 94-105.
25. Tsyganok V.V. Simulation of Expert Judgements for Testing the Methods of Information Processing in Decision-Making Support Systems / V.V.Tsyganok., S.V.Kadenko, O.V.Andriichuk // Journal of Automation and Information Sciences. – 2011. – v.43, issue 12. – p.21-32.
26. Tsyganok, V. Significance of expert competence consideration in group decision making using AHP. / V. Tsyganok, S. Kadenko, O.Andriichuk // International Journal of Production Research. Vol. 50, Issue 17, 2012 – pp. 4785-4792
27. Тоценко В.Г. Методы и системы поддержки принятия решений. Алгоритмический аспект / В.Г. Тоценко. – К.: Наук. думка, 2002. – 382 с.
28. Totsenko V.G. Matching and Aggregation of Experts Estimates Taking into Account Experts' Competence while Group Estimation of Alternatives for Decision Making Support / V.G. Totsenko // Journal of Automation and Information Sciences. – 2002.– v.34. – No. 8

СОДЕРЖАНИЕ

Национальная академия наук Украины
Институт проблем регистрации информации

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ И БЕЗОПАСНОСТЬ

Материалы XVII Международной научно-практической конференции

Выпуск 17

Підп. до друку 7.12.2017. Ум. друк. арк. 12,65. Обл.-вид. арк. 23,66.
Наклад 100 пр. Зам. № 15-200.

ТОВ "Інжиніринг"
ISBN 978-966-2344-59-2