НАЦІОНАЛЬНА АКАДЕМІЯ НАУК УКРАЇНИ
ІНСТИТУТ ПРОБЛЕМ РЕЄСТРАЦІЇ ІНФОРМАЦІЇ НАН УКРАЇНИ

ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ І БЕЗПЕКА

МАТЕРІАЛИ XX МІЖНАРОДНОЇ НАУКОВО-ПРАКТИЧНОЇ КОНФЕРЕНЦІЇ

ВИПУСК 20

Київ – 2020

Рекомендовано до друку Вченою радою
Інституту проблем реєстрації інформації НАН України
(протокол № 12 від 15 грудня 2020 р.)

Інформаційні технології і безпека. Матеріали XX Міжнародної науково-практичної конференції ІТБ-2020. – Київ: Інжиніринг. – 174 с. ISBN: 978-966-2344-77-6

До збірника увійшли матеріали доповідей, представлених на XX Міжнародній науково-практичній конференції «Інформаційні технології і безпека» (ІТБ-2020, 10 грудня 2020 року, м. Київ, Україна).

У збірнику представлені матеріали, присвячені питанням створення і впровадження інформаційних технологій, актуальним проблемам забезпечення інформаційної та кібербезпеки, протидії інформаційним операціям і кібертероризму, проведенню аналітичних досліджень на основі аналізу контенту мережі Інтернет.

Для фахівців в області інформаційних технологій, інформаційної і кібернетичної безпеки, а також для аспірантів і студентів старших курсів вищої школи відповідних спеціальностей.

ISBN 978-966-2344-77-6

ДИНАМІЧНА РЕКОНФІГУРАЦІЯ В АВТОМАТИЗОВАНИХ СИСТЕМАХ ОРГАНІЗАЦІЙНОГО УПРАВЛІННЯ

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

У доповіді розглядаються питання підвищення функціональної стійкості автоматизованих систем організаційного управління (СОУ) із застосуванням методу динамічної структурно-функціональної реконфігурації. Обговорюються методологічні аспекти розробки і впровадження процедур реконфігурації в автоматизовані СОУ, що функціонують в умовах перманентних змін зовнішнього середовища. Обґрунтовується доцільність створення спеціалізованого комп'ютерного моделюючого комплексу при автоматизації СОУ для відпрацювання базових системних, конструкторських і технологічних рішень, зокрема, процедур реконфігурації. На спеціалізованому моделюючому комплексі можливо напрацювати шаблони дій управлінців в умовах появи небажаних змін внутрішнього і зовнішнього середовища функціонування СОУ та розвитку нештатних ситуацій на об'єктах управління.

Ключові слова: автоматизована система організаційного управління, реконфігурація, комп'ютерний моделюючий комплекс.

Вступ

Автоматизовані системи організаційного управління (СОУ) являють собою складні системи збору, аналізу та опрацювання інформації стосовно об'єкта управління, його внутрішнього середовища і взаємодії із зовнішнім середовищем [1], що мають соціальну складову, елементи якої є активними підсистемами із власними локальними цілями, власним уявленням про навколишнє середовище і модель управління. СОУ за своїм призначенням повинні адекватно визначати необхідний рівень безпеки об'єктів управління, забезпечувати керованість об'єктів і досягнення визначеної в СОУ цілі.

Прийняття обґрунтованих управлінських рішень сьогодні потребує опрацювання великої кількості даних, які часто слабо структуровані або взагалі неструктуровані, отримані з різних джерел, від різних датчиків, контролерів і т. ін. Так, у 2019 році із необхідністю аналізу великих обсягів даних (потоків даних) стикнулись більше 50 % компаній по всьому світу, хоча у 2015 році таких компаній було лише 17% [2].

Динамічність і непередбачуваність середовища функціонування, складність і неоднозначність процесів, що мають місце на об'єктах управління, зростання ризиків неотримання бажаного результату, багатофакторність, що призводить до неоднозначності розуміння

інформації особами, які приймають рішення, в умовах відсутності достовірного знання про динаміку ситуації на об'єкті управління потребує впровадження у практику керування модельно-алгоритмічних способів і підходів до забезпечення необхідного рівня ефективності функціонування об'єкта управління і власне самої СОУ. Функціональні відмови, що можуть статись і в самій автоматизованій СОУ, іноді загрожує переходом у небезпечний стан не лише об'єкта управління, а й порушенням безпеки функціонування деякої критичної інфраструктури.

Реконфігурація системи – це, у загальному випадку, процес зміни структури, параметрів, технологій функціонування для відновлення до потрібного рівня показників працездатності і ефективності системи або забезпечення мінімального зниження рівня цих показників в умовах функціональної деградації системи. Практика свідчить, що технологія реконфігурації серед засобів підвищення функціональної стійкості складних систем займає одне з першорядних місць.

Реконфігурація в системах організаційного управління: призначення, задачі, моделювання

Автоматизація СОУ забезпечує автоматизацію вирішення функціональних задач на різних рівнях управління, наявна сукупність математичних методів, моделей, алгоритмів, програмно-технічних засобів і спеціальних програмних модулів, що реалізують функціональні підсистеми, комплекси задач, функціональні задачі та АРМ посадових осіб СОУ та персоналу. Автоматизована СОУ характеризується наступними особливостями:

- наявністю функціональних підсистем різної природи, складності та призначення, що відтворюють процеси управління підрозділами об'єкта і забезпечують, взаємодіючи між собою, управління об'єктом в цілому;
- наявністю інтенсивних потоків інформації, різноманітної й неоднорідної за складом, призначенням, способом кодування тощо;
- широким діапазоном зміни станів об'єкту управління та високою динамічністю зміни цих станів;
- високим рівнем автоматизації задач, що вирішуються в процесі функціонування;
- одночасним вирішенням різних функціональних комплексів задач, пов'язаних між собою;
- функціонуванням у реальному масштабі часу.

До неможливості виконання комплексів функціональних задач на АРМ посадових осіб всіх рівнів управління і порушення процесу управління об'єктом в цілому можуть призвести:

- непрацездатність АРМ чи втрата зв'язків між ними внаслідок їх фізичного руйнування або порушення цілісності даних, відсутності посадової особи;
- погіршення технічних характеристик комп'ютерної системи (швидкості, продуктивності, пропускної спроможності та ін.);

- спотворення алгоритмів функціонування;
- зменшення структурної надлишковості, запасу ресурсів;
- погіршення функціонування елементів і керованості автоматизованої СОУ;
- фатальна втрата працездатності елементів або системи в цілому.

Загальний стан системи управління вимірюється параметрами та визначається за індикаторами, які характеризують: наявність та функціонування програмно-апаратних ресурсів системи; наявність та функціонування посадових осіб-користувачів АРМ та їх спроможність виконувати свої завдання; відповідність поточного процесу вирішення функціональних задач управління до заданого плану; дотримання планових часових показників вирішення функціональних задач управління; отримані результати (у порівнянні з нормативними або граничними значеннями); зовнішнє та внутрішнє середовище системи.

Розробка і впровадження автоматизованої СОУ являє собою складний ітераційний процес, для реалізації якого доцільно використовувати комп'ютерні моделюючі комплекси (КМК), де розгортати інструментарій для розв'язання завдань розробки, дослідження та налаштовування автоматизованих робочих місць (АРМ) посадових осіб СОУ та персоналу, відпрацювання елементів автоматизації управління, проведення навчання і тренажу посадових осіб та персоналу. Засоби КМК мають дозволяти напрацювати базові системні, конструкторські, технічні й технологічні рішення, проводити апробацію проектних рішень, програмного, інформаційного забезпечення. Надати можливість дослідити фактори впливу на СОУ як ззовні, так і з внутрішнього середовища, моделювати зміни у функціонуванні СОУ і об'єкта управління, розробляти засоби запобігання переходам їх у небажані стани, відпрацював процедури компенсації відмов, реконфігурації (гнучкого перерозподілу в СОУ виконуваних задач і функцій між працездатними або частково працездатними АРМ, наявних ресурсів і засобів).

В КМК можна застосовувати різноманітні системи та засоби моделювання, імітації управлінських процесів, регламентів взаємодії, провести системно впорядкування формалізованих відображень процесів функціонування об'єкта управління, розробити спеціалізовані засоби для навчання та тренажу персоналу і посадових осіб. Гнучке й розширюване середовище моделювання в КМК дозволяє моделювати функції управління максимально наближено до наявної практики прийняття і реалізації рішень у конкретній предметній сфері.

Архітектура КМК залежить від обраної моделі предметної сфери, моделей систем і процесів, які задіяні у реалізаціях конкретних завдань управління. Кожна функція СОУ, що виконується при реалізації певного управлінського процесу, моделюється на КМК як окрема функціональна задача. Вхідні дані для процесу можуть бути або початковим управлінським впливом, або вихідними чи проміжними даними іншого управлінського процесу. Виконання процесу управління передбачає:

- підготовку і напрацювання рішень щодо упорядкування дій, необхідних для виконання функціональної задачі, у деяку послідовність операцій, що реалізуються у рамках відповідної технології;
- визначення, які люди (співробітники), у який час, які технологічні процеси (операції) виконують для того, щоб отримати спільний бажаний кінцевий результат (досягти загальносистемної цілі функціонування СОУ).

Виконання технологічного процесу потребує наявності в автоматизованій СОУ не лише фахівців з відповідним рівнем кваліфікації, а й технічних засобів, методик та інструкцій по їх застосуванню, програмно-технічного, інформаційного та іншого забезпечення, необхідного і достатнього для виконання набору завдань у конкретній предметній сфері.

Функціональна структура КМК, як правило, є проекцією управлінської структури, що керує виконанням певного бізнес-процесу, із збереженою ієрархією (підпорядкованістю), системою взаємодій та регламентованостей. Модель СОУ являє собою впорядковану сукупність управлінських процесів із узгодженням необхідних форм документів й структур баз даних (інформаційним забезпеченням), що вимагає документування у конкретній предметній сфері [3].

В автоматизованій СОУ у якості функціональних компонентів виступають автоматизовані робочі місця управлінців (АРМ). Кожний АРМ є підсистемою, структура якої визначається її функціональним призначенням. Спеціалізація АРМ відбувається шляхом інсталяції відповідного програмного забезпечення і налагодження зв'язків між складовими системи. Модульний принцип розробки програмного забезпечення дозволяє досить легко сформувати необхідну конфігурацію АРМ, як підсистеми СОУ, для виконання певних управлінських функцій. Функціональність АРМ можна розширювати у разі потреби, підключаючи нові програмні модулі. Таким чином створюється гнучке масштабоване середовище реалізації управлінських функцій.

Впровадженню технології динамічної конфігурації в автоматизованих СОУ має передувати оцінка доцільності цього на основі аналізу витрат ресурсів і досягнутих змін показників функціональної стійкості автоматизованої СОУ. Головним є забезпечення безперервності процесу управління, який реалізується виконанням комплексу завдань $\Phi = (\phi_1, \phi_2, ..., \phi_n)$ на АРМах посадових осіб-користувачів АРМ у наявних умовах функціонування.

Для раціонального перерозподілу функцій між працездатними компонентами автоматизованої СОУ необхідно врахувати поточні характеристики вирішуваних задач і функцій, що виконуються на об'єктах, аналізувати і оцінювати поточний стан інфраструктури взаємодії АРМ; здійснювати оперативні розрахунки, оцінювання і аналіз інформаційно-технічних можливостей СОУ. Реконфігурація в СОУ – це не лише технологічне рішення щодо компенсації відмов, а й управлінський процес по забезпеченню оперативного перерозподілу функцій управління і необхідних

ресурсів між АРМами посадових осіб для досягнення необхідної ефективності функціонування СОУ в цілому по управлінню об'єктом.

Розробка і впровадження технології динамічної реконфігурації в автоматизовану СОУ передбачає формулювання і вирішення із застосуванням засобів КМК наступних задач:

- моделювання можливих ситуації ураження об'єктів управління і компонентів СОУ для виявлення тих, вплив на які найбільш значущий для процесів управління;
- оцінка доцільності застосування однієї з трьох можливих стратегій зниження значень ризиків для управління об'єктом, а саме:
  - зменшення ймовірності виникнення небажаної події,
  - зменшення збитків, що матимуть місце у разі небажаної події,
  - одночасне зменшення ймовірності виникнення і величини збитків від небажаної події;
- побудова і моделювання сценаріїв динамічної реконфігурації СОУ при виникненні небажаної ситуації на об'єкті управління чи наявності різноманітних деструктивних впливів, що можуть призвести до порушень у функціонуванні об'єкта управління;
- розробка і відпрацювання технології динамічної ре конфігурації автоматизованої СОУ в умовах динамічної зміни стану об'єкта управління;
- імітаційне моделювання реалізації різних сценаріїв реконфігурації СОУ для навчання і тренажу персоналу і посадових осіб.

Основою для побудови комп'ютерної моделі структурно-функціональної реконфігурації може слугувати узагальнена теоретико-множинна модель реконфігурації складної системи, запропонована у [4], у вигляді математичної структури вибору:

$$\left(Q(s,w), \Delta, \left\{ \mathrm{r}_i^{\alpha}(w) \right\}_{i \in \Gamma}, \left\{ \mathrm{r}_j^{\beta}(w) \right\}_{j \in \Gamma_1}, \left\{ \mathrm{F}^k(w) \right\}_{k \in \Gamma_2}, \Omega = \{w\} \right),$$

де $Q(s,w)$ - деяка вихідна структура типу S, яка визначає тип моделі (статична, динамічна, математична, логіко-алгебраїчна, детермінована, з невизначеністю тощо);

Δ – множина альтернатив (планів, способів) структурно-функціональної реконфігурації, серед яких відбувається вибір;

$\left\{ \mathrm{r}_i^{\alpha}(w) \right\}_{i \in \Gamma}$ – множина α відношень, обмежуючих вибір, які вводяться безпосередньо при постановці задач вибору і відбивають основні просторово-часові, технічні і технологічні обмеження, пов'язані з процесом функціонування складної системи;

$\left\{ \mathrm{r}_j^{\beta}(w) \right\}_{j \in \Gamma_1}$ – множина β відношень переваг, що задаються на $\Delta \times \Omega$ і характеризують різні переваги при виборі раціонального рішення;

$\left\{\mathrm{F}^k(w)\right\}_{k\in\Gamma_2}$ – множина узгоджуючих правил, які дозволяють визначити результуюче відношення переваги задачі структурно-функціональної реконфігурації системи;

$\Omega=\{w\}$ – множина подій (множина невизначеності).

Така модель надає широкі можливості для різноманітних постановок задач реконфігурації складної системи. У якості вихідної структури $Q(s,w)$ при моделюванні реконфігурації автоматизованої СОУ можуть виступати відповідна організаційна структура, технологічна (функціональна) структура автоматизованої СОУ, структура програмно-математичного та інформаційного забезпечення, технічна структура тощо. Обмеження на вибір альтернатив залежатимуть від важливості виконуваних функціональних завдань і функціонального навантаження АРМ посадових осіб та персоналу СОУ, рівня повноважень, доступності даних про поточний стан об'єкта управління, внутрішнього і зовнішнього середовища функціонування тощо. У постановку задачі бажано привнести інформацію, яка б дозволила зняти критеріальну та модельну невизначеність і звести задачу з невизначеними факторами до її детермінованого варіанту [4]. Моделювання динамічної реконфігурації автоматизованої СОУ на КМК дозволяє відпрацювати не лише процеси, що реалізуються на основі функціонального резервування чи побудови відмовостійких розкладів виконання задач для множини працездатних станів системи, а й ті, що враховують структурну і функціональну розподіленість системи і дозволяють задіяти реальні й віртуальні ресурси для виконання управлінських функцій.

Висновки

Досвід розробки, впровадження і супроводу автоматизованих СОУ свідчить, що всі базові системні, конструкторські, програмні й технологічні рішення для створюваної СОУ мають бути попередньо апробовані, а управлінці мають пройти етап навчання та тренажу.

Налаштування та апробацію АРМ посадових осіб та персоналу доцільно проводити на спеціалізованому КМК, де можна змоделювати притаманне відповідній професійній сфері управлінське завдання.

У середовищі КМК можна не лише відпрацювати основні процеси управління конкретним об'єктом, а й провести вибір й апробацію практично придатних процедур динамічної реконфігурації СОУ з наочністю результатів і врахуванням існуючої системи підпорядкованості й взаємодії в СОУ. Можливим стає побудова послідовності процедур реконфігурацій, яка може бути використана у разі критичності часу.

При відпрацюванні на КМК процесів виконання функціональних завдань та взаємодії між посадовими особами-користувачами АМР можливий перехід від інтуїтивних оцінок до кількісних, формалізація

досвіду експертів, що об'єктивізує управлінські рішення і сприятиме підвищенню їх якості.

Напрацьований при розробці АРМ СОУ аналітичний ресурс, формалізовані методики підтримки безпеки функціонування критичних інфраструктур дозволять підвищити оперативність і обґрунтованість управлінських рішень, наочність результатів управління навіть в умовах змін в системі підпорядкованості й взаємодії в СОУ , що викликані небажаними змінами у функціонуванні критичної інфраструктури.

Список літератури

1. Горбачик О.С. Технологія реконфігурації у системах організаційного управління критичних інфраструктур //*Реєстрація, зберігання і обробка даних:* зб. наук. праць за матеріалами Щорічної підсумкової наукової конференції 16-17 травня 2018 року ІПРІ НАН України. Київ, 2018, С. 105-108.
2. Аналитика Big Data – реалии и перспективы в России и мире [Електронный ресурс]. – Режим доступу: https://habr.com/ru/company/mailru/blog/449370/
3. Додонов О.Г. Комп'ютерне моделювання процесів організаційного управління // Вісник НАН України, 2016.- №1.- С.69-77.
4. Павлов А.Н. Комплексное моделирование структурно-функциональной реконфигурации сложных объектов //SPIIRAS Proceedings. 2013. Issue 5(28). P. 143-168. ISSN 2078-9181 (print), ISSN 2078-9599 (online), www.proceeding.spiiras.nw.ru

МОДЕЛЮВАННЯ ЖИВУЧОСТІ МЕРЕЖЕВИХ СТРУКТУР

О.Г. Додонов[0000-0001-7569-9360], Д.В. Ланде[0000-0003-3945-1178]
Інститут проблем реєстрації інформації НАН України, Київ, Україна

dodonov@ipri.kiev.ua, dwlande@gmail.com

У роботі моделюється і досліджується структурна живучість систем. Вводиться пороговий критерій живучості системи, що залежіть від розміру максимальної зв'язаної компоненти мережевої моделі після деструктивного впливу на неї. Цей показник складніше показника канонічної структурної живучості, яка враховує тільки порушення зв'язності мережі. Досліджується стан мереж з різною топологією при видаленні їх елементів (зв'язків). Очевидно, що введений показник залежить від топології мережі і її розмірів, разом з тим, добре наближається кубічними поліномами.

Ключові слова: Моделювання живучості, Модель мережі, Показник живучості.

Введення

До найважливіших фундаментальних властивостей систем відноситься живучість, під якою розуміють її здатність адаптуватися до нових умов функціонування, протистояння небажаним впливам при реалізації основної цільової функції. Ця робота присвячена моделюванню структурної живучості систем. Одне із завдань цієї роботи - дати чітку кількісну оцінку живучості.

Структурна живучість розглядається як властивість системи зберігати свою функціональність при пасивній протидії пошкодженням окремих елементів. Критерій структурної надійності - число елементів, що відмовили, при тому, що не порушуються працездатність системи [1].

1 Загальноприйняті моделі

Схема функціонування складної системи може бути задана за допомогою мережі, сукупності вузлів і зв'язків, яка визначає фізичну структуру цієї системи.

У загальноприйнятих моделях [2] вважається, що видалення всіх зв'язків, інцидентних деякому вузлу, ізолює його, перериваючи всі шляхи до інших вузлів – мережа стає незв'язною, живучість мережі – рівною нулю.

У цій роботі приймається інший критерій, пороговий, а саме, розглядається розмір найбільшою зв'язковий компоненти мережі. Зв’язність всієї мережі може бути порушена, але система залишається функціонально здатною, якщо відповідна максимальна зв'язкова компонента за обсягом (кількістю вузлів) незгірш від деякого заданого заздалегідь порога.

У роботах [3], [4] пропонується канонічне визначення коефіцієнта живучості, при якому ніяке руйнування зв'язків (ребер графа) не приводить до втрати зв'язності.

Канонічна живучість мережі *R(G, p)* визначається як ймовірність того, що граф мережі *G* залишиться зв'язним після того, як кожний зв'язок (ребро) буде видалений з однаковою ймовірністю *p*. *R(G, p)* можна розрахувати за допомогою перерахування остовів *G*. На практиці канонічна живучість тісно пов'язана з многочленом Татте-Уїтні (Tutte-Whitney) [5], який є інваріантом графа, що описує його комбінаторні властивості.

При цьому точний розрахунок канонічної живучості системи являє собою *NP*-складну задачу, витрати на вирішення якої зростають експоненційно з ростом числа вузлів і зв'язків, так як для розрахунку живучості полінома графа, що складається з n ребер, необхідно пройтися по всіх остовних підграфів графа *G* [6]. Тому в багатьох роботах пропонуються альтернативні наближені підходи для оцінки живучості систем [7].

2 Метод, що пропонується

На відміну від традиційного підходу в даній роботі пропонується підхід, який базується на імітаційному моделюванні:

1. Модель системи – граф, що складається з вузлів і зв'язків (ненаправлених) $S = (V, E)$. Вузли – однорідні функціональні компоненти.

2. «Потужність» системи – кількість вузлів в найбільшій зв'язковий компоненті V_s.

3. Система знаходиться в стані «жива», функціонально здатна, якщо питома «потужність» системи не менше деякого порогу τ, тобто $|V_s|/|V_o| \geq \tau$, де V_o – первинний розмір мережі.

4. На зв'язки мережі проводиться деструктивний вплив. Кожна зв'язок може бути видалена з ймовірністю p.

Для кожної конкретної системи можна визначити міру живучості системи при заданому порозі τ, тобто ймовірність видалення окремих елементів (зв'язків) $p*$, при якій система виходить із стану «жива», тобто $|V_s|/|V_o| < \tau$.

3 Еталонні мережі

Для моделювання як приклад досліджуються три артефактних мережі, а саме, мережі Барабаші-Альберт, Ердеша-Ренї та Уаттса-Строгатца. Ці мережі можуть розглядатися як прототипи багатьох реальних мереж.

3.1 Мережа Барабаші-Альберт: модель «переважного приєднання»

Більшості артефактних мереж притаманний степеневий закон розподілу. Такий розподіл пояснюється ефектом переважного приєднання (preferential attachment). До таких мереж відносяться мережі Барабаш-Альберта

(Barabási, Albert). Для побудови цих мереж використовується спеціальна процедура, яка полягає в тому, що до початку малій кількості вузлів поступово додаються нові вузли, зв'язки від яких з більшою вірогідністю приєднуються до тих вузлів, у яких зв'язків більше [8]. Модель переважного приєднання Барабаші-Альберт реалізована, зокрема, на мові R в пакеті igraph [9].

3.2 Мережа Ердеша–Ренї

Мережу Ердеша–Ренї (P. Erdős, A. Rnyi) [10] можна побудувати, розподіливши випадковим чином M зв'язків між N вузлами. Її іноді викликають моделлю пуассонівського випадкового графа (Poisson random graph model) через пуассонівського розподілу ступенів при $N \to \infty$, або іноді просто моделлю випадкового графа (random graph model). Еквівалентна модель, в якій значення кількості ребер M замінюється у відповідності із ймовірністю p появи нового ребра в графі. Модель випадкового графа реалізована на мові R в пакеті igraph.

3.3 Мережа «малого світу» (Small World)

Д. Уаттс і С. Строгатц (D.J.Watts, S. Strogatz) виявили феномен, характерний для багатьох реальних мереж, названий ефектом малих світів (Small Worlds) [11]. Ними була запропонована процедура побудови наочної моделі мережі, якій притаманний цей феномен. Ця модель являє собою одномірну регулярну решітку, що складається з N вузлів, де кожен вузол з'єднаний тільки зі своїми 4 найближчими сусідами і накладені періодичні граничні умови – решітка згорнута в кільце. Потім виконується така процедура: з імовірністю p відбувається перемикання (rewiring) невеликої кількості зв'язків (ребер), в ході якого вони видаляються і замінюються іншими зв'язками, що з'єднують два випадково обраних вузла. У початковому стану цієї мережі – вона регулярна – кожен вузол якої з'єднаний з чотирма сусідніми. Потім у цій мережі деякі «ближні» зв'язку випадковим чином замінені «далекими» – саме в цьому стані виникає феномен «малих світів» (при $p \in (0.01,\ 0.1)$). При подальшому збільшенні p утворюється мережа, яка за властивостями близька до випадкової мережі Ердеша-Ренї.

4 Модель

Було проведено імітаційне моделювання процесу руйнування трьох мереж, а саме, Барабаші-Альберт, Ердеша-Ренї, Уаттса-Строгатца. Моделювання проводилося на мові програмування R з використанням бібліотеки igraph. Результати моделювання наведені у Таблиці 1 і на Рис. 1-3.

Якщо задати поріг руйнування, наприклад, наступним чином, мережа функціональна, якщо розмір найбільшої зв'язної компоненти V_s становить 0.2 від початкового розміру мережі V_0, тобто: $|V_s|/|V_o| \geq \tau,\ \tau = 0.2$ то, відповідно, отримуємо значення порогової ймовірності $p*$ для мереж:

Ердеша-Рен`ї: ≈0.8
Уаттса-Строгатца: ≈0.7
Барабаші-Альберт: ≈0.5.

Таблиця 1. Результати моделювання.

Назва моделі	**Параметри**	**Наближена формула**	**Точність R^2**
Ердеша-Рен`ї	N=200, M=500	$-3*10^{-8}x^3 + 10^{-5}x^2 - 0,0011x + 1,0091$	0,99
Уаттса-Строгатца	N=200	$10^{-7}x^3 - 6*10^{-5}x^2 + 0,0061x + 0,8768$	0,98
Барабаші-Альберт	N=200	$-4*10^{-7}\ x^3 + 0,0001\ x^2 - 0,0187x + 1,0029$	0,97

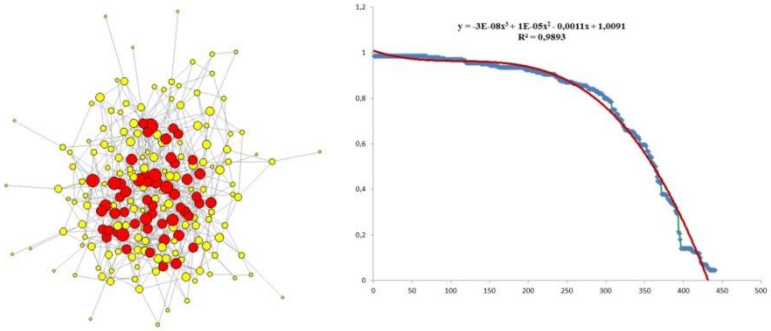


Рис. 1: Мережа Ердеша-Рен`ї та графік «потужності» мережі (вертикальна вісь) від кількості вилучених зв'язків (горизонтальна вісь).

Як видно, в кожному разі, криві з високою точністю апроксимуються кубічними многочленами, тобто для досить точного наближення досить трьох ступенів многочлена Татте-Уїтні. З огляду на топологію мереж і наведені раніше аналітичні оцінки, можна зробити висновок, що найбільша структурна живучість серед трьох розглянутих мереж властива випадкової мережі Ердеша-Рен`ї (в даному випадку ця мережа має найбільшу кількість ребер). На другому місці - мережа малого світу, ця мережа, в якій вузли мають середню ступінь 2 з розподілом, близьким до пуассонівського. І найгірші показники живучості у мережі Барабаші-Альберт.

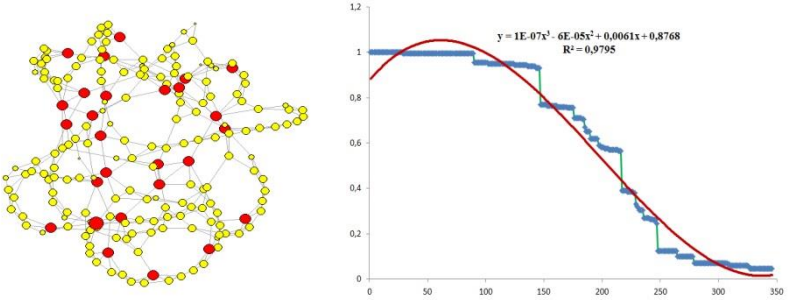


Рис. 2: Мережа Уаттса-Строгатца і графік «потужності» мережі (вертикальна вісь) від кількості вилучених зв'язків (горизонтальна вісь).

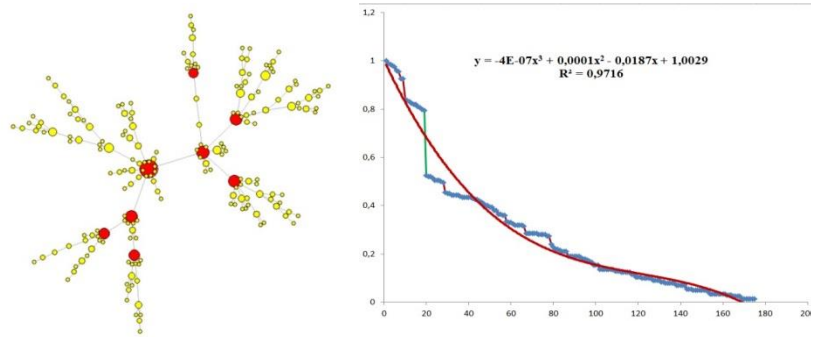


Рис. 3: Мережа Барабаші-Альберт і графік «потужності» мережі (вертикальна вісь) від кількості вилучених зв'язків (горизонтальна вісь).

Слід звернути увагу на те, що в останньому випадку, на відміну від інших, спостерігається «опукла вниз» функція живучості, це свідчить про те, що розглянута в цій роботі «потужність» різко знижується вже навіть при невеликих значеннях ймовірності деструктивних впливів.

Висновки

У цій роботі був введений новий показник структурної живучості мережевої структури, який ґрунтується на питомому розмірі максимальної компоненти зв'язності мережі при деструктивному впливі на неї.

Розвиток запропонованої моделі можливо шляхом урахування нерівнозначності вузлів мережевої моделі або зміни функції «потужності»

мережевої структури. Також дана модель може розширюватися в напрямку вивчення мереж, в яких зв'язки «регенеруються», або можуть встановлюватися нові зв'язки.

Список літератури

1. Величко В.В., Попков Г.В., Попков В.К.: Модели и методы повышения живучести современных систем связи. Москва: Горячая линия–Телеком (2014).
2. Громов Ю.Ю., Драчев В.О., Набатов К.А., Иванова О.Г.: Синтез и анализ живучести сетевых систем : монография. Москва: «Издательство Машиностроение-1» (2007).
3. Oxley, J.G.: Matroid Theory. Oxford Science Publications (1992).
4. Sekine K., Imai H., Tani S.: Computing the Tutte Polynomial of a Graph of Moderate Size. In: 6th International Symposium on Algorithms and Computation (ISAAC'95). Lecture Notes in Computer Science. 1004. pp. 224–233 (1995).
5. Tutte, W.T.: A Contribution to the Theory of Chromatic Polyno-mials. Canadian Journal of Mathematics, 6. 80-91 (1954).
6. Филлипс Д., Гарсиа-Диас А.: Методы анализа сетей. Москва, Мир (1984).
7. Долгов А.А.: Исследование живучести сетевых информационных систем с использованием неросетевых моделей. Психолого-педагогический журнал Гаудеамус, 2(16), 285-287 (2010).
8. Albert-László Barabási, Réka Albert: Emergence of scaling in random networks. Science, 286, 5439. 509-512 (1999).
9. Люк Д.А.: Анализ сетей (графов) в среде R. Руководство пользователя. Москва: ДМК Пресс (2017).
10. Erdős, P.; Rényi, A.: On Random Graphs. I. Publicationes Mathematicae, 6. 290–297 (1959).
11. Watts D.J., Strogatz S.H.: Collective dynamics of “small-world” networks. Nature. 393. pp. 440-442 (1998).

ENSURING SECURE LONG-TERM DATA STORAGE

V.V. Petrov, A.A. Kryuchyn, Ie.V. Beliak, O.V. Shihovets
Institute for Information Recording of NAN of Ukraine
E-mail: kryuchyn@gmail.com

In this presentation, technologies used to create long-term data storage systems have been reviewed. The objectives of long-term data storage systems of strategic importance for the development of society have been identified. It has been shown that network-based data storage systems (DSSs) are becoming a promising trend in this field. The most promising types of storage media that can be used to create long-term data storage systems have been identified.

The problem of long-term storage of electronic documents is of high importance, having its impact on various fields of economy, science, and culture. While producing enormous amounts of data, modern society faces challenges related to organising systems for their storage, as well as methods for securing these storages against unauthorised access. Ensuring secure storage of and access to data is an important element of an information security system. The number of electronic documents increases drastically, and therefore, long-term storage will become even more challenging over time. The key challenges of long-term storage are: preserving the authenticity of a stored document throughout the whole storage period; ageing of storage media; inevitable updates of hardware and software storage environment; as well as the interpretability and presentation of stored electronic documents. The following requirements apply to long-term data storage systems: a system for durable long-term storage of electronic documents has to be created and maintained, preserving all content-related and functional characteristics of source documents, as well as ensuring transparent search and access to the documents for users, for the purpose of both reading, analysis, and research. As the body of knowledge available to humanity continue to expand, and so do data recording technologies, the importance of developing long-term storage methods increases. Several types of data require continuous storage for decades, like archival storage, and even for hundreds or thousands of years, when dealing with ancestral data or data that can impact the survival of future generations.

1. Creating long-term data storage systems is a scientific and technological objective of high importance, having its impact on the progress in many fields of today's industry, as well as on the communication of knowledge to future generations.
2. Continuous increase in amounts of data subject to long-term storage and physical limitations of storage device capacities result in network-based data storage systems becoming the primary trend in creating long-term data storage systems.
3. When creating long-term data storage systems based on various architectures, special media for long-term data storage are used.

УДОСКОНАЛЕННЯ ЗАДАЧ ТЕСТУВАННЯ З ВИКОРИСТАННЯМ ЕКСПЕРТНИХ ТЕХНОЛОГІЙ ПРИЙНЯТТЯ РІШЕНЬ

Гнатієнко Григорій Миколайович, Снитюк Віталій Євгенович,
Тмєнова Наталія Пилипівна, Волошин Олексій Федорович
Київський національний університет імені Тараса Шевченка
G.Gna5@ukr.net, snytyuk@gmail.com,
tmyenovox@gmail.com, olvoloshyn@ukr.net

Розглянуто нові підходи до оцінювання у задачах тестування. Досліджено запитання закритого типу та запропоновано обґрунтовані формули визначення результатів тестування. Для формалізації тестування успішно використовується апарат теорії прийняття рішень та експертні технології. Наведено класифікацію запитань при проведенні тестування. Введено єдиний підхід до формалізації різних видів тестових завдань.
Ключові слова: формалізація тестування, експертне оцінювання, прийняття рішень, класифікація задач тестування.

Вступ

На сьогодні тестування є неоднозначним інструментом. Існує широкий діапазон відношення до цього інструменту. Існують спеціалісти, які упевнені, що тестування загубило освіту, зокрема американську. Інші спеціалісти не менш упевнено стверджують про широкі можливості та перспективи цього напрямку діяльності.

Істина, як завжди, знаходиться десь посередині. Слід використовувати та розвивати кращі риси цього підходу і відкидати ті, які негативно впливають на застосування інструменту.

Класифікація типів запитань у тестових завданнях

Типи закритих запитань	Тестові запитання закритої форми	Типи відкритих запитань	Відкриті запитання
Тип 1	Дихотомічні запитання: вибір альтернативної відповіді	Тип 1	На доповнення
Тип 2	Запитання з вибором однієї правильної відповіді	Тип 2	З короткою відповіддю
Тип 3	Запитання з вибором декількох правильних відповідей	Тип 3	З розгорнутою відповіддю (есе)
Тип 4	Встановлення відповідності	Тип 4	На обчислення результату
Тип 5	Встановлення правильної послідовності	Тип 5	Визначення інтервалу значень

На сьогодні нараховується близько 300 видів тестових запитань. Існують різні підходи до класифікації тестових запитань. У цій статті будемо дотримуватися такого підходу до класифікації запитань при тестуванні.

Формалізація типів запитань, які використовуються при контролі знань

Для оптимізації контролю знань застосовуються такі процедури:

– уніфікація оцінки;
– забезпечення інформаційної єдності оцінювання;
– забезпечення системності оцінювання;
– однозначність трактування запитань;
– однозначність формування оцінки на основі відповідей.

Реалізація цих процедур пов'язана з формалізацією типів запитань, процедурами їх оцінювання та критеріями визначення комплексної оцінки [1].

Нехай загальна кількість запитань, які охоплюють предметну область, відповідно до онтології, і утворюють логічну схему тестування, дорівнює *n*. Усі вони розподілені на k видів, залежно від того, який тип відповіді має відповідати запитанню.

Множину запитань опишемо таким чином:

$$Q = (Q_1, Q_2, ..., Q_n) = \left(Q_1, Q_2, ..., Q_{n_1}, Q_{n_1+1}, ..., Q_{n_{k-1}}, Q_{n_{k-1}+1}, ..., Q_{n_k}\right) \quad (1)$$

Запитання типу 1 формально представляються таким чином:

$$Q^1 = \left\langle Q_i, A_i\left(\{0,1\}\right), 1\right\rangle, \quad (2)$$

де Q^1 – підмножина запитань типу 1, Q_i^1 – запитання першого типу, $Q_i^1 \in Q^1$, $i = \{1, ..., n_1\}$, A_i^1 – множина можливих відповідей на i – е запитання, 1 – кількість відповідей, які необхідно вибрати з множини значень A_i^1, $i = \{1, ..., n_1\}$.

Запитання типу 2 формально представляються у вигляді:

$$Q^2 = \left\langle Q_i, A_i\left(a_1, a_2, ..., a_{p_1}\right), 1\right\rangle, \quad (3)$$

де $i = \{n_1 + 1, ..., n_2\}$, a_j – можливі варіанти відповідей на запитання 2-го типу, $j = \{1, ..., p_i\}$, p_i – кількість варіантів відповідей.

Запитання типу 3 визначаються таким формалізмом:

$$Q^3 = \left\langle Q_i, A_i\left(a_1, a_2, ..., a_{p_i}\right), d_i\right\rangle, \quad (4)$$

де $i = \{n_2 + 1, ..., n_3\}$, a_j – можливі варіанти відповідей на запитання 3-го типу, d_i – можлива кількість відповідей, причому $d_i = const \in \{2, p_i\}$, що визначається апріорно, або d_i – змінна величина, $d_i \in \{1, p_i\}$.

Запитання типу 4, де елементи однієї множини слід поставити у відповідність до елементів іншої множини, називають завданнями на встановлення відповідності, і формально описуються таким чином:

$$Q^4 = \left\langle Q_i, A_i\left(A_i^1, A_i^2\right), f\left(D_i^1, D_i^2\right)\right\rangle, \quad (5)$$

де $i = \{n_3 + 1, ..., n_4\}$, A_i^1, A_i^2 – множини елементів, між якими слід встановити відповідність, $f : D_i^1 \to D_i^2$ – відповідь у вигляді встановленої відповідності між множинами A_i^1, A_i^2.

Запитання типу 5, де слід встановити правильну послідовність дій або слів (варіантів відповідей тощо), використовуються у тестових завданнях на встановлення правильної послідовності, і описуються таким формалізмом:

$$Q^5 = \left\langle Q_i, A_i\left(a_1, a_2, ..., a_{p_i}\right), R_i\right\rangle, \quad (6)$$

де $i = \{n_4 + 1, ..., n_5\}$, a_j – варіанти відповідей на запитання 5-го типу множини елементів, для яких слід встановити правильну послідовність, тобто побудувати ранжування $R_i = a_{i_1} \succ a_{i_2} \succ ... \succ a_{i_{n_5}}$.

Запитання типу 1 та типу 2 є тривіальними, тому детальніше розглянемо підходи до формалізації трьох останніх типів запитань. Кожен із запитань типів 3-5 містить у собі неоднозначність та невизначеність, тому для однозначного сприйняття таких запитань їх слід формалізувати.

4 Оцінювання результатів тестування при аналізі відповідей типу 3 – при множинному виборі варіантів відповідей

Розглянемо формальне описання множинного вибору у закритих запитаннях при тестуванні. Зазначимо, що моделі та методи множинного вибору на базі аксіоми незміщеності досліджувалися також у монографії [1].

Нехай задано множину варіантів відповідей $a_i \in A, \ i \in I = \{1, ..., n\}$, кількість яких дорівнює $n, \ n = |A|$. Частина відповідей $n_1, \ n_1 < n,$ є вірними і вони складають підмножину $A^1, \ A^1 \subset A,$ а інша частина відповідей $n_0, \ n_0 < n,$ є помилковими і вони складають підмножину $A^0, \ A^0 \subset A,$ причому $A^1 \cup A^0 = A.$ Крім того, будемо вважати, що усі запропоновані тестовим завданням відповіді $a_i \in A, \ i \in I,$ є рівноцінними.

Для багатьох тестових завдань така постановка є природньою та логічною. Наприклад, вибрати серед заданих варіантів чисел ті, які є дільниками заданого числа. Можна навести багато варіантів такого роду завдань. Тобто, такий підхід має місце у повсякденному житті і задача його

формалізації при тестуванні є актуальною. Особливість таких задач полягає у тому, що вони відображують відому істину: «Скільки людей – стільки думок». Тому розв'язок має бути обґрунтованим такою мірою, яку підказує логіка його побудови, політика оцінювання, визначена організаторами тестування, здоровий глузд тощо.

Запропоновано застосовувати алгебраїчний підхід до визначення результатів оцінювання, який успішно використовується у теорії прийняття рішень та при застосуванні технологій експертного оцінювання. При алгебраїчному підході формалізація передбачає обчислення та обґрунтування усіх можливих варіантів відповідей. Максимальна кількість балів за достовірно вибрану підмножину варіантів дорівнює B. Кількість балів за правильно вибраний елемент підмножини вірних відповідей $b = B / n_1$.

Оцінювання результатів тестування при аналізі відповідей типу 4 – на встановлення відповідності

Зрозуміло, що при такому типу тестового завдання можуть виникати різні співвідношення: ін'єкція, сюр'єктія, а у ідеальному випадку – бієкція. Про тип співвідношення може бути повідомлено завчасно, або не повідомлено – тоді завдання ускладнюється.

Позначимо через $A_{il} \in A,\ i \in I, l = \{1,...,\lambda\}$, підмножину об'єктів, яку j – й студент відніс до l – го класу. Під виразом $A_{il},\ i \in I, l \in \{1,...,\lambda\}$, будемо розуміти кількість об'єктів множини $A_{il},\ i \in I, l \in \{1,...,\lambda\}$. Формули визначення мір схожості між тестовим співвідношенням та виконаним студентом, які запропоновано в роботі [2], є такими:

$$\mu_{ij}^{(1)} = 2\sum_{l=1}^{\lambda}\left(\left|A_{il} \cap A_{jl}\right| / \left(\left|A_{il}\right| + \left|A_{jl}\right|\right)\right) / \lambda,$$

$$\mu_{ij}^{(2)} = 2\sum_{l=1}^{\lambda}\left(\left|A_{il} \cap A_{jl}\right| / \max\left(\left|A_{il}\right|, \left|A_{jl}\right|\right)\right) / \lambda,$$

$$\mu_{ij}^{(3)} = 2\sum_{l=1}^{\lambda}\left(\left|A_{il} \cap A_{jl}\right| / \left|A_{il} \cup A_{jl}\right|\right) / \lambda,$$

$$\mu_{ij}^{(4)} = \sum_{l=1}^{\lambda} d^l(i, j),$$

де величини $d^l(i, j),\ i, j \in L, l \in \{1,...,\lambda\}$, визначаються таким чином:

$$d^l(i, j) = \begin{cases} 1/\lambda, \text{ якщо експерти з індексами } i, j \\ \quad \text{віднесли } l\text{-й об'єкт до одного класу,} \\ 0, \text{ якщо до різних класів.} \end{cases}$$

Оцінювання результатів тестування при аналізі відповідей типу 5 – на встановлення правильної послідовності

Будемо застосовувати, де це доцільно та обґрунтовано, алгебраїчний підхід. Тобто значення оцінки опитуваного $C(R*)$ буде пропорційно залежати від відстані наданої ним відповіді $R*$ до вірної (або – ідеальної, еталонної) відповіді R^0, і символьно позначати це:

$$C(R*) = B \cdot (1 - d(R^0, R*) / d^M),$$

де B – максимально можлива оцінка за відповідь, d^M – максимально можлива відстань, тобто відстань від вірної відповіді до повністю невірної, протилежної до ідеальної. Наприклад, коли вірна відповідь $R^0 = a_1 \succ a_2 \succ a_3 \succ a_4$, то найбільш віддалена від неї, «найгірша» відповідь: $R* = a_4 \succ a_3 \succ a_2 \succ a_1$.

Відповідно до робіт [3, 4], відстані у порядкових (рангових) шкалах, зокрема, у ранжуваннях, вимірюються з використанням різних метрик, зокрема:

– метрика Кука неспівпадання рангів (місць, позицій) елементів списку

$$d^K(R^0, R*) = \sum_{i \in I} \left| r_i^0 - r_i^* \right|, \tag{7}$$

де r_i^0 – ранг i – го елемента списку у еталонному ранжуванні елементів списку R^0, r_i^* – ранг i – го елемента списку у ранжуванні елементів R^*, заданому опитуваним;

– метрика Хемінга

$$d^H(R^0, R*) = \sum_{i \in I} \sum_{j \in I} \left| b_{ij}^0 - b_{ij}^* \right|, \tag{8}$$

де $b_{ij}^0 = 1$, $i, j \in I$, тоді і тільки тоді, коли у вірній відповіді R^0 має місце співвідношення $a_i \succ a_j$, $i, j \in I$; $b_{ij}^0 = -1$, $i, j \in I$, – коли у вірній відповіді R^0 має місце співвідношення $a_i \prec a_j$, $i, j \in I$; $b_{ij}^0 = 0$, $\forall i, j \in I$; $b_{ij}^* = 1$, $i, j \in I$, тоді і тільки тоді, коли у відповіді опитуваного R^* має місце співвідношення $a_i \succ a_j$, $i, j \in I$; $b_{ij}^* = -1$, $i, j \in I$, – коли опитуваний визначив у своїй відповіді R^* такий порядок: $a_i \prec a_j$, $i, j \in I$; $b_{ij}^* = 0$, $\forall i, j \in I$;

– метрика Евкліда

$$d^E(R^0, R^*) = \left(\sum_{i \in I} \left(r_i^0 - r_i^* \right)^2 \right)^{1/2}, \quad i \in I;$$

– вектор переваг, елементами якого є кількість альтернатив, які передують кожній альтернативі у ранжуванні.

Максимально можливі відстані між еталоном та найгіршим ранжуванням:

– для метрики Кука (7) $d^{KM} = n^2 / 2$ – для парних n; а для непарних обчислюється формулою $d^{KM} = (n^2 - 1) / 2$;

– для метрики Хемінга (8) максимальна відстань $d^{HM} = n(n-1) / 2$.

Висновки

У роботі запропоновано нові підхід до обчислення оцінки при тестуванні з використанням різних типів запитань [5]. Підходи є обґрунтованими та формалізованими, тому можуть бути застосовані у різних предметних областях. Позитивною рисою запропонованих підходів є прозорість апріорно заданих організаторами тестування правил, відсутність ситуацій невизначеності під час проведення процедури оцінювання. Крім того, описані підходи допускають можливість подальшого розвитку та удосконалення.

Джерела

1. Снитюк В.Є., Юрченко К.Н. Интеллектуальное управление оцениванием знаний. – Черкассы, 2013. – 262 с.
2. Гнатієнко Г.М. Визначення міри схожості експертних роз-поділів об’єктів за належністю до кластерів // Вісн. Київ. ун-ту. Фіз.-мат. науки. – 2001. – № 3. – С. 220–223.
3. Гнатієнко Г.М., Снитюк В.Є. Експертні технології прийняття рішень: Монографія. – К.: ТОВ «Маклаут», 2008. – 444с.
4. Волошин О.Ф., Мащенко С.О. Моделі та методи прийняття рішень : навч. посіб. для студ. вищ. навч. закл. – 2-ге вид. – К. : Видавничо-поліграфічний центр "Київський університет", 2010. – 336 с.
5. Гнатієнко Г.М., Снитюк В.Є. Математичне та програмне забезпечення задач обробки експертної інформації при проведенні іспитів //Искусственный интеллект. – 2010. – №3. – С.638-647.

METHOD FOR IDENTIFYING TWITTER ACCOUNTS THAT HAVE CHANGED THEIR OPINION ABOUT POLITICIANS

Taras Rudnyk[0000-0001-9492-0374] and Oleg Chertov[0000-0003-0087-1028]
NTU of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute",
tarasrudnyk@gmail.com, chertov@i.ua

Nowadays, social networks provide an opportunity to communicate with the world and share political views. Therefore, each politician has an account and many followers. Revealing political opinions on Twitter accounts can be a good alternative to regular social surveys. Identifying those who changed their opinions will help to understand the reasons for disappointment in a particular political figure. In this paper we introduce our approach to collect data from Twitter, detect accounts political views, opinion changes and the reasons people get frustrated. The experimental results showed that such research was successful in the analysis of Ukrainian President supporters.

Keywords: Twitter, social network, Ukraine, politics, machine learning, natural language processing, sentiment analysis.

In this research work, we have demonstrated an approach to collect Twitter dataset and classify accounts accordingly to their opinion about Volodymyr Zelenskyy and his political party "Servant of the People" during the period of his Twitter account creation to the 29th of November. Proposed approach allowed to detect four groups of people who initially were supporting Ukrainian president. Majority of accounts – 34.5 percent stopped any activity right after they replied to Volodymyr Zelenskyy first posts. It seems that they were created only with purpose of president popularization. Second group – 31 percent became politically not active and we could not classify their opinion. There is still big group of accounts – 27.4 percent which continue support of Volodymyr Zelenskyy. The last group – 7.1 percent are those who completely changed opinion about Ukrainian president which reflects results of local elections in Ukraine and social polls. Such approach could be used not only for politics but in all fields where popularity matters such as companies and brands.

Another part of this research work is the analysis of the dynamics of subscriptions on the page of Volodymyr Zelenskyy. At the moment we have identifiers of subscribers for the last 4 weeks. In future work, we will extend the observation period. In addition, we plan to download personal information and all tweets of unsubscribed accounts in order to more accurately find out the reasons.

Acknowledgment. This study is funded by the NATO SPS Project CyRADARS (Cyber Rapid Analysis for Defense Awareness of Real-time Situation), Project SPS G5286.

THE DECODING METHOD OF THE AUGMENTED REALITY MOSAIC SUSTAINABLE MARKER

Igor Ruban[1[0000-0002-4738-3286]], Hennadii Khudov[2[0000-0002-3311-2848]], Oleksandr Makoveichuk[1[0000-0003-4425-016X]] and Rostyslav Khudov[3[0000-0002-6209-209X]]
[1] Kharkiv National University of Radio Electronics, Kharkiv, Ukraine
ruban_i@ukr.net, omakoveychuk@gmail.com
[2] Kharkiv National Air Force University, Kharkiv, Ukraine
2345kh_hg,khizh_ia@ukr.net
[3] V. N. Karazin Kharkiv National University, Kharkiv, Ukraine
rhudov@gmail.com

Abstract. The subject of study in the article are augmented reality markers. The purpose is to develop the decoding method of the augmented reality mosaic sustainable marker. Objectives: analysis of basic operations in augmented reality marker systems, analysis of the main existing types of augmented reality markers, development of the decoding method of the augmented reality mosaic sustainable marker. The methods were used: methods of digital image processing, probability theory, mathematical statistics, cryptography and information security, mathematical apparatus of matrix theory. The following results were obtained. It is determined that one of the main operations in augmented reality marker systems is the decoding of markers in the video stream in order to distinguish virtual objects from the real world. The decoding method of the augmented reality mosaic sustainable marker has been developed. Conclusions. For the first time the decoding method of the augmented reality mosaic sustainable marker was obtained, which on the basis of the proposed system of indicators determines the size of the marker bit matrix, builds a marker bit matrix from the transformed image of the bit container, determines the offset in the full bit matrix. the full matrix of bits implements the filtering of the permuted image. Areas of further research are the development of a method for designing virtual objects on the plane of the augmented reality marker; development of information technology for the use of mosaic stochastic markers in augmented reality systems.

Keywords: Marker, Augmented Reality, Decoding Method.

Introduction

It is known [1–6] that augmented reality is based on visual markers and involves the use of a camera and special visual markers, such as QR-code (quick response code - quick response code). In [6] a new type of augmented reality markers is proposed. One of the main operations in augmented reality marker systems is the decoding of markers in a video stream in order to distinguish virtual objects from the real world. It is important to determine the position and orientation of the

camera, which is determined by computer vision [6–7]. The purpose of the paper is to develop a method of decoding a mosaic stochastic augmented reality marker.

Problem Analysis

The main existing types of AR markers are given in [1, 2, 6–9]: template markers – black and white markers that have a simple image inside a black frame; 2D barcode markers – markers consisting of black and white cells that encode data in bits, and sometimes frames or synchronization areas. QR codes are most often used as bar code AR markers; circular markers – similar to bar code markers, only the bits are encoded not by rectangular cells, but by black and white circular sectors; image markers – ordinary color images are used as markers. Image markers may contain a frame or other landmarks for detecting and finding a position. Image markers are usually identified by searching by pattern or by image features. In [3] a new type of mosaic sustainable augmented reality marker is proposed (see Fig. 1).

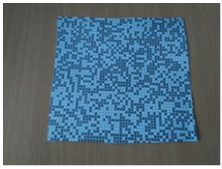


Fig. 1. A new type of mosaic sustainable augmented reality marker

So, let's develop the decoding method of the augmented reality mosaic sustainable marker has been developed.

Main Text

To decode the marker, it is necessary to solve the problem of determining the number of rows and columns in the work area, which is convenient to do on the transformed image of the mask of the AR-code (see Fig. 1). To do this, define the functions (1)–(2):

$$S_1(x) = \frac{1}{h}\sum\nolimits_y b_{x,y},$$

$$S_2(y) = \frac{1}{w}\sum\nolimits_x b_{x,y},$$

where w, h are the length and width of the image, respectively.

Thus, to determine the number of columns W, it is necessary to count the number of cross-sections from the bottom up as a function $S_1(x)$ of the threshold line $Q_1(x)$; similarly, to determine the number of rows H, count the number of crossings from the bottom up as a function $S_2(y)$ of the threshold line $Q_2(y)$.

$$W = \sum\nolimits_x \frac{d}{dx}\operatorname{sign}\left(S_1(x) - Q_1(x)\right),$$

$$H = \sum_y \frac{d}{dy} \text{sign}\left(S_2(y) - Q_2(y)\right), \tag{4}$$

where sign(x) is the function that returns the sign of the number x.

The filled matrix of the bit container is shown in Fig. 3, the value of bits 1 is encoded in white, bits 0 - in gray, and indeterminate values (no cells) - in black. In Fig. 3 shows the complete matrix of the bit container

Fig. 2. A filled bit container matrix

Fig. 3. A complete bit container matrix

The next step is to fill in the undefined pixels. This operation is performed for each block of pixel size, the value of the block is assigned to the most common bit (0 elements are not considered). The final result of decoding the marker is presented in Fig. 4, all bits, despite a significant part of the gaps (see Fig. 3), are restored correctly.

Fig. 4. A complete bit container matrix

Conclusion

A method for decoding a mosaic sustainable augmented reality marker is obtained. The method determines the size of the marker bit matrix, builds a marker bit matrix from the transformed image of the bit container, determines the offset in the full bit matrix, based on the application of inverse permutation to the full bit matrix implements the filtering of the permuted image.

References

1. Facebook Research. AR/VR-Facebook Research, https://research.fb.com/category/augmented-reality-virtual-reality, last accessed 2020/10/12.
2. Siltanen, S.: Theory and applications of marker-based augmented reality, Espoo (2012).
3. Bielievtsov, S., Ruban, I., Smelyakov K., Sumtsov, D.: Network technology for transmission of visual information. Selected Papers of the XVIII International Scientific and Practical Conference on Information Technologies and Security (ITS 2018). – CEUR Workshop Processing. – Kyiv, Ukraine, November 27. pp. 160-175 (2018).
4. Arsenov, A., Ruban, I., Smelyakov, K., Chupryna, A.: Evolution of Convolutional Neural Network Architecture in Image Classification Problems. Selected Papers of the XVIII International Scientific and Practical Conference on Information Technologies and Security (ITS 2018). – CEUR Workshop Processing. – Kyiv, Ukraine, November 27. pp. 35-45 (2018).
5. Smelyakov K., Chupryna, A., Hvozdiev, M., Sandrkin, D.: Gradational Correction Models Efficiency Analysis of Low-Light Digital Image. 2019 Open Conference of Electrical, Electronic and Information Sciences (eStream), 25-25 April 2019, Vilnius, Lithuania. pp. 34-39 (2019).
6. Ruban, I., Khudov, H., Makoveychuk, O., Khizhnyak, I., Khudov, V., Lishchenko, V.: The model and the method for forming a mosaic sustainable marker of augmented reality. In 2020 IEEE 15th Inter. Conf. on Advanced Trends in Radioelectronics, Telecommunications and Engineering (TCSET), pp. 402–406. (2020). https://doi.org/10.1109/TCSET49122.2020.235463.
7. Khudov, H., Makoveychuk, O., Khizhnyak, I., Yuzova, I., Irkha, A., Khudov, V.: The Mosaic Sustainable Marker Model for Augmented Reality Systems, International Journal of Advanced Trends in Computer Science and Engineering 9(1), 637–642 (2020). DOI: https://doi.org/10.30534/ijatcse/2020/89912020.
8. Lowe, David G.: Object recognition from local scale-invariant features. Proceedings of the International Conference on Computer Vision 2. pp. 1150–1157. (1999).
9. Hartley, R., Zisserman, S.: Multiple View Geometry in Computer Vision. Cambridge University Press New York, NY, USA. (2003).

OPTIMIZATION OF CRITICAL INFORMATION RESOURCES PROTECTION SYSTEM

Bogdan Korniyenko [1][0000-0002-2521-0878], Liliia Galata [2][0000-0002-7978-3954], Lesya Ladieva [1][0000-0002-1706-0072]
[1] National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine,
[2] National Aviation University, Kyiv, Ukraine

bogdanko@gmx.net, galataliliya@gmail.com, lrynus@yahoo.com

The main approaches to the development of a method for optimizing the protection system of critical information resources are considered. The transition from a multicriteria optimization problem to a single-criteria one is proposed. According to formulated concept of system security, the optimization task is to ensure the maximum level of security in view of cost restrictions of the protection system and the impact on system performance. A method for assessing the information security based on the task has been developed. The method allows to assess the protection level of the information and communication system. The result of the method is a quantitative assessment of the protection level. At the first stage the list of threats and probabilities of threats and threats reflection by information protection system are defined. The probabilities of repelling threats by the information security system are calculated using a mathematical model. The elements of the transition intensity matrix were found using a simulation model of critical resource protection on GNS3. At the second stage, cost restrictions of the information protection system and restrictions on reducing the level of information system productivity are introduced. At the third stage, an assessment of the system protection overall level is provided according to selected protection tools.

Keywords: Optimization; Protection System; Critical Information Resources.

The method of optimization of the information protection system for the corporate network is offered. The transition from a multi-criteria optimization problem to a single-criteria one has been made. According to the formulated definition of security system, the optimization task is to ensure the maximum level of protection (as a function of the protected information and probability of penetration) with limits on the cost of the protection system and influence to the productivity of the system.

The task of assessing the information security have been implemented. The usage of a systematic approach is a solution to the problem of comprehensive assessment of the effectiveness of the protection system of critical information resources. It will allow at the design stage to quantify the level of security and create a risk management mechanism. The level of protection system of critical information resources is estimated. The method is based on the idea that the level of risk in a protected system should be minimal in relation to the level of protection system without security tools.

A method for assessing the information security based on the task has been developed. The method allows to assess the information and communication system security level. The result of the method is a quantitative protection level assessment. At the first stage the list of threats, probabilities of threats and threats reflection by information protection system are defined. At the second stage, cost restrictions of the information protection system and restrictions on reducing the level of information system productivity are introduced. At the third stage, an assessment of the system protection overall level is provided according to selected protection tools. At the fourth stage, from the set of protection options, the one that best meets the requirements and does not go beyond these limits is selected.

ЕФЕКТ СИМЕТРІЇ ПРИ ВІДНОВЛЕННІ ЗНАЧЕНЬ МЕРЕЖЕВИХ ХАРАКТЕРИСТИК ВУЗЛІВ ПІСЛЯ ЗОВНІШНІХ ЗБУРЕНЬ

А.О. Снарський [1,2][0000-0002-4468-4542], Д.В. Ланде [1,2][0000-0003-3945-1178], О.О. Дмитренко [1][0000-0001-8501-5313]

[1] Інститут проблем реєстрації інформації НАН України, Київ, Україна
[2] Національний технічний університет України "Київський політехнічний інститут імені Ігоря Сікорського", Київ, Україна
asnarskii@gmail.com, dwlande@gmail.com, dmytrenko.o@gmail.com

В цій роботі досліджуються числові характеристики вузлів мережевих структур – показник часу релаксації мережі та індивідуальний показник часу релаксації вузла, які характеризують стійкість складної мережі та, відповідно, кожного вузла окремо до зовнішніх збуджень. Обчислення показників часу релаксації здійснюється за допомогою уповільнених ітераційних алгоритмів (HITS або PageRank, наприклад). Окрім вже відомих аналогій цих показників з часом релаксації Максвелла, що відіграє важливу роль у фізиці твердого тіла, у цій роботі також показано ефект симетрії при відновленні значень мережевих характеристик вузлів після зовнішніх збуджень. Апробацію показника часу релаксації та індивідуального показника часу релаксації було проведено на прикладі простої ненаправленої ієрархічної мережі.

Ключові слова: складна мережа, показник часу релаксації, індивідуальний показник релаксації, HITS, PageRank.

Вступ

Складні мережі відіграють важливу роль у природі. Такі мережі є поширеними та володіють нетривіальними топологічними властивостями [1], [2]. І саме ці властивості складних мереж істотно визначають їх функціональні можливості та є предметом їх дослідження.

Наукові роботи вчених П. Ердоша, М. Е. Дж. Ньюмана, А.-Л. Барабаші, С. Г. Строгаца, Дж. Клейнберга, та інших дослідників зробили значний внесок у розвиток теорії, що займається дослідженням характеристик складних мереж – теорії складних мереж (від англ. – Complex Networks).

Існують різні мережеві характеристики [3], зокрема такі найважливіші як ступінь вузла та показники, що відповідають алгоритмам HITS [4] та PageRank [5]. Проте незважаючи на вже існуючі характеристики складних мереж, важливе значення також має розробка та дослідження нових, зокрема таких, які б характеризували стійкість складної мережі до зовнішніх збурень.

Показник часу релаксації

Під час дослідження складних мереж було помічено, що значення мережевих характеристик вузлів після того, як на них подіяли зовнішні збурення, в результаті перерахунку відновлюються до своїх початкових рівноважних значень за деякий індивідуальний для кожного вузла час. Порівнюючи показник часу релаксації із самою мережевою характеристикою, значення якої піддавалося збуренню, простежується нечітка залежність. Наприклад, у роботі [6] показано, що найбільші числові значення показника часу релаксації (характеризує повільне відновлення до початкових рівноважних значень) досягаються для середніх значень степеня вузла. Або ж навпаки, швидке відновлення числових значень визначеної мережевої характеристики вузлів після збурення цих числових значень характерне для вузлів, що мають, наприклад, низький або високий степінь. Це означає, що показник часу релаксації є унікальною та неподібною до інших числовою характеристикою вузлів мережі.

Для вищеописаної нової характеристики простежується аналогія з часом релаксації Максвелла, який відіграє важливу роль у фізиці твердого тіла. Час релаксації Максвелла τ для електричного заряду, який "розсмоктується" у середовищі з питомою електричною провідністю σ та діелектричною проникністю ε є характерним часом переходу цього середовища в рівноважний стан, де зменшення щільності заряду ρ з часом t має вигляд $\rho(t) \sim e^{\frac{-t}{\tau}}$, де $\tau = \varepsilon / \sigma$. У випадку дослідження складних мереж таким часом буде називатися кількість ітераційних кроків τ відповідного алгоритму, які потрібні, щоб з деякою наперед заданою точністю μ (зазвичай μ=10^{-4}) досягти початкових рівноважних числових значень певної характеристики після її збурення. Іншими словами, показником часу релаксації мережі після збурення певного m-го вузла $\tau^{(m)}$ називають кількість ітераційних кроків, які необхідні, щоб відновилось значення кожного вузла, або $\max_k\left(\tau_k^{(m)}\right)$ (k = 1,….,n, де n – кількість вузлів мережі). А кількість ітерацій $\tau^{(m)}$, які необхідні для відновлення саме того вузла, числове значення характеристики якого було збурене, називатиметься індивідуальним показником часу релаксації вузла. Тобто перший показник характеризує вузол в термінах стійкості всієї мережі до наданого йому зовнішнього збурення, останній – в термінах його індивідуальної стійкості до цього збурення.

Встановлено [6], що на відновлення після зовнішнього збурення числових значень мережевих характеристик вузлів, наприклад, таких як HITS або PageRank, впливає топологія мережі, а також попередньо задане наближення μ – так звана умова досягнення початкового рівноважного стану. Також на відновлення числових значень впливає вибір коефіцієнта

уповільнення алгоритмів β, який застосовувався у випадках швидкої збіжності [6].

$$p(i) \rightarrow p1(i) \rightarrow p2(i) \rightarrow .... \rightarrow pequilibrium(i), \quad (1)$$

$$\hat{A}p_1(i) = p_2, \quad \hat{A}p_n(i) = p_{n+1}(i), \quad (2)$$

де $\hat{A}$ – оператор алгоритму HITS або PageRank, тобто

$$p_{n+1}(i) \leftarrow \beta p_{n+1}(i), \quad (3)$$

Ефект симетрії

Під час досліджень було помічено, що на час (кількість ітерацій відповідного ітеративного алгоритму), який потрібний на відновлення початкового рівноважного стану мережі і кожного вузла окремо, окрім умови збіжності алгоритму μ та коефіцієнта уповільнення ітеративного алгоритму β впливає також величина збурення, яка надається одному із вузлів мережі.

Так наприклад, для ненаправленої ієрархічної мережі, що представлена на Рис. 1, були проведені дослідження показника часу релаксації мережі та індивідуального показника часу релаксації вузла при різних числових значеннях збурення. У якості числових значень мережевих характеристик вузлів, яким надавалося певне числове збурення, були взяті значення отримані за допомогою алгоритму PageRank.

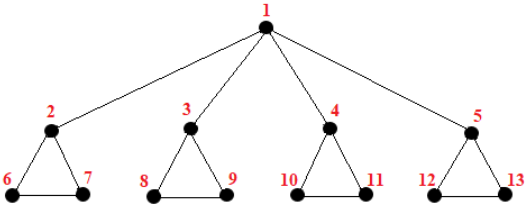


Рис. 5. Приклад простої ненаправленої ієрархічної мережі

Якщо узагальнити отримані результати для різних значень збурення у вигляді графіка (Рис. 2), то можна помітити симетрію відносно додатних та від'ємних значень збурення.

Висновки

В цій роботі було проведено аналогію між часом релаксації Максвелла та новими числовими характеристиками вузлів мережевих структур – показником часу релаксації мережі та індивідуальним показником часу релаксації вузла. Симетрія при відновленні значень мережевих характеристик вузлів відносно додатних та від'ємних значень зовнішніх збурень показує, що на показник часу релаксації мережі впливає лише абсолютна величина збурення, а не її знак.

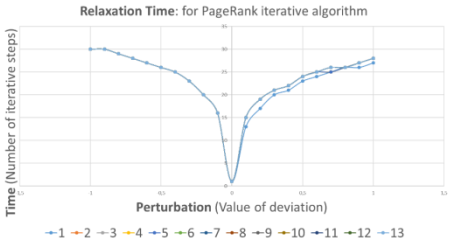


Рис. 6. Ефект симетрії для показника часу релаксації мережі (алгоритм PageRank)

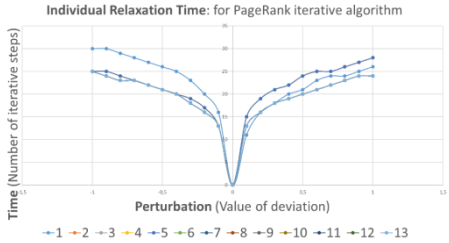


Рис. 7. Ефект симетрії для індивідуального показника часу релаксації вузла (алгоритм PageRank)

Отже, наявність ще однієї аналогії з часом релаксації Максвела підтверджує адекватність запропонованих характеристик вузлів та дає можливість використовувати їх для дослідження стійкості мережі та окремих вузлів до зовнішніх збурень.

Література

1. Newman, M. E. J.: The structure and function of complex networks. SIAM Review, vol. 45. pp. 167–256. (2003). DOI:10.1137/S003614450342480
2. Strogatz, S. H. (2001). Exploring complex networks. nature, 410(6825), 268-276.
3. Snarskii, A., Lande, D.: Modeling of complex networks: tutorial. K.: Engineering, (2015).
4. Kleinberg, J. M.: Authoritative sources in a hyperlinked environment. In Processing of ACM-SIAM Symposium on Discrete Algorithms, 46(5), pp. 604–632 (1998).
5. Brin, S., Page, L.: The anatomy of a large-scale hypertextual web search engine. Computer networks and ISDN systems, 30(1-7), 107-117 (1998). DOI:10.1016/S0169-7552(98)00110-X
6. Lande D., Snarskii A., Dmytrenko O., Subach I.: Relaxation time in complex network. ARES '20: Proceedings of the 15th International Conference on Availability, Reliability and SecurityAugust 2020. Article No.: 99 1-6. (2020). DOI: https://doi.org/10.1145/3407023.3409231

AN OVERVIEW OF DECISION SUPPORT SOFTWARE: STRATEGIC PLANNING PERSPECTIVE

Sergii Kadenko[1][0000-0001-7191-5636], Vitaliy Tsyganok[1][0000-0002-0821-4877],
Oleh Andriichuk[1] [0000-0003-2569-2026], Aleksandr Karabchuk[1],
Minglei Fu[2][0000-0002-4371-8667]
[1] Institute for Information Recording of the National Academy of Sciences of Ukraine, Kyiv, Ukraine
[2] College of Information Engineering, Zhejiang University of Technology, Hangzhou, China

Abstract. The paper features a review of available software tools and systems that utilize expert data for decision support in weakly-structured subject domains. Existing tools are analyzed and compared among themselves from the standpoint of implemented mathematical approaches and functions, particularly, those related to strategic planning. We outline the key trends of decision support tools development, witnessed during the last decades. We show that existing automated decision support tools have a bunch of limitations and drawbacks. Based on conducted analysis, we formulate relevant requirements to modern decision support software and specific recommendations as to selection of decision support tools for strategic planning and further improvement of existing decision support products.
Keywords: Decision-making Support, Expert Estimation Scale, Target-oriented Hierarchic Decomposition, Strategic Planning, Resource Allocation, Scenario Analysis.

Introduction: Relevance of Decision Support Tools Usage for Strategic Planning in Weakly Structured Subject Domains

Decision support tools using expert knowledge are mostly applied in weakly structured subject domains. These domains are characterized by a set of features, induced by high level of uncertainty [1]. Due to these features, it is often the case that only expert knowledge and methods of its processing allow us to get a detailed formal description of a weakly structured subject domain and provide the decision-maker (DM) with recommendations, which are needed to make the decision better substantiated and more credible.

In this paper we focus mainly on decision-making support problems, emerging during strategic planning in the weakly-structured subject domains. Building of strategies in such domains, usually, calls for engagement and processing of expert knowledge. In order to analyze the capacities of modern decision support (DS) tools, we suggest to use a widely-approbated original technology of strategic planning [2] as basis. The technology involves the following conceptual phases:

1) The DM formulates the main goal of the strategic plan in the given subject domain and recruits the expert group (for example – using co-authorship networks [3]).

2) The experts hierarchically decompose the main goal into factors, influencing its achievement.

3) The experts estimate relative impact of each factor from the hierarchy.

4) Finding the optimal (rational) distribution of limited available resources among projects, which maximizes the strategic goal achievement degree.

The technology demonstrated its high efficiency in multiple applications.

In our paper we try to review the features of the most well-known tools in the context of their ability to solve the problems, that emerge during strategic planning in weakly structured subject domains (based on expert and other information).

An Analytic Overview of the Most Widely-used DS Tools

Due to expansion of DS tools usage, we are witnessing an increase in demand for intelligent systems intended to solve DS problems. World-known decision support systems (DSS), developed in recent years, include the following ones: ExpertChoice, SuperDecisions, DecisionLens, D-Sight, Promethee, ОЦЕНКА И ВЫБОР (Estimate and choice), Solon, Consensus, and others. Authors of [4, 5] provide a comparative analysis of some DSS, including those, not listed in our review. Based on our own research and additional information on DSS [4, 5], we are able to obtain aggregate comparison data on DS software, particularly, from the standpoint of mathematical and technological solutions implemented in them (see Table 1). If we take just the 4 conceptual phases of the above-mentioned strategic planning cycle as DSS comparison criteria, the comparison makes no sense: all the systems we consider do allow users to rate alternatives, factors, and projects (incorporate 3 phases out of 4), while almost none of them support resource allocation functionality. That is why, in order to compile an illustrative comparison table, we have selected more constructive (specific) criteria.

Our review is not an exhaustive one, however it illustrates the general trends in the development of modern DSS, using expert data (including pair-wise comparisons and other types of estimates in various scales).

None of the listed systems allows its users to organize the whole strategic planning cycle (outlined in the introduction) using expert information in remote mode. The review allows a DM, an expert session organizer, or a knowledge engineer, to select DS tools, which are most suitable for specific problems.

Conclusions

The paper presents a brief overview and comparative analysis of available DS tools from the standpoint of their applicability to strategic planning problem-solving in weakly structured subject domains. Not a single automated DSS, featured in the review, can solve all the range of problems, emerging in the process of strategic planning based on expert and other information.

Table 2. Comparison of functional capabilities of modern DSS.

DSS name	Implemented methods	Pair-wise compa-risons	Time lag analysis	Sensiti-vity analysis	Group estima-tion	Support for diffe-rent scales	Cloud (web) version
1000Minds	PAPRIKA	**Y**	**N**	**Y**	**Y**	**N**	**Y**
AIRM Online	AIRM	**N**	**N**	**Y**	**N**	**N**	**Y**
Analytica		**N**	**Y**	**Y**	**N**	**N**	**Y**
Decision Lab		**Y**	**N**	**Y**	**N**	**N**	**N**
Decision Lens	AHP, ANP	**Y**		**Y**	**Y**		**Y**
D-Sight	MAUT, PROMETHEE	**Y**	**N**	**Y**	**Y**	**N**	**Y**
Expert Choice	AHP	**Y**	**N**	**Y**	**Y**	**N**	**Y**
Logical Decisions	AHP, MAUT	**Y**	**N**	**Y**	**Y**	**N**	**N**
Make It Rational	AHP	**Y**	**N**	**Y**	**Y**	**N**	**Y**
Mind Decider	AHP	**N**	**Y**	**Y**	**Y**	**N**	**N**
PROME-THEE Visual	PROMETHEE	**Y**	**N**	**Y**	**Y**	**Y**	**N**
RFP	AHP, ANP	**Y**	**Y**	**Y**	**Y**	**N**	**Y**
Super Decisions	AHP, ANP	**Y**	**N**	**Y**	**N**	**N**	**N**
TreeAge Pro		**N**	**N**	**Y**	**N**	**N**	**N**
Very Good Choice	ELECTRE	**Y**	**N**	**Y**	**Y**	**N**	**N**
ОЦЕНКА И ВЫБОР (Estima-tion and choice)	AHP, utility function, Pareto Analysis	**Y**	**N**		**Y**	**N**	**Y**
SVIR’	Multi-criteria optimiza-tion, Classification, AHP	**Y**	**N**	**Y**	**Y**		**N**
SOLON	MTDEA	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**

The problem of development and improvement of DS tools, allowing to obtain and process both expert and other information most thoroughly and without distortions, remains relevant. The most promising directions of further research on

the paper's topic include improvement of DSS tools' features, related to organization of remote expert sessions and incorporation of data on subject domain from all available sources.

References

1. Kadenko, S.: Prospects and Potential of Expert Decision-making Support Techniques Implementation in Information Security Area. In CEUR Workshop Proceedings. Vol-1813, pp. 8-14 (2016).
2. Tsyganok V.V., Kadenko S.V., Andriichuk O.V. Using Different Pair-wise Comparison Scales for Developing Industrial Strategies. Int. J. Management and Decision Making. Vol. 14, No 3. – pp. 224–250 (2015).
3. Balagura I., Kadenko S., Andriichuk O., Gorbov I. Defining Potential Academic Expert Groups based on Joint Authorship Networks Using Decision Support Tools. Selected Papers of the XIX International Scientific and Practical Conference on Information Technologies and Security (ITS 2019) Kyiv 2019 – pp. 222-233 (2019).
4. Weistroffer, H.R., Smith, C.H. and Narula, S.C. Multiple criteria decision support software, Ch 24 in: Figueira J., Greco S. and Ehrgott M. eds, Multiple Criteria Decision Analysis: State of the Art Surveys Series / N.Y.: Springer, (2005).
5. Buckshaw D. Decision analysis software survey / OR/MS Today. Volume 37, #5. – P. 44-53, (2010).

CODING FOR INFORMATION SYSTEMS SECURITY AND VIABILITY

Bohdan Zhurakovskyi[1][0000-0003-3990-5205], Serhii Toliupa [2][0000-0002-1919-9174], Serhiy Otrokh [1][0000-0001-9008-0902], Valeriy Kuzminykh[1][0000-0002-8258-0816], Hanna Dudarieva[3][0000-0002-9887-021X] and Vladislav Zhurakovskyi[4][0000-0002-6510-9969]
[1]National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine
[2]Taras Shevchenko National University of Kyiv, Ukraine
[3]State University of Telecommunications, Kyiv, Ukraine
[4]National Aviation University, Kyiv, Ukraine
zhurakovskiybyu@tk.kpi.ua, tolupa@i.ua, 2411197@ukr.net, vakuz0202@gmail.com, Annett.13.86@gmail.com, vbzh116@gmail.com

In this article, models of discrete channels are analyzed.It is proposed to use the L.P. Purtov for building channels of information systems. Redundant binary and non-binary codes and their error-correcting capability are analyzed. Formulas are derived for finding the probability of an uncorrected error for error-correcting codes, for codes that find errors, and for codes that detect and correct errors. The coefficient of increasing the information array is calculated for different probabilities of the detection of one element of the codeword in the channel and different error rates. The calculation of the redundancy of the error burst correction codes is performed. The use of Fire codes as anti-noise codes in the channels of the information system is proposed.

Keywords: error-correcting codes, error bursts, discrete channel model, code redundancy.

Introduction

The use of combined types of modulation in information channels allows to significantly increase the data transmission rate in comparison with binary systems. And the use of error-correcting codes with q□2 alphabet, which allowyou to identifyand correct errors, in such systems is quite a logical step. Such codes are called Non-binary (multi-positional, multi-base or q-ary) [3, 4, 5, 7, 8, 15]. Despite the fact that some redundant nonbinary codes, such as the generalized Hamming code, the Reed-Solomon code, and others, have been widely used in information channels [6, 9, 10, 11, 12, 13, 14], the development issues of the nonbinary coding theory remain quite relevant.

Statement of research problem

The intensity of the grants and the nature of the rise in the hour

Error detection and correction codes increase the noise immunity of information transmission.A careful distribution of the statistical characteristics of the sequence of errors in real communication channels showed that the errors are dependent and

have a tendency to grouping (batching), that is, there is a correlation between them.With the group nature of the distribution of errors, one parameter - the probability of error - does not fully characterize the channel; additional parameters are needed that reflect the degree of grouping of errors in various types of channels.In practice, communication channels are characterized by the dependence of the probability of interference of the next transmitted symbol on the distortion of the previous one, as well as seasonal and daily changes in the weather, the presence of industrial interference that changes the intensity at the time of the day and days of the week, mutual interference, etc. All this leads to an abrupt nature errors in information transmission channels.The calculation of error probabilities for batch distribution is carried out according to the formula:

$$P_{\text{int}} = \frac{p_s}{q} \sum_{b=1}^{b_{\max}} \left(1 + \frac{n-1}{b}\right) \cdot \frac{bP_b}{\sum_{b=1}^{b_{\max}} bP_b}, \quad (1)$$

where b is the length of the error packet; p_b - conditional probability of occurrence of a packet of errors of length b; p_s is the probability of distortion of a binary symbol; q is the density of errors in a packet, which is equal to the ratio of the number of errors in a packet to the length of this packet b.

Approximate formula for determining p_{ncor} for codes that detect and correct errors look like:

$$P_{ncor} = \frac{\sum_{i=0}^{t} C_n^i}{2^r} \left(\frac{n}{d-1}\right)^{1-s}. \quad (2)$$

These formulas give good results if the number of errors e in combination with n symbols satisfying the requirements $e < 0.3n$ [6].

When transmitting information with a simple non-redundant code, the reliability of reception depends on the type of channel and the type of interference in it. In most cases, the reliability that is found is insufficient. It must be increased so that the probability of erroneous reception of the message by the consumer is less than the probability of errors in the message without special measures.

Thus, the choice of a method for increasing the reliability of information transmission depends on many factors: the reliability of reception, the permissible transmission rate, and dependence on errors in the communication channel are necessary.

The degree of information protection from errors by the appropriate coding method depends mainly on the minimum coding distance dminof the given code.

After determining the characteristics of the communication channel [1], further selection of the code is characterized by the error probability [2].

Compared to other codes, the best code that not only detects but also corrects independent (single) errors and has recommendations of international organizations and is relatively simple to implement is Hamming code, both binary and generic. As for both packets and independent errors, the BCH, Fire and Reed-Solomon codes are the best for the same parameters.

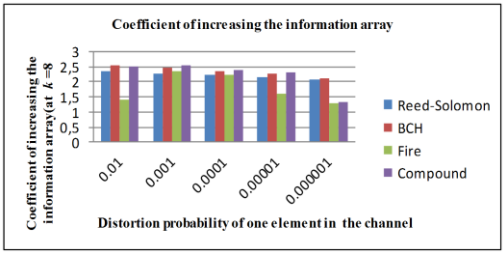


Fig1. Coefficient of increasing the information array

The following restrictions are most typical for such a control information transmission channel:

- control information arrives at the channel input in the form of messages (blocks) with a length that is a multiple of 8 bits (due to the specifics of the development of a modern element base);
- the block nature of the control information determines the use of block codes (more often Bose-Chowdhury-Hawkingham (BCH) codes or their non-binary subclass - Reed-Solomon (RS) codes);
- the physical layer provides the error probability for the binary symbol y, but errors are grouped into packets.

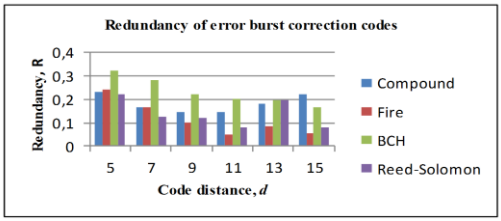


Fig2. Dependence of the code redundancy on the code distance

Conclusions

After analyzing the data obtained, we can conclude that the Fire code provides a lower probability of an uncorrected error than the BCH, compound, or Reed-Solomon codes at the same coding distance, that is, these codes require greater message redundancy to provide the same probability uncorrected error as the Fire code.

So, after considering the codes that allow you to correct single bursts of errors, you can propose the use of the Fire code instead of the BCH or Reed-Solomon code, since it has less redundancy compared to other noise immunity codes that correct bursts of errors.

References

1. B. Zhurakovskiy, N. Tsopa Assessment. Technique and Selection of Interconnecting Line of Information Networks. —IEEE, 3rd International Conference on Advanced Information and Communications Technologies (AICT). —2019. — pp. 71-75.
2. B. Zhurakovskyi, Y. Boiko, V. Druzhynin, I.Zeniv, O.Eromenko. Increasing the efficiency of information transmission in communication channels. Indonesian Journal of Electrical Engineering and Computer Science. – 2020. – Vol 19, №3 - С. 1306–1315. DOI: doi.org/10.11591/ijeecs.v19.i3.pp1306-1315
3. Жураковський Ю. П. Полторак В. П. Теорія інформації та кодування. – К.: «Вища школа», 2001. – 256 с.
4. Березюк Н. Т. Кодирование информации. – Харьков: «Вища школа», 1978. – 252 с.
5. Richard E. BlahutTheory and Practice of Error Control CodesHardcover – January 1, 1984
6. Y. Wu. "Implementation of parallel and serial concatenated convolutional codes." Dissertation submitted to the Faculty of theVirginia Polytechnic Institute and State University. April, 2000. - 206 p.
7. Берликэмп Э. Алгебраическая теория кодирования. Пер.с англ.- М.: Мир, 1972. – 478 с.
8. G. Caire and E Biglieri, "Parallel concatenated codes with unequal error protection," IEEE Transactions on Communications, vol. 46, No. 5, May 1998, pp. 565-567.
9. P. Robertson, K. Villebrun, and P. Hoeher, "A comparison of optimal and sub-optimal MAP decoding algorithms operating in the log domain," in Proc. IEEE Int. Conf on Commun., 1995. - pp. 1009- 1013.
10. Tilavat, V., &Shukla, Y. (2014). Simplification of procedure for decoding Reed–Solomon codes using various algorithms: an introductory survey. International Journal of Engineering Development and Research, 2(1), 279-283.
11. Sarwate, D. V., & Morrison, R. D. (1990). Decoder malfunction in BCH decoders. Information Theory, IEEE Transactions on, 36(4), 884-889.
12. Richard E. Blahut, "Algebraic Codes for Data Transmission", 2003, chapter 7.6 "Decoding in Time Domain"
13. Lin, S. J., Chung, W. H., & Han, Y. S. (2014, October). Novel polynomial basis and its application to reed-solomon erasure codes. In Foundations of Computer Science (FOCS), 2014 IEEE 55th Annual Symposium on (pp. 316-325). IEEE.
14. Otrokh S., Kuzminykh V., HryshchenkoO.. Method of forming the ring codes // CEUR WorkshopProceedings - Vol-2318, - 2018, pp. 188-198.
15. Berkman, L., Otrokh, S., Kuzminykh, V., Hryshchenko O. Method of formation shift indexes vector by minimization of polynomials // CEUR Workshop Proceedings -Vol-2577, -2019, pp. pp. 259-269.

OPENEMR BASED MODEL OF A ORGANIZATIONAL MANAGEMENT SYSTEM FOR A MEDICAL INSTITUTION

Sergiy Zagorodnyuk [0000-0003-3415-7746], Bogdan Sus[0000-0002-2566-5530]
and Oleksandr Bauzha [0000-0002-4920-0631]
Taras Shevchenko National University of Kyiv
szagorodniuk@gmail.com, bnsuse@gmail.com, asb@univ.kiev.ua

To organize the productive activity of specialists, medical institutions use the popular web program OpenEMR, which allows planning the working hours of specialists and patients. It provides authorized access only for employees of the institution. Patients of a medical institution usually cannot individually register and make an appointment with specialists. The article suggests a mechanism of combining the existing OpenEMR web application with a newly created web portal for a patient self-registration. Also, the article proposes methods to eliminate errors and shortcomings of the interface of the web program OpenEMR, which arose as a result of the fallacious translation of elements of this interface from English to other regional languages.

Keywords: Medical office web portal, Authorized patient web portal, Web program, Software interface.

Introduction

For the broad majority of modern commercial enterprises operating in a competitive market, it is forced to use an effective CRM system (Customer Relation Management) [1]. But due to the specifics of medical services provision the medical institutions should use specialized EHR-systems (Electronic Health Records) [2], EMR-systems (Electronic Medical Records) [3] instead of CRM-systems.

EHR-systems and EMR-systems have much in common and, conversely, do not have a clear line between their purpose and functionality [4]. EHR systems are designed for the long-term collection and reliable storage the whole data about patient medical observation and diagnosis of the in a strict chronological order [5], including data obtained from remote monitoring of the patient's condition [6].

On the contrary, EMR systems also contain the necessary information about patients' health, but this information is as complete and detailed as it is necessary for the patient's medical care in a particular medical institution.

The vast majority of EMR systems are commercial. Open source EMR systems usually provide authorized access only for competent healthcare professionals who need to manually enter patients' personal information from their words or questionnaires completed by them.

The article offers a model of deployment and configuration of a logistics information complex for a medical institution based on the popular web program OpenEMR [7], coded in PHP. The model assumes the two self-sufficient and independent modules that are exchanged using specialized software interfaces

(Fig. 1). The first module is a service web portal for competent employees of the medical institution, including administrators, coordinators and medical consultants of this institution. The second module is the web portal of patients.

Content and logistics of the infocommunication model

By the means of web portal, another important module of this service patient users can log in, register, make the necessary modifications in any field of the patient's questionnaire, place an application and obtain an electronic appointment for a medical consultation. This module consists of the web interface and system service.

The patient deals directly with the web interface. System service is a permanent resident program that can periodically execute the program code, completely hidden and invisible for the user. In Figure 1, the system service is represented as a gear.

Access to the web portal of patients can only be authorized, and access to the registration procedure itself can be organized, for example, only for the patients from a particular institution.

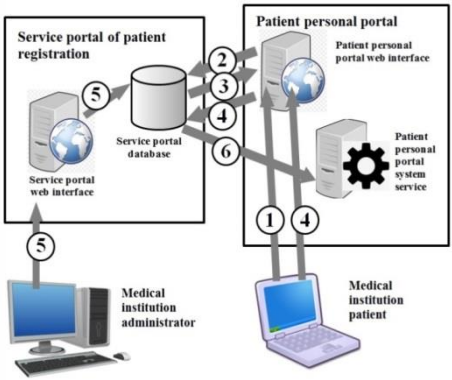


Fig. 8. Interaction stages of Patient personal portal and Service portal of patient registration in the medical institution information complex.

To get to the consultation or interview with the consulting doctor, a patient must initiate the procedure, which in Fig. 1 is presented as a sequence of stages from 1 to 6, and wait for its successful completion.

OpenEMR web interface correction

One of the definite advantages of the OpenEMR web program is it's open-source, coded in PHP. The user could not view the PHP program code, which allows the programmer to safely store in PHP code even passwords and other confidential information.

Often changes to the PHP program are required. For example, the patient's paternal type looks as shown in Fig. 3, ie only the first four letters of the word

"Borisovich" are visible. To fix the interface, it was necessary to change the line #391 in the php-file /var/www/html/openemr/library/options.inc.php
specifically, in the assignment operator to the variable **$fldlength** it was necessary to add three characters "15+", as a result of which the current value of the variable **$fldlength**, which is the length of the text field (field length) changed from 2 characters to 17 characters. In the corrected interface of the patient's father's questionnaire "Borisovich" is now displayed completely (Fig. 3.).

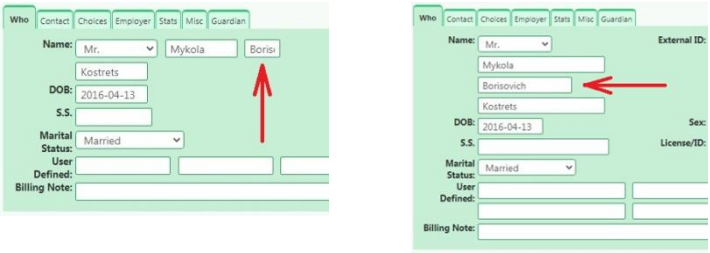


Fig.3. The original patient card interface (left), in which "Borisovich" is not completely visible and the corrected patient card interface (right), in which "Borisovich" is completely visible.

Another feature of the OpenEMR web program is the inaccurate translation of menu items from English into other languages, including Ukrainian. For example, the state of the consulting physician, in which he is at work in his office and is ready to receive the patient in English can be conveyed in only two letters - "In" (English "in the room"). The American translators decided that in Ukrainian such a state of the consulting doctor also could be conveyed not by two, but even by one Ukrainian letter "У" which is a lexical error for Ukrainian.

The special MySQL table “ lang_definitions” contains a translation of each word or phrase into all the languages including Ukrainian. For example, to illustrate the translation of the phrase "IN" in all languages, it is essential to execute the following SQL query:

```
SELECT * FROM `lang_definitions` WHERE `cons_id`=4649
```

As a result of updating the Ukrainian translation of the English phrase "IN" in the table “lang_definitions” the consultant's state can be easily changed from the incorrect value of Ukrainian "У" to the correct value "PATIENT ADMISSION".

Conclusions

1. The information complex of the medical institution consists of two interacting parts. The first is the official web portal, which provides authorized access exclusively for employees of the medical institution. The database of this web portal stores complete information about

administrators, doctors, patients, a common calendar of events with a schedule of admissions, a list of applications and patient records.

2. Another part of the information complex is a web portal for patient users, which can be authorized students and employees of any educational institution, institution or enterprise. Users who have successfully passed the authorization can individually add themselves as patients to the official web portal, update information, place a request for a consultation, review the status of accomplishment of submitted requests.
3. Opensource OpenEMR implies a wide range of options for enhancing the web user interface. In particular, the words and phrases of this interface can be replaced as a result of Update queries to MySQL tables lang_definitions and lang_languages, and the size of text fields can be changed by analyzing the dynamically generated HTML markup and comparing it with the corresponding PHP program that generates the markup.

References

1. Baashar, Y. et al. Customer relationship management systems (CRMS) in the healthcare environment: A systematic literature review. Computer Standards & Interfaces 71, 103442 (2020).
2. Bottle, A. et al. How an electronic health record became a real-world research resource: comparison between London's Whole Systems Integrated Care database and the Clinical Practice Research Datalink. BMC Med Inform Decis Mak 20, 71 (2020).
3. Janett, R. S. & Yeracaris, P. P. Electronic Medical Records in the American Health System: challenges and lessons learned. Ciênc. saúde coletiva 25, 1293–1304 (2020).
4. Electronic Medical Record Systems, https://digital.ahrq.gov/key-topics/electronic-medical-record-systems
5. Carayon, P., Smith, P., Hundt, A. S., Kuruchittham, V. & Li, Q. Implementation of an electronic health records system in a small clinic: the viewpoint of clinic staff. Behaviour & Information Technology 28, 5–20 (2009).
6. Bauzha, O., Sus', B., Zagorodnyuk, S. & Stuchynska, N. Electrocardiogram Measurement Complex Based on Microcontrollers and Wireless Networks. in 2019 IEEE International Scientific-Practical Conference Problems of Infocommunications, Science and Technology (PIC S&T) 345–349 (IEEE, 2019). doi:10.1109/PICST47496.2019.9061528.
7. Kiah, M. L. M., Haiqi, A., Zaidan, B. B. & Zaidan, A. A. Open source EMR software: Profiling, insights and hands-on analysis. Computer Methods and Programs in Biomedicine 117, 360–382 (2014).

ПОБУДОВА ТРАЄКТОРІЇ РУХУ В ДИФЕРЕНЦІАЛЬНІЙ ГРІ ПЕРЕСЛІДУВАННЯ

Барановська Л.В.[1][0000-0003-0024-8180], Довжаниця К.Г.[2][0000-0001-8529-1553], Барановська Г.Г.[3][0000-0002-6878-6973]
[1, 2, 3] КПІ ім. Ігоря Сікорського, Київ, Україна
lesia@baranovsky.org

У даній роботі розглянуто диференціальну гру переслідування. За допомогою методу розв'язуючих функцій А.О. Чикрія знайдено достатні умови для закінчення гри і одержано рівняння для чисельного знаходження розв'язуючої функції. Сформульовано достатні умови для закінчення гри. Для контрольного прикладу Л.С. Понтрягіна реалізовано візуалізацію траєкторії рухів переслідувача і втікача на площині. Для цього був створений програмний продукт на мові програмування Python. Даний програмний продукт є прототипом системи моделювання «переслідувачі-втікач», яку можна буде застосовувати для вибору керування в задачах переслідування. У роботі було розглянуто конкретний приклад. Дана задача було розв'язана аналітично за допомогою методу розв'язуючих функцій, побудовано керування переслідувача та знайдено розрахунковий час закінчення гри. Програмний продукт було протестовано на контрольному прикладі. Результат показав співпадіння розрахункового часу та фактичного часу закінчення гри.

Ключові слова: конфліктно-керовані процеси, диференціальні ігри, ігри переслідування.

Вступ

У наш час теорія конфліктно-керованих процесів відносять до математичних та комп'ютерних наук, що розвиваються досить активно. Основним об'єктом дослідження є задачі керування динамічними процесами в умовах конфлікту, який передбачає наявність двох або більше сторін, здатних діяти на процес з протилежними або неспівпадаючими цілями, а також оптимізація функціональних властивостей процесу. Динамічні процеси можуть описуватися диференційними, інтегральними, різницевих, гібридних та іншими рівняннями. Конфліктно-керовані процеси, що описуються диференційними рівняннями, називають диференційними іграми. Цей термін було введено Р. Айзеком – одним з засновників теорії диференційних ігор [1].

Найбільш простим та досить ефективним для розв'язку конкретних задач переслідування є перший метод Л.С. Понтрягіна. Перший метод Л.С. Понтрягіна тісно пов'язаний з методом розв'язуючих функцій А.О. Чикрія [2], який і буде застосовано у даній роботі. В рамках цього методу достатні умови розв'язності задачі переслідування в класі контрстратегій можна

зручно перевірити. Метод розв'язуючих функцій розроблено і для диференціально-різницевих ігор в роботах [3–9]. Ігрові проблеми зближення при відмові керуючих пристроїв розглянуто у роботах [10, 11]. Для нестаціонарних процесів метод розв'язуючих функцій представлений у роботах [12–15].

Контрольний приклад Л.С. Понтрягіна

Рух переслідувача і втікача задається системою диференціальних рівнянь:

$$\begin{cases} \ddot{x} + \dot{x} = 2u, & x \in R, & \left|\left|u\right|\right| \le 1, & \alpha, \rho > 0, \\ \ddot{y} + 2\dot{y} = v, & y \in R, & \left|\left|v\right|\right| \le 1, & \beta, \sigma > 0\,. \end{cases}$$

Нехай $\alpha = 1,\ \beta = 2,\ \sigma = 1,$ $\begin{cases} x(0) = 6,\ \dot{x}(0) = 1, \\ y(0) = 2,\ \dot{y}(0) = 1. \end{cases}$

Переслідування вважається завершеним, якщо $x = y$. Перейдемо до системи рівнянь:

$$\begin{cases} \dot{z_1} = z_2 - z_3, \\ \dot{z_2} = -z_2 + 2u, \\ \dot{z_3} = -\dot{2}z_3 + v. \end{cases}$$

Термінальна множина $M^{\cdot} = \{z : z_1 = 0\}$, при чому, $M^{o} = \{z : z_1 = 0\}$, $M = \{z : z_1 = z_z = z_3 = 0\}$. Тоді $L = \{z: z_2 = z_3 = 0\} = \{R, 0, 0\}$. Оператор ортогонального перетворення $\pi: R^s \to L$ задається матрицею

$$\pi = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Матриця A системи $\dot{z} = Az + u - v$ та області управління:

$$A = \begin{pmatrix} 0 & 1 & -1 \\ 0 & -1 & 0 \\ 0 & 0 & -2 \end{pmatrix}, U = \left\{ \begin{pmatrix} 0 \\ 2u \\ 0 \end{pmatrix} : \left|\left|u\right|\right| \le 1 \right\}, V = \left\{ \begin{pmatrix} 0 \\ 0 \\ v \end{pmatrix} : \left|\left|v\right|\right| \le 1 \right\}.$$

Фундаментальна матриця відповідної однорідної системи буде мати вигляд:

$$e^{At} = \begin{pmatrix} 1 & \frac{1-e^{-t}}{1} & -\frac{1-e^{-2t}}{2} \\ 0 & e^{-t} & 0 \\ 0 & 0 & e^{-2t} \end{pmatrix}.$$

Покладемо $\gamma(t) \equiv 0,\ z(0) = \begin{pmatrix} 4 \\ 1 \\ 1 \end{pmatrix}$. Тоді $\xi(t,z,0) = 4 + \frac{1-e^{-t}}{1}1 - \frac{1-e^{-2t}}{2}1.$

Розв'язуюча функція $\alpha(t,\tau,z,v,0)$ буде більшим додатним коренем квадратного рівняння

$$\left\| \left| \frac{1-e^{-2(t-\tau)}}{2} v - \alpha\xi(t,z(0),0) \right| \right\| = \frac{1-e^{-(t-\tau)}}{1}2$$

відносно α. Далі знайдемо час затримання втікача $T(z(0),0)$ як $\min\{t \geq 0 : \int_0^t \frac{w(t-\tau)}{||\xi(t,z,0)||} d\tau\}$, звідки одержуємо $t = 5$. Будуємо керування переслідувача:

$$u = v(\tau) - \frac{\alpha(T,\tau,z^{\circ},v(\tau),0)[\,\xi(T,z^{\circ},0)]}{\pi e^{A(T-\tau)}}, \qquad (1)$$

$$T = 5, \qquad u1 = v(\tau) - \frac{\alpha(5,\tau,z^{\circ},v(\tau),0)[\,\xi(5,z^{\circ},0)]}{\pi e^{A(5-\tau)}},$$

$$u = lex\min u1.$$

Візуалізація гри на площині

Для представлення траєкторій рухів переслідувача та втікача у грі, описаній у контрольному прикладі Л.С. Понтрягіна та для випадку площини, був створений програмний продукт на мові програмування Python. Для програми вхідними даними є наступні відомості: значення параметрів $\alpha, \beta, \sigma, \rho$; початкові координати всіх учасників.

Нехай $\alpha = 1,\ \beta = 2,\ \rho = 2,\ \sigma = 1,$ $\begin{cases} x(0) = 6,\ \dot{x}(0) = 1, \\ y(0) = 2,\ \dot{y}(0) = 1. \end{cases}$

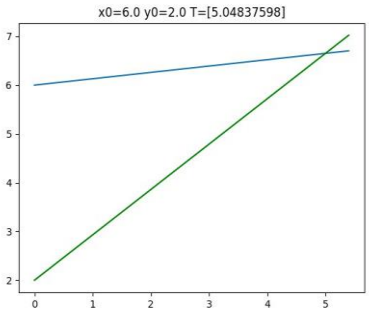


Рис. 1. Побудова графіків руху втікача та переслідувача

Ввівши дані у програму одержали графічний результат (Рис. 1) руху втікача та переслідувача, підтвердили співпадіння фактичного часу закінчення гри з розрахунковим ($T = 5$).

Алгоритм, за яким працює програмний продукт:

1. Перевіряємо умови $\rho \geq \sigma, \frac{\rho}{\alpha} \geq \frac{\sigma}{\beta}$.
2. Знаходимо фундаментальну матрицю e^{At}.
3. Перевіряємо умову $w(t) \neq 0$ та знаходимо $w(t)$.
4. Знаходимо $\xi(t, z, 0) = z_1 + \frac{1-e^{-\alpha t}}{\alpha} z_2 - \frac{1-e^{-\beta t}}{\beta} z_3$.
5. Знаходимо час затримки втікача як $\min\left\{t \geq 0 : \int_0^t \frac{w(t-\tau)}{\|\xi(t,z,0)\|} d\tau\right\}$.
6. Будуємо керування переслідувача у вигляді (1).
7. За допомогою метода Рунге-Кутта будуємо розв'язки диференціальних рівнянь в момент часу *t*.

Література

1. Айзекс, Р.: Дифференциальные игры. Москва, Мир (1967).
2. Chikrii, A.: Conflict-Controlled Processes. Springer Science &Business Media (2013).
3. Baranovska, L.V.: Quasi-Linear Differential-Deference Game of Approach. In: Sadovnichiy, V., Zgurovsky, V. (eds.) Modern Mathematics and Mechanics. Understanding Complex Systems, pp. 505–524. Springer, Cham (2019).
4. Baranovska, L.V.: On Quasilinear Differential-Difference Games of Approach. Journal of Automation and Information Sciences 49(8), 53–67 (2017).
5. Baranovska, Lesia V.: Pursuit differential-difference games with pure time-lag. Discrete and Continuous Dynamical Systems – Series B 24(3), 1024–1031 (2019).
6. Baranovskaya, G.G., Baranovskaya, L.V.: Group Pursuit in Quasilinear Differential-Difference Games. Journal of Automation and Information Sciences 29(1), 55–62 (1997).
7. Барановская, Л.В., Барановская, Г.Г.: О дифференциально-разностной игре группового преследования. Доповіді Національної академії наук України (3), 12–15 (1997).
8. Baranovskaya, L.V.: A method of resolving functions for one class of pursuit problems. Eastern-European Journal of Enterprise Technologies, vol.2, 4(74), 4–8 (2015).
9. Baranovska, L.V.: Method of resolving functions for the differential-difference pursuit game for different-inertia objects. In: Sadovnichiy, V., Zgurovsky, V. (eds.) Advances in Dynamical Systems and Control, vol.69, pp. 159–176 (2016).

10. Chikrij, A.A., Baranovskaya, L.V., Chikrij, Al.A.: The game problem of approach under the condition of failure of controlling devices. Problemy Upravleniya I Informatiki (Avtomatika) (4), 5–13 (1997).
11. Baranovskaya, L.V., Chikrii, Al.A.: Game Problems for a Class of Hereditaly Systems. Journal of Automation and Information Sciences 29(2-3), pp. 87–97 (1997).
12. Baranovskaya, L.V., Chikrij, A.A., Chikrij, Al.A.: Inverse Minkowski functionals in a nonstationary problem of group. Izvestiya Akademii Nauk. Teoriya i Sistemy Upravleniya (1), 109–114 (1997).
13. Baranovskaya, L.V., Chikrij, A.A., Chikrij, Al.A.: Inverse Minkowski functionals in a nonstationary problem of group. Journal of Computer and Systems Sciences International 36(1), 101–106 (1997).
14. Pepelyaev, V.A., Chikrii, Al.A.: On the game dynamics problems for nonstationary controlled processes. Journal of Automation and Information Sciences 49(3), 13–23 (2017).
15. Chikrii, Al.A.: On nonstationary game problem of motion control. Journal of Automation and Information Sciences 47(11), 74–83 (2015).

PRINCIPLES OF CONSTRUCTING AN OVERLAY NETWORK BASED ON CELLULAR COMMUNICATION SYSTEMS FOR SECURE CONTROL OF INTELLIGENT MOBILE OBJECTS

Vitalii Tkachov, Andriy Kovalenko, Mykhailo Hunko and Kateryna Hvozdetska
Kharkiv National University of Radio Electronics, Kharkiv, Ukraine
tkachov@ieee.org, andriy.kovalenko@nure.ua, gunko@ieee.org,
kateryna.hvozdetska@nure.ua

The current state of the problem regarding secure control of intelligent mobile objects using existing telecommunication infrastructures as a data transmission medium is described in the paper. The scientific and technical task of developing the principles of using cellular communication systems for the organization of secure remote control of Intelligent Mobile Objects has been set and successfully solved. The solution lies in the use of overlay computer networks based on VPN tunneling. Proposals have been developed for constructing an overlay computer network based on VPN taking into account traffic aggregation as well as dividing network elements according to their purposes. The related problems of connection stability at the access, distribution and core levels are considered. The possibility of using nested VPN tunneling in high-speed cellular communication systems to improve the security of transmitted data is analyzed. The result of a number of model and experimental studies, including modeling various network attacks in order to intercept and replace traffic in the "Intelligent Mobile Object – control node" chain, is the proof of the proposed principles efficiency. Recommendations have been developed for using the proposed principles of constricting an overlay network based on cellular communication systems for the secure control of Intelligent Mobile Objects in solutions related to the implementation of the "smart city" concept.

Keywords: Virtual private networks, Network security, Overlay networks, Intelligent Mobile Object.

Introduction

At the early development stages of methods for constructing communication systems for transferring data between Intelligent Mobile Objects (IMOs) and support IMO control nodes, the theory of constructing a private communication system dominated [1, 2]. For example, one of these methods is based on the presence of a relay node, which is physically located in the IMO grouping and while exchanging data between them, it ultimately transmits useful data to a stationary object or to a higher-level data transmission subsystem. In such a data exchange scheme, the direct data transmission from the IMO to the control center is impractical, since it causes a high energy consumption with limited energy reserves in the IMO. If we consider, for example, the performance of the IMO in urban conditions, then it is possible for a relay node to not always be in the

availability zone of an autonomous IMO, and that already violates the principle of constant data exchange with the control center [3]. On the contrary, constructing a stationary communication system within a city, for example, when implementing the "smart city" concept, is economically unprofitable [4,5].

The use of the existing infocommunication infrastructure of mobile operators is considered as a data transmission medium between the IMO and the control center. The modern development of technologies of cellular communication systems has long passed into the field of using seamless principles of access to the Internet. This allows to maintain data constant exchange with the IMO in the process of its movement in space.

However, since cellular communication systems are assumed to be used by all network participants, there arises a problem of ensuring the secure transmission of IMO data in such a network. In particular, we are talking about the problem of intercepting data in a mobile network in order to steal useful data (video, photo content, substitution of control commands, etc.) [5]. In other words, the attacker is interested in obtaining data of the IMO's performance results.

Known ways to solve this problem are described in [6-8]. However, the lack of described practical results on the application of the proposed solutions raises doubts about the effectiveness of these approaches, since under conditions of high IMOs mobility, the data transfer protocols used in stationary computer networks behave very differently, which requires deep experimental studies.

One of the ways to ensure security in mobile networks is to use virtual overlay networks, that is, networks that are built over the existing network infrastructures of telecommunication service providers. The basic method for implementing an overlay network is to use virtual private networks. In this regard, there is an urgent task of building an overlay computer network based on VPN tunneling technology to ensure secure data exchange between the IMO and the control center through cellular communication systems.

Principles

To ensure the functioning of an overlay computer network, it is necessary to consider the principles of building a hierarchical model of a virtual network in the form of a three-level structure [9]. Dividing a large network into small modules (levels) contributes to the grouping of functions of network elements, as well as faster localization of emerging problems associated with the movement of IMOs and reconstructing VPN tunnels. In addition, when creating a fault-tolerant segment of an overlay network, in which data that requires confidentiality circulates, it is necessary to protect data during its transfer to external networks, as well as to differentiate access when transferring it between overlay network segments within the IMO group.

Based on the approaches discussed above, it is possible to point out the elements of the overlay network and define their functions: access level (network segmentation, connection of terminal equipment to the network, switching to the distribution level); distribution level (routing functions, load balancing, packet

filtering); core layer (routing functions in tunnels, connecting objects in VPN, routing from tunnels to tunnels).

The access layer can be represented by virtual switches. The distribution layer can be represented by virtual routers performing firewalling functions. The core layer can be represented by virtual routers that perform functions of creating VPN cryptographic protection. With a large number of IMOs in the overlay network, it may not be fully connected and some nodes may perform the functions of transit ones. In this regard, for such IMOs, it may be necessary to take into account not only the presence of tunnels between nodes, but also their bandwidth. There could be situations when it is required to use several tunnels between the IMO and the control center within the overlay network and combine them (aggregation).

The aggregation condition will be:

$$v_{inVPN} > v_{core-agg}, \quad (1)$$

where v_{inVPN} is the bandwidth of uplink tunnels from one IMO core towards the external network; $v_{core-agg}$ is the maximum bandwidth of one virtual tunnel between the core and aggregation levels. With a large number of internal control commands when working in low-speed cellular communication systems ($v_{IMO \to IMO}$), it is advisable to aggregate virtual channels also between the IMO and virtual switches of the distribution level. The aggregation condition will be:

$$v_{IMO \to IMO} + v_{core \to IMO} > v_{IMO \to agg}, \quad (2)$$

where $v_{core \to IMO}$ is the data coming from the control server to other IMOs in the overlay network. $v_{IMO \to agg}$ is the maximum bandwidth of one tunnel between the IMO sublayer and the aggregation scheme. The use of the bandwidth of one tunnel, and not all, between the two levels within the segment of the IMO group in (1) and (2) is performed for the worst case. It is advisable to use other virtual channels connecting these segments not only for balancing, but also for reserving. The considered principle requires a rather complex IMO configuration at all levels, more use of IP addresses. Simpler options are possible, in which IMOs at the distribution and aggregation level are combined into a fault-tolerant virtual cluster. Thus, one virtual tunnel is functioning, while the other is in reserve at this time.

At the same time, this option requires the use of more ports at the IMO at all levels of the overlay network. In case of a shortage of ports, such a scheme can have a fully-connected cluster topology. However, this will reduce the object connectivity indicator to two.

The proposed principles were investigated experimentally by setting up a number of model and field experiments, including modeling various network attacks in order to intercept and replace traffic in the “Intelligent Mobile Object – control node” chain. The results show the reliability of the secure way of data exchange. This provides guidance on how to use these principles to securely control Intelligent Mobile Objects in solutions related to realization of the “smart city” concept.

Conclusions

Using the principles of constructing an overlay network based on cellular communication systems for the secure control of Intelligent Mobile Objects, it is possible to create full-fledged segments in the schemes of the “smart city” concept, providing data transmission of varying degrees of confidentiality with a sufficient degree of reliability.

At the same time, these principles have their advantages and disadvantages and can be applied based on the specific situation at the object and the capabilities of cellular communication systems.

It is worth paying particular attention to the logical structure of the overlay computer network, since the use of all the advantages of the supporting data transmission medium depends on its correct definition.

References

1. Dodonov, A.G., Gorbachyk, O.S., Kuznietsova, M.G.: Management organization of mobile technical objects group. Selected Papers of the XVII International Scientific and Practical Conference on Information Technologies and Security (ITS 2017), pp. 1-7. CEUR Workshop Proceedings (2017).
2. Merlac, V., Smatkov, S., Kuchuk, N., Nechausov, A.: Resourses Distribution Method of University e-learning on the Hypercovergent platform. In 9th Int. Conference on Dependable Systems, Service and Technologies (DESSERT’2018), pp. 136-140. IEEE (2018).
3. Dodonov, A.G., Gorbachyk, O.S., Kuznietsova, M.G.: Survivability of organizational management systems and the maintenance of critical infrastructure security. Selected Papers of the XIX International Scientific and Practical Conference "Information Technologies and Security" (ITS 2019), pp. 1-10. CEUR Workshop Proceedings (2019).
4. Kuchuk, H., Kovalenko, A., Ibrahim, B.F., Ruban, I.: Adaptive compression method for video information. International Journal of Advanced Trends in Computer Science and Engineering, 8(1.2), 66-69 (2019).
5. Tiutiunyk, V., Kalugin, V., Pysklakova, O., Levterov, A., Zakharchenko, Ju.: Development of Civil Defense Systems and Ecological Safety. In 2019 IEEE International Scientific-Practical Conference Problems of Infocommunications, Science and Technology (PIC S&T), pp. 295-299. IEEE (2019).
6. Aljehani, M., Inoue, M.: Communication and autonomous control of multi-UAV system in disaster response tasks. In KES International Symposium on Agent and Multi-Agent Systems: Technologies and Applications, pp. 123-132. Springer, Cham (2017).
7. Liu, D., Xu, Y., Wang, J., Xu, Y., Anpalagan, A., Wu, Q., Shen, L.: Self-organizing relay selection in UAV communication networks: A matching game perspective. IEEE Wireless Communications, 26(6), 102-110 (2019).

8. Li, B., Fei, Z., Zhang, Y., Guizani, M.: Secure UAV Communication Networks over 5G. IEEE Wireless Communications, 26(5), 114-120 (2019).
9. Tkachov, V., Hunko, M., Volotka, V.: Scenarios for Implementation of Nested Virtualization Technology in Task of Improving Cloud Firewall Fault Tolerance. In 2019 IEEE International Scientific-Practical Conference Problems of Infocommunications, Science and Technology (PIC S&T), pp. 759-763. IEEE (2019).

ENHANCED SOFTWARE VULNERABILITY MODEL

Serhii Semenov[1(0000-0003-4472-9234)], Cao Weilin[2(0000-0001-8230-5235)],
Zhang Liqiang[2(0000-0001-7493-7670))], Viacheslav Davydov[1(0000-0002-2976-8422),
Inna Petrovskaya[1(0000-0002-1425-3426)]
[1]National Technical University «KhPI», Kyrpychova str., 2, Kharkiv, Ukraine
s_semenov@ukr.net, vyacheslav.v.davydov@gmail.com, innap6699@gmail.com
[2]Neijiang Normal University, Neijiang, China
caowl@njtc.edu.cn, zhangiq@njtc.edu.cn

Software security testing enables the identification of vulnerabilities at all stages of the software life cycle, thereby maintaining the necessary level of security. Consideration of possible software cryptographic security, as well as fuzziness of input data, will greatly improve the level of accuracy of final test results. The aim of the report is to improve the model for detecting software vulnerabilities. The report outlines the main stages of the process of identifying vulnerabilities, and has developed a framework for the preparation of safety testing studies and safety testing studies. It was noted that the main features of the proposed model are: the consideration of additional procedures for checking the source code for cryptographic and other software security at all stages of testing; Consideration of the fuzzy input factor in the research process, by using the mathematical apparatus and fuzzy data logic, as well as appropriate fuzzy decision-making methods to confirm software vulnerabilities. This will expand the range of controlled data (control results) and input data in the second phase of testing and improve the accuracy of software vulnerability decisions.

Keywords: software vulnerabilities, security testing, data protection.

Introduction

Motivation

Ensuring the computer systems security, in conditions of cyberattacks increased intensity, is associated with the need to carry out prompt and accurate control of the software security level.

One of the direct mechanisms for controlling the software security level are vulnerabilities identifying methods and means. It should be noted that vulnerabilities for cyber attacks, as a negative factor, are present in the software initial versions in most practical cases.

Research has shown that technologies, models, methods and procedures that implement the software vulnerability threat identification process currently have a number of weaknesses: Limited scope due only to expertise in cyber threats and the ability to implement known models for their detection, neglecting the inconsistency and vagueness of the test input data; Low speed and incomplete control of the software real state; Poor reliability of vulnerability identification results due to poor timeliness, completeness lack and controls accuracy.

Analysis of related works

An analysis of the literature [1-10] showed that different models, methods and methodologies for testing and certification are being proposed to identify vulnerabilities and enhance the security of the computer systems and software.
For example, the work [9] proposes a vulnerability model based on the logic of firmware automata. Furthermore, the problem of mathematical modeling was reduced to the synthesis of the control firmware automaton in the adaptive-control module of the system for detecting vulnerabilities in an unstable network environment.

The problems this development is: low adaptability of models to real changes in system behavior; significant complexity of implementation algorithms in the case of a possible minor change in the behavior of at least one site (agent).

Some authors attribute the elimination of these deficiencies to the use of intelligent modelling techniques. Thus, the results of the research of intelligent methods of mathematical formalisation of data protection processes are given in the works [1, 2, 4, 5]. At the same time, the lack of inputs accepted for recording, as well as their negligence, reduces the accuracy of the identification of vulnerabilities. In addition, the shortcomings related to the slow learning of neural networks at this stage of the development of intelligent methods of mathematical formalisation have not been eliminated.

The problem of increasing the adaptability of the resulting mathematical model has been attempted by the authors [3, 7-10]. In these works, models of identifying software vulnerabilities using network methods of mathematical formalisation are presented. However, linking the proposed models to specific types of cyberattacks, As well as the lack of theoretically sound proposals for the use of methods of mathematical waiting of individual transitions from state to state of the system reduces the practical value of the development and casts doubt on the accuracy of the simulation results.

Goals

Thus, it remains a challenge to develop a vulnerability model and method that meets the requirements of speed, completeness and reliability of software security controls at all stages of software development.

Major part

A structured process model (Fig. 1, 2) has been developed for schematic description of software vulnerability identification procedures.

The purpose of the first phase is analysis, preparation of test documentation and preliminary analysis, and evaluation of software security.

A distinguishing feature of this phase of research is the introduction of additional procedures to test the source code for cryptographic and other protection of the code. This will allow for the selection of the data format, changes to the relevant test documentation, and the development of additional encryption analysis (encoding) tools.

An outline of the second phase of the security test is presented in Fig. 2. The ultimate goal of the second stage is to create a multitude of vulnerabilities and undeclared software capabilities, as well as tools and algorithms for confirming vulnerabilities.

The distinguishing feature of the second stage of the process of identifying vulnerabilities in the presented scheme is the use of the mathematical apparatus of fuzzy data for confirming the vulnerabilities of software.

The availability of a separate class of fuzzy input data determines the need to use the appropriate mathematical apparatus in modelling, selection of dynamic analysis methods (see fig. 2), taking into account the logic of fuzzy data, as well as the use of unclear decision-making methods to confirm potential software vulnerabilities.

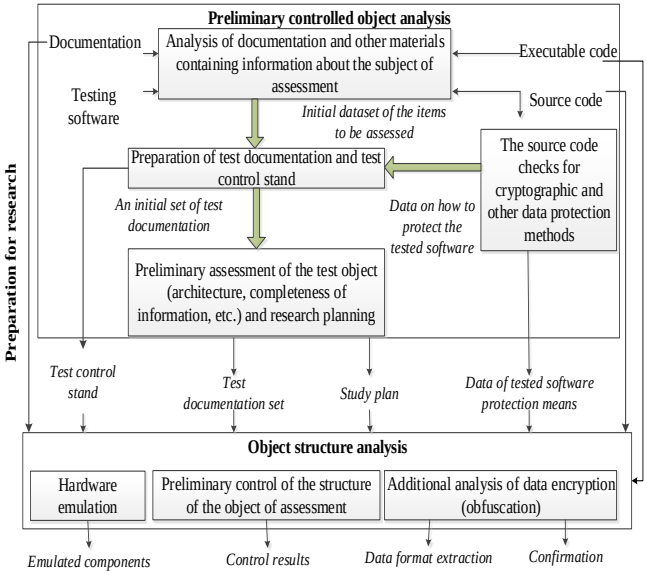


Fig. 1. Preparatory scheme for security Test Research

Conclusions.

The work presents a model for identifying software vulnerabilities. The distinctive features of the model are: 1. Perform additional procedures to verify the source code for cryptographic and other software security methods, taking this factor into account at all stages of testing. This will increase the number of controlled data (control results). 2. Integration of fuzzy input data into the research process, using mathematical tools and fuzzy data logic, as well as appropriate fuzzy decision-

making methods to validate software vulnerabilities. This will improve the accuracy of software vulnerability decisions.

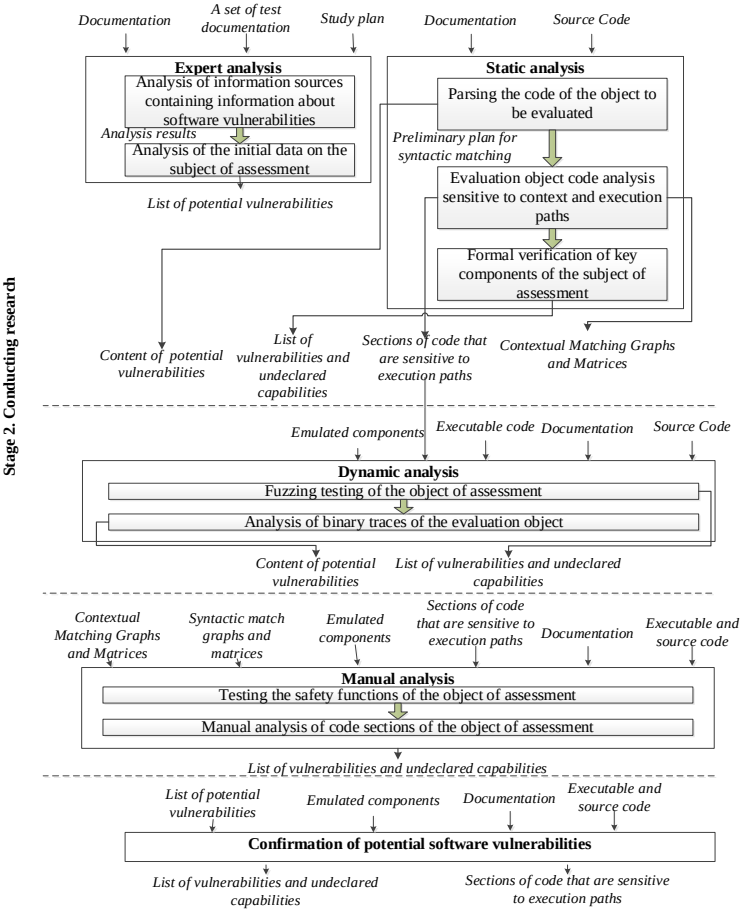


Fig. 2. Security Test Research Scheme

References

1. Adetunji Adebiyi Johnnes Arreymbi Chris Imafidon A Neural Network Based Security Tool for Analyzing Software DoCEIS 2013, IFIP AICT 394, pp. 80–87.
2. Holik F., Horalek, J., Marik, O. Neradova S. and Zitta S. Effective penetration testing with Metasploit framework and methodologies, presented at the CINTI 2014 - 15th IEEE International Symposium on Computational

Intelligence and Informatics, Proceedings, 2014, pp. 237–242.
3. Kuchuk, G., Kovalenko, A., Komari, I.E., Svyrydov, A., Kharchenko, V.: Improving Big Data Centers Energy Efficiency. Traffic Based Model and Method. In: Kharchenko, V., Kondratenko Y., Kacprzyk J. (eds) Green IT Engineering: Social, Business and Industrial Applications, Studies in Systems, Decision and Control, vol 171. Springer, Cham, DOI: doi.org/10.1007/978-3-030-00253-4_8 (2019).
4. Michael Felderer, Matthias Büchler, Martin Johns, Achim D. Brucker, Ruth Breu, Alexander Pretschner. Security Testing: A Survey Advances in Computers Volume 101, pp. 1-51 DOI doi:10.1016/bs.adcom.2015.11.003
5. Narayan Goela Jai, Mehtreb B.M. Vulnerability Assessment & Penetration Testing as a Cyber Defence Technology 3rd International Conference on Recent Trends in Computing 2015 (ICRTC-2015) pp.710-715 DOI: doi.10.1016/j.procs.2015.07.458
6. Ruban, I., Bolohova, N., Martovytskyi, V., Lukova-Chuiko, N., Lebediev, V. Method of sustainable detection of augmented reality markers by changing deconvolution. International Journal of Advanced Trends in Computer Science and Engineering, 2020, 9(2), pp. 1113-1120
7. Ruban, I., Martovytskyi, V., Lukova-Chuiko, N. Approach to Classifying the State of a Network Based on Statistical Parameters for Detecting Anomalies in the Information Structure of a Computing System. Cybernetics and Systems Analysis, 2018, 54(2), pp. 302-309
8. Semenov S., Davydov V., Lipchanska O., Lipchanskyi M. Development of unified mathematical model of programming modules obfuscation process based on graphic evaluation and review method. Eastern-European Journal of Enterprise Technologies, Vol 3, No 2 (105), 2020 pp. 6-16
9. Semyonov, S.G., Gavrilenko, S.Y., Chelak, V.V.: Information processing on the computer system state using probabilistic automata. In: Proceedings - 2017 2nd International Ural Conference on Measurements, UralCon 2017, 2017, 2017-November, pp. 11-14.
10. Yan D., Liu F., and Jia K, Modeling an information-based advanced persistent threat attack on the internal network, in ICC 2019-2019 IEEE International Conference on Communications (ICC). IEEE, 2019, pp. 1–7. DOI: 10.1109/ICC.2019.8761077

DEVELOPMENT OF SCIENTIFIC-TECHNICAL BASIS FOR CREATING AN AUTOMATICALLY CONTROLLED SYSTEM OF DETECTING AND IDENTIFYING THE SEISMIC SIGNAL BULK WAVES FROM HIGH POTENTIAL EVENTS OF TECHNOGENIC AND NATURAL ORIGIN

Yurii Hordiienko[1][0000-0002-4690-5126], Vadym Tiutiunyk[2][0000-0001-5394-6367]
Leonid Chernogor[3][0000-0001-5777-2392], Vladimir Kalugin[2][0000-0002-6899-1010]
[1] S.P. Korolev Zhytomyr Military Institute, 22, Mira Ave., Zhytomyr, Ukraine
[2] National University of Civil Defence of Ukraine, 94, Chernishevska Str., Kharkov, Ukraine
[3] V.N. Karazin Kharkiv National University, 4, Svobody Sq., Kharkov, Ukraine
tutunik_v@ukr.net

In order to develop a scientific-technical basis for creating an automatically controlled seismic monitoring system, this paper studies angular characteristics peculiarities of the seismic signal components resulting from the events with foci in the regional zone. Based on the angular characteristics analysis, the approaches are proposed for detecting and identifying automatically of the seismic signal bulk waves arising as a result of high potential events of technogenic and natural origin with foci in the regional zone. The proposed approaches implementation makes it possible to reduce the time of seismic record processing with appropriate reliability.

Keywords: earthquake, seismic waves, seismic signal, signal detection, signal identification, seismic monitoring, automatically controlled monitoring system, three-component seismic station.

Introduction

In order to create an automatically controlled system for the detection and identification of the seismic signal bulk waves resulting from high potential events of technogenic and natural origin, the angular characteristics peculiarities of the components of seismic signal from the events with foci in the regional zone are studied in the paper. Thus, one of the main stages of processing a seismic signal detected according to the seismic observation results using a three-component seismic station (TCSS) is the identification of the main seismic waves types.

At present, the seismic waves types identification in the institutions of the Main Center for Special Control of the State Space Agency of Ukraine (SSAU) is performed by the operator according to the analysis results of the total seismic signal. The accuracy of determining the seismic signal components and their parameters depends on the operator qualification level. In addition, approaches to determine the types of waves in manual processing cannot be controlled automatically.

Thus, the issue of methodological approaches development for detecting and identifying the main seismic waves types in an automatic mode in order to

increase the decision-making efficiency relative to a seismic event, as well as timely warning of relevant bodies and services is acute.

Unresolved issues

Currently, the automatic identification algorithms of the seismic signal components, which are based on the frequency characteristics of the main seismic waves types, are widely used for three-component seismic station [1–11]. However, seismic detectors with a wide frequency range are required for their implementation. For the currently existing network of seismic observations at the Main Center for Special Control of the State Space Agency of Ukraine, the seismic detectors of this type account for less than 25% of the total quantity.

Due to the seismic observation network peculiarities of the Main Center for Special Control of the State Space Agency of Ukraine, today there is an acute question of methodological approaches development for identifying the main seismic waves types resulting from hazardous events of technogenic and natural origin in the nearest zone on the basis of additional information criteria. As such criteria, the angular characteristics of the seismic signal components are promising.

Main part

Currently, the angular characteristics are calculated only for the first seismic wave entry (P−waves) – the azimuth of the seismic wave entry α and the angle of the wave exit to the daylight surface γ, which are used to calculate the location of the seismic event source (SES). For sources with significant depth, such as, for example, earthquakes in the Vrancea region (Romanian part of the Carpathians) with depth of foci $H = 80 \div 170$ km, the use of this approach, when processing measurement information in automatic mode, can lead to erroneous determination of the seismic event source location. This is conditioned by the fact that the information about the azimuth and the angle of the seismic signal exit to the daylight surface, obtained from the results of processing of the first seismic signal entry at the observation point (OP), determines the parameters of the seismic signal propagation path, but does not determine the seismic event source (SES) location on it (Fig. 1).

To increase the accuracy of determining the seismic event source location according to the observation results of three-component seismic station, it is necessary to use additional information about the distance between the observation point and the event epicenter Δ. Such information can be obtained by detecting and identifying the main seismic signal components, with subsequent using a residual time curve.

From the seismic event epicenter (earthquake or explosion), the soil elastic vibrations are propagated in all directions. These vibrations are propagated over considerable distances from the epicenter and are an important source of information for identifying a seismic phenomenon and assessing its parameters. The orthogonality peculiarities of the calculated azimuths for bulk waves can be the basis for the seismic signals automatic identification algorithm.

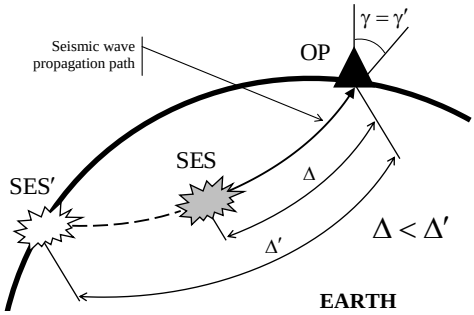


Fig. 1. The scheme of an error occurrence when determining the SES position based on the angular characteristics assessment result of the first seismic signal entry.

Given the foregoing, the algorithm for the automatic processing of measurement information at the three-component seismic station output, using the angular characteristics peculiarities, will include the following stages: 1 stage – signal detection; 2 stage – assessing the first entry angular characteristics (P-waves) – α_p and γ_p; 3 stage – search for areas of a seismic signal recording with angular characteristics, corresponding to the S-wave – $\alpha_p + 90°$ and γ_p; 4 stage – assessing the distance to a seismic source using a residual time curve; 5 stage – assessing the seismic source coordinates; 6 stage – assessing the seismic source parameters – time, magnitude and seismic event type.

As it can be seen from Fig. 2, based on the angular characteristics estimation of the first entry (P-wave), the signal phase has been determined that corresponds to the S-wave. The time difference between the P and S-waves entry, calculated by this algorithm, is 1 minute 6 seconds, which coincides with the results obtained using a residual time curve.

Thus, the methodology presented in the work for processing the seismic signal from emergencies sources of various origins allows, within the framework of the proposed algorithm, to diagnose the seismic phenomenon in automatic mode based on a comprehensive analysis of the geographic coordinates of the seismic wave source, its nature and type, amplitude-frequency characteristics and other parameters. It is performed with an increased degree of reliability and a significant time reduction in the processing information on hazard and making managerial decisions aimed at minimizing the emergency consequences.

Conclusion

A relation has been established between the characteristics of the seismic signal main components depending on the events with the foci in the regional zone, as

well as between the angular characteristics of the seismic event focus relative to the observation point.

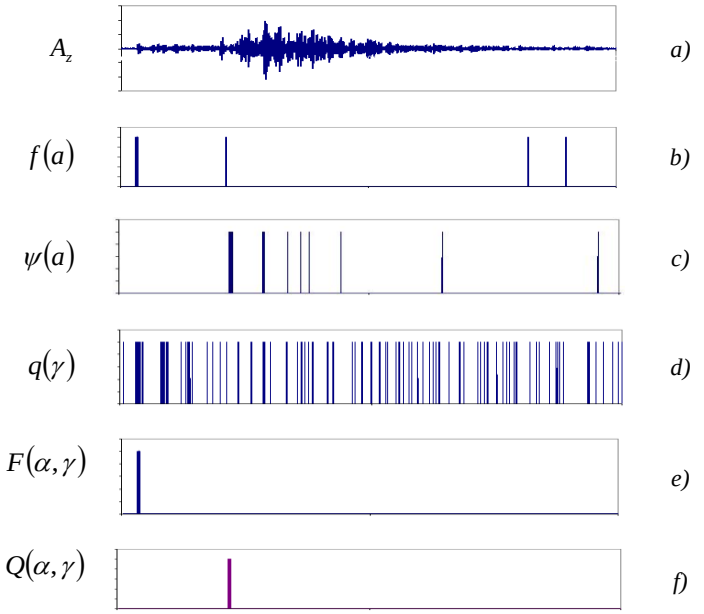


Fig. 2. The results of determining bulk waves for a seismic signal from an earthquake in the Vrancea region: a – vertical component A_z; b – $f(a)$; c – $\psi(a)$; d – q(γ); e – $F(\alpha,\gamma)$; f – $Q(\alpha,\gamma)$.

The bases are formed and an approach is proposed for detecting and identifying the seismic signal components in an automatic mode, depending on the angular characteristics peculiarities of the seismic signal components.

An algorithm has been developed for the automatically controlled processing of measurement information at the three-component seismic station output using the angular characteristics peculiarities of the seismic event focus determined in the work.

Implementation of the proposed approaches makes it possible with appropriate reliability to significantly reduce the time of processing a seismic record as compared to manual processing.

References

1. Tiutiunyk V., Kalugin V., Pysklakova O., Levterov A., Zakharchenko Ju. Development of Civil Defense Systems and Ecological

Safety. IEEE Problems of Infocommunications. Science and Technology, pp. 295–299 (2019).
2. Boschi L., Weemstra C. Stationary-phase integrals in the cross correlation of ambient noise. Reviews of Geophysics, No.53, pp.411–451 (2015).
3. Huang H.-H., Lin F.-C., Tsai V.C., Koper K.D. High-resolution probing of inner core structure with seismic interferometry. Geophysical Research Letters, No.42, pp.10622–10630 (2015).
4. Wang T., Song X., Xia H. H. Equatorial anisotropy in the inner part of Earth's inner core from autocorrelation of earthquake coda. Nat. Geosci., No.8, pp.1–4 (2015).
5. Poli P., Campillo M., de Hoop M. Analysis of intermediate period correlations of coda from deep earthquakes. Earth Planet. Sc. Lett., No.477, pp.147–155 (2017).
6. Phạm T.-S., Tkalčić H., Sambridge M., Kennett B. Earth's Correlation Wavefield: Late Coda Correlation. Geophys. Res. Lett., No.45, pp.3035–3042 (2018).
7. Santos J., Catapang A.N., Reyta E.D. Understanding the Fundamentals of Earthquake Signal Sensing Networks. Analog Devices, Vol.53, No.3, (2019), https://www.analog.com/en/analog-dialogue/articles/understanding-the-fundamentals-of-earthquake-signal-sensing-networks.html
8. Tkachov V., Hunko M. Quest Method for Organizing Cloud Processing of Airborne Laser Scanning Data. 2019 IEEE 8th International Conference on Advanced Optoelectronics and Lasers (CAOL), Sozopol, Bulgaria, pp.565–569 (2019).
9. Tiutiunyk V., Kalugin V., Pysklakova O., Yaschenko O., Agazade T. Hierarchical clustering of seismic activity local territories Globe. EUREKA: Physics and Engineering, No.4, pp.41–53 (2019).
10. Lei Li, Pierre Boué, Michel Campillo. Observation and explanation of spurious seismic signals emerging in teleseismic noise correlations. Solid Earth, No.11, pp.173–184 (2020).
11. Zheng-Yi Feng, Chia-Ming Hsu, Shi-Hao Chen. Discussion on the Characteristics of Seismic Signals Due to Riverbank Landslides from Laboratory Tests. Water, Vol.12, No.83, pp.1–19 (2020).

DEVELOPMENT AND COMPARATIVE ANALYSIS OF COMPUTER SYSTEM STATE IDENTIFICATION METHODS BASED ON ENSEMBLE ALGORITHMS

Svitlana Gavrylenko[0000-0002-6919-0055] and Illia Sheverdin[0000-0002-7881-0658]
National Technical University «Kharkiv Polytechnic Institute», Kharkiv, Ukraine
gavrilenko08@gmail.com, illia.sheverdin@gmail.com

The subject of this report are methods for identifying a computer system state.
The purpose of this report is to develop and compare methods for identifying a computer system state.
Tasks to investigate the use of classifiers with and without a teacher and ensembles of classifiers to identify the abnormal state of a computer system.
The methods were used are: artificial intelligence, machine learning, and ensemble classification algorithms.
The results were obtained: methods for identifying the abnormal state of computer systems based on classifiers with and without a teacher were learned. An ensemble method for identifying the state of a computer system without a teacher was developed, which can be distinguished by using the procedure for grouping unlabeled source data and using machine learning technology based on the "Isolation Forest" algorithm, which makes it possible not only to identify the state of a computer system, but also to highlight the name of the process that caused the abnormal state. A method for identifying a computer system functioning state was proposed by applying an ensemble of Boosting algorithm based on the C4.5 decision tree classifier and a procedure for selecting the main classification parameters: the number of trees included in the algorithm; the number of leaves in the tree; the number of splits in a node; the depth of the tree. The effectiveness of the developed classifiers was evaluated.
Conclusions. The scientific novelty of the results obtained consists in creating ensemble methods for classifying the state of a computer system without a teacher and with a teacher. The method based on the "Isolation Forest" algorithm can be used as an express method for analyzing a computer system state. This will allow not only to identify the state of a computer system state, but also to highlight the name of the abnormal processes. This method can also be used to generate labeled data and use it as the source data of the ensemble algorithm with a teacher. The algorithm with a teacher built according to the C4.5 algorithm is more accurate and can be used to refine the result of identifying a computer system state using the method based on the "Isolation Forest" algorithm.

Keywords: Computer system; operating system events; anomalous state, Decision trees; Isolation Forest, Ensemble methods; Bagging; Boosting.

Introduction

Today, a computer system functioning can be characterized by a large number of processes, which lead to difficulties in their selection and development of evaluation criteria that meet the selected indicators. One of the most common methods of big data analysis is machine learning methods, which are built to work directly with huge amounts of information.

Problem statement and scientific publications review

The basis of machine learning is data mining (Data mining) - a multi-stage automated iterative process of knowledge identification in databases, based on the analysis of huge arrays of information [1].

The most popular machine learning algorithms are given in [2-5].

Different approaches are used to improve the quality of algorithms with classifiers.

Classifiers with and without a teacher can be part of ensemble classifiers. Classifiers without a teacher are faster because they do not require training. Such methods do not require labeled data and try to find patterns directly from the source data.

Thus, the research provided an opportunity to identify a number of limitations in the use of existing methods, which, combined with the presence of different types of data (which characterize the state of the computer system), leads to significant differences in quality and poor practicality of individual classifiers. In addition, the known methods perform only the computer system state identification and do not determine the process that caused the abnormal state.

Input data formation

The name of the process, read and write operations and the path to the execution file are important indicators of the operation system (OS) and provides an opportunity to find out which process and how many times initiated the operation and on which keys or files it performed.

A procedure based on multifactor data grouping is proposed for the formation of the input dataset. To combine the two categories into a single dataset, read and write operations are combined into a single array of event groups and mapped to a process name, which further identifies the process that is causing the computer system to be abnormal.

Development of a computer system state identification method based on machine learning algorithms without a teacher

To assess the state of a computer system (CS) in case of absence of posted data, a method of identifying the state of the CS was developed based on the procedure of grouping unlabeled source data and using machine learning technology based on the algorithm "Isolation Forest" (IF).

To simulate the anomalous state of the CS, a large number of processes were launched, including archiving files with the application "7Zip" (process "7zG.exe") and malware (Virus Petya).

The method for identifying a computer system state based on IF algorithm has identified a different number of abnormal events depending on the process name. In addition, the developed software made it possible not only to identify a computer system state, but also to highlight the process name that caused this abnormal condition.

For comparison, in order to identify anomalies in the functioning of the CS, the KNN algorithm was also used. The IF algorithm detected more abnormal processes on average. The anomalies identification speed based on the KNN algorithm is lower compared to the IF algorithm.

Thus, according to the results of research on the algorithm without a teacher, the IF was chosen for the computer systems state identification.

A method development for identifying a CS state based on machine learning algorithms with a teacher

To identify an anomalous a computer system state in the presence of labeled data, the use of algorithms for constructing decision trees: REPTree, Random Tree, C4.5, HoeffdingTree, DecisionStump and ensembles of decision trees based on running and boosting to identify an anomalous a CS state was investigated [12].

At the training stage, both ensemble boost and begging algorithms are effective, with the best results obtained when using decision trees: C4.5, RETTree, and Random Tree.

Testing of the developed methods was carried out by cross-validation on a test sample that contained some of the initial data that were not used in training classifiers.

Testing confirmed the high efficiency of the ensemble algorithms with a teacher based on Boosting and Bagging for determining the state of CS. At the same time, the most effective components of the ensemble are Decision Trees based on the C4.5 algorithm. As a result of research, the method based on the ensemble Boosting algorithm and decision trees built using the C4.5 algorithm is proposed to identify a CS state in the presence of labeled data.

The developed methods comparative analysis for classifying a CS state

To assess the performance of the developed ensemble methods with and without a teacher, a comparative analysis of the proposed methods was performed. Since the ensemble Boosting meta-algorithm based on the C4.5 decision tree requires training, the time to identify the CS state using this method is on average almost 21 times longer compared to the method based on the “Isolation forest” algorithm. Same time, it is obtained that the identification time also depends on the number of events. Thus, Petya virus events set with an average 500,000 events count, the identification speed is 100 times higher than the identification speed based on the C4.5 decision tree and the ensemble Boosting meta-algorithm.

Thus, when evaluating the classification accuracy based on real data, the unknown data presence for the C4.5 algorithm was revealed.

The unknown events number is significant for some processes, such as virus Petya, which can be a sign of running uncertified processes. An increase in the number of unknown events can also be a sign of the emergence of new legitimate processes. If there is a large amount of unknown data, the quality of the training model decreases and it needs to be retrained.

Since the IF algorithm does not require training, there is not unknown data problem for it. The quality of the developed methods was evaluated using ROC-analysis. The area under the ROC-curve for the IF algorithm is 0.8143, and the classification accuracy is 89.83%. This shows that this accuracy is high enough. The area under the ROC-curve for the IF algorithm is 0.8510, and the classification accuracy is 98.32. This indicates that the accuracy is high enough, it is higher than for the IF algorithm.

CONCLUSIONS

Thus, the problem of increasing the efficiency of identifying the state of the functioning of a computer system was solved in the research made. The scientific novelty of the obtained results lies in the fact that for the first time a method for identifying the state of the CS was proposed, which differs in the use of the procedure for grouping unlabeled initial data and the use of machine learning technology based on the Isolation Forest algorithm, which makes it possible not only to identify a computer system state, but also provides the process name that caused the abnormal condition.

The method for identifying the state of the functioning of a computer system has been improved by using an ensemble of boosting methods based on the C4.5 decision tree classifier and the procedure for selecting the main classification parameters.

The method based on boosting and decision trees built according to the C4.5 algorithm made it possible to reveal the features of using the method associated with the presence of unknown data. A comparative analysis of the performance and accuracy of identification has been carried out.

It was found that the performance of the ensemble meta-algorithm for boosting and decision trees based on the C4.5 algorithm is on average almost 21 times higher than the method based on the "Isolation Forest" algorithm.

The accuracy of the developed methods was assessed using ROC analysis.

The experiments carried out confirmed the efficiency of the proposed methods on real data, which makes it possible to recommend them for practical use. At the same time, the method based on the "Isolation Forest" algorithm can be used as an express method for analyzing the CS state. This allows not only to identify the state of the CS, but also to highlight the name of the anomalous process. This method can also be used to generate labeled data and use them as input data for a supervised classifier. The boosting algorithm with the classifier C4.5 algorithm is more accurate and can be used to refine the result of identifying the state of the CS by the method based on the "Isolation Forest" algorithm.

Further research prospects may consist in creating fuzzy ensemble classifiers based on the proposed methods and optimizing methods implementation and improving the classification accuracy.

References

1. John D. Kelleher, Brian Mac Namee and Aoife D'Arcy. Fundamentals of Machine Learning for Predictive Data Analytics, Second Edition. The MIT Pres, 2020, p. 642.
2. Semenov S., Sira O., Gavrylenko S., Kuchuk N. Identification of the state of an object under conditions of fuzzy input data. Eastern-European Journal of Enterprise Technologies, 2019, 4(97), pp. 22-29.
3. Tiutiunyk V., Ruban I., Tiutiunyk O. Cluster analysis of the regions of Ukraine by the number of the arisen emergencies. IEEE Problems of Infocommunications. Science and Technology (2020).
4. Ruban, I., Martovytskyi, V., Lukova-Chuiko, N. Approach to Classifying the State of a Network Based on Statistical Parameters for Detecting Anomalies in the Information Structure of a Computing System. Cybernetics and Systems Analysis, 2018, 54(2), pp. 302-309.
5. Gavrylenko S., Chelak V. & Kazarinov M. Method of Identifying the State of Computer System under the Condition of Fuzzy Source Data.International Journal of Computer Science and Security (IJCSS), 2020, 14(5), pp.174-186.
6. Ruban, I., Martovytskyi, V., Lukova-Chuiko, N. Designing a monitoring model for cluster supercomputers. Eastern-European Journal of Enterprise Technologies, 2016, 6(2), pp. 32-37.
7. V´ıctor Blanco, Justo Puerto and Antonio M. Rodrıguez-Chia. On lp-Support Vector Machines and Multidimensional Kernels. Journal of Machine Learning Research,2020, 14(20), pp.1−29.
8. Havard Kvamme, Ornulf Borgan and Ida Scheel. Time-to-Event Prediction with Neural Networks and Cox Regression, Journal of Machine Learning Research, 2019, 20(129), pp.1−30.
9. Salvatore Ruggieri. Complete Search for Feature Selection in Decision Trees. Journal of Machine Learning Research, 2019, 20 (104), pp.1−34.
10. Vinutha H.P., Basavaraju Poornima. Analysis of Feature Selection and Ensemble Classifier Methods for Intrusion Detection. International Journal of Natural Computing Research, January 2018, pp.57-72.
11. Sebastian Buschjäger; Philipp-Jan Honysz and Katharina Morik. Generalized Isolation Forest: Some Theory and More Applications Extended Abstract. IEEE 7th International Conference on Data Science and Advanced Analytics (2020).
12. Niranjan A., Nutan D., Nitish A., Shenoy P., Venugopal R. Ensemble of Random Committee and Random Tree for Efficient Anomaly Classification Using Voting. IEEE 3rd International Conference for Convergence in Technology (2018).

ЗАСОБИ СТРУКТУРУВАННЯ ТА СЕМАНТИЧНОГО АНАЛІЗУ МЕТАДАНИХ ДЛЯ ІНТЕРПРЕТАЦІЇ BIG DATA

Рогушина Юлія[1][0000-0001-7958-2557] , Гладун Анатолій[2][0000-0002-4133-8169]
[1] Інститут програмних систем НАН України, Київ, Україна,
ladamandraka2010@gmail.com,
[2] Міжнародний науково-навчальний центр інформаційних технологій та систем НАН та МОН України, Київ, Україна, glanat@yahoo.com

Розглянуто сучасні засоби подання метаданих, проаналізовано специфіку їх застосування для опису Big Data. Обґрунтовано доцільність застосування семантичних технологій для аналізу метаданих Big Data. Використання фонових знань ПрО та знань щодо змісту окремих елементів метаданих дозволяє більш ефективно обирати набори даних з Big Data. За допомогою онтологій користувачі можуть формально описувати сферу своїх інформаційних потреб, формувати структуру інформаційних об'єктів та явно виділяти ті аспекти ПрО, які використовуються ними. Онтології метаданих дозволяють визначити семантику окремих елементів, їх рівень структурованості та відношення між ними.
Ключові слова: Big Data, онтологія, метадані, семантична розмітка

Метадані та їх властивості

Метадані дозволяють охарактеризувати контекст, контент і структуру Big Data, а також методи керування ними. Оскільки метадані накопичуються з плином часу, то вони документують історію Big Data. Ціль цієї роботи полягає в аналізі напрямків структурування метаописів Big Data з використанням існуючих стандартів.

Метадані — це структурована інформація, що робить дані корисними. Вони відображають контент, контекст та структуру даних [1]. Таке визначення описує сферу застосування метаданих, але є надто загальним для практичного використання. Метадані – це структуровані дані, які описують характеристики деякого ресурсу. Вони не тільки описують склад даних, їхню структуру та ознаки (місце та час збереження, формат), але й характеризують інформаційні технології, що підтримують ці дані, методи доступу та користувачів. Використання метаданих безпосередньо пов'язане із проблематикою обробки неструктурованої та напівструктурованої інформації і з визначенням рівня структурування даних. Метадані дозволяють описувати не тільки інформаційні ресурси (ІР), але й типові для певної предметної області (ПрО) інформаційні об'єкти (ІО), які описуються у цих ІР та є їх елементами. Ці ІО можуть описувати як об'єкти віртуального світу (документи, елементи БД, мультимедійну інформацію), так і об'єкти реального світу (персоналій, організації, географічні об'єкти тощо).

Метадані використовують для визначення семантики інформації, отже, для покращення її пошуку, вибору, розуміння і використання і залежно від

цілей анотування можуть застосовувати онтології різної складності (від контрольованих словників і глосаріїв до онтологій зі складними відношеннями інверсії, неперетину тощо) [2].

Для анотування метаданих документів та ресурсів Web використовують багато різних типів онтологій. Dublin Core (http://www.dublincore.org/) є прикладом легкої онтології, яку широко використовують, щоб вказати характеристики електронних документів, і саме цю онтологію найчастіше застосовують для семантизації метаданих та забезпечення їх інтероперабельності [3]. Вона може застосовуватися й для опису метаданих Big Data. Метадані, які характеризують Big Data, описують джерело даних; автора і дату створення документа; розмір і формат даних; кількість записів у наборі даних; опис цих даних тощо. Як правило, елементи метаданих, що характеризують семантику Big Data, – це анотації та коментарі, що представлені у неструктурованому вигляді.

В обробці Big Data аналіз метаданих має ключове значення, тому що метадані містять інформацію не тільки про походження даних, але й про їх зміст, тоді як безпосередній аналіз всієї сукупності Big Data є надто складним та потребує багато часу.

Стандартизація метаданих

Стандартизація метаданих – основа інтероперабельності та повторного використання як самих метаданих, так і тих IP, що характеризують ці метадані. Міжнародні організації зі стандартизації приділяють велику увагу розробці форматів метаданих та специфікують набір полів (атрибутів, властивостей, елементів метаданих) для формального опису різних типів IP та IO)г. В Україні на сьогодні діє три міжнародні стандарти, що стосуються метаданих, – ISO 15489-1:2016, ISO 15836-1:2017, ISO 15836-2:2019, які прийняті як національні стандарти методом підтвердження.

ISO 15489-1:2016 Information and documentation — Records management — Part 1: Concepts and principles (Інформація і документація. Керування документами. Частина 1: Поняття і принципи) визначає основні поняття і принципи керування документами і інформацією. Цей стандарт може бути застосований для відображення основних властивостей Big Data: 1) автентичності; 2) достовірності; 3) цілісності; 4) придатності їх до обробки. В стандарті описано інформаційні поля, що входять в структуру метаданих. Для Big Data ці поля дозволяють відобразити наступну інформацію: опис контенту Big Data – це структура даних (форма, формат, зв'язки між блоками Big Data); середовище створення; взаємозв'язок з іншими блоками Big Data (шардинг, реплікація) і метаданими; ідентифікатори та іншу інформацію, що потрібна для видобутку і подання даних; дії і події, що пов'язані з цими Big Data (дата, час дій, зміна метаданих тощо). Big Data, які не супроводжуються такими метаданими, не можуть використовуватися повноцінно.

ISO 15836-1:2017 Information and documentation — The Dublin Core metadata element set — Part 1: Core elements (Інформація та документація.

Набір елементів метаданих «Дублінське ядро». Частина 1: Основні елементи) описує 15 елементів Dublin Core, які використовують для опису ресурсів. В цьому стандарті під ресурсом розуміють будь-який об'єкт, який можна ідентифікувати (наприклад, у сфері комп'ютерних наук ресурсами виступають окремі документи, тексти, аудіо- та відео-файли, Web-сторінки, бази даних тощо). Big Data та їх метадані теж відповідають такому визначенню і можуть розглядатися як ресурси

ISO 15836-2:2019 Information and documentation — The Dublin Core metadata element set — Part 2: DCMI Properties and classes (Інформація та документація. Набір елементів метаданих «Дублінське ядро». Частина 2: DCMI властивості і класи) є розширенням і доповненням першої частини цього стандарту ISO 15836-1. Розширення полягає у тім, що він надає програмістам загальну універсальну мову для створення та аналізу метаданих. Така універсальна мова забезпечує розширений опис елементів метаданих, використовуючи їх оновлені властивості та класи. Стандарт ISO 15836-2 збільшує початковий набір з 15 основних властивостей до 40 властивостей і 20 класів для підвищення точності і виразності описів в стандарті Dublin Core. Основна увага цього стандарту зосереджена на опису загальних властивостях елементів метаданих, що необхідні для базової інтероперабельності між різними мовами програмування та ПрО їх застосування. Це забезпечує користувачів та програмістів засобами для створення описів ресурсів, які взаємодіють на загальному рівні, розширює сферу застосування стандарту Dublin Core.

Використання семантичних технологій для аналізу метаданих Big Data

На сьогодні створено багато методів, що забезпечують здобуття знань з різних типів ІР – структурованих, напівструктурованих та неструктурованих.

Семантичні технології дозволяють не шукати заново вже відомі користувачам закономірності та семантично збагатити зв'язки між параметрами (властивостями об'єктів, що аналізуються) за рахунок наявних знань щодо відношень між ними. Для цього можуть використовуватися методи Data Mining та його підвидів – Text Mining та Web Mining []. Для аналізу метаданих Big Data важливими вимогами до цих методів є швидкість обробки та наявність евристик, що дозволяють значно скоротити час аналізу. Наприклад, знання щодо відношення "клас-підклас" між параметрами метаданих дозволяє вдосконалити навчальну вибірку. Це обумовлює необхідність отримання таких фонових знань: 1) пошук ІР, що пертинентні задачі користувача; 2) здобуття з цих ІР необхідних фонових знань; 3) використання отриманих знань для аналізу даних.

У випадку аналізу Big Data ці задачі конкретизуються наступним чином: 1) вибір сховища Big Data, в якому здійснюється пошук; 2) пошук або створення онтології ПрО, що містить фонові знання щодо задачі користувача; 3) аналіз метаданих Big Data з метою вибору набору даних, пертинентних задачі користувача, використовуючи онтологію ПрО;

4) генерація потрібного набору даних (підмножини Big Data за визначеними умовами), використовуючи онтологію ПрО; 5) здобуття з онтології ПрО тих термінів та відношень між ними, які потрібні для більш ефективного аналізу великого обсягу інформації; 6) використання отриманих знань для аналізу отриманого набору даних та для інтерпретації результату. Онтології дозволяють як аналізувати семантично метадані, що описують Big Data, так і аналізувати самі дані, але виникає потреба у засобах пошуку та генерації онтологій, що відповідають інформаційним потребам конкретного користувача.

Пропонується використовувати для цього семантично розмічені Wiki-ресурси, які підтримують автоматизоване створення онтологій – як усього ресурсу в цілому, так і довільної підмножини, що може визначатися явним переліком Wiki сторінок або генеруватися за семантичним запитом.

Висновки

Метадані – основне джерело інформації про зміст Big Data. Аналіз метаданих дозволяє відбирати набори даних з Big Data, що неявно містять відомості, необхідні для рішення задачі користувача. Для цього необхідно знаходити фонові знання, що характеризують потреби користувача, та співставляти їх із метаданими на семантичному рівні. Джерелом таких фонових знань можуть бути семантизовані Wiki-ресурси, які дозволяють створювати легкі онтології. На жаль, до значних проблем призводить відсутність специфічних для Big Data стандартів подання метаданих, розробка яких значно підвищила б ефективність здобуття знань з Big Data.

Література

1. Baca, M. (Ed.). (2016). Introduction to metadata. Getty Publications. http://d2aohiyo3d3idm.cloudfront.net/publications/virtuallibrary/0892368969.pdf
2. J. Rogushina, A. Gladun, S.Pryima “Use of Ontologies for Metadata Records Analysis in Big Data” In: CEUR Workshoop Proceedings, 2019. http://ceur-ws.org/Vol-2318/paper5.pdf
3. Weibel, S. L., & Koch, T. (2000). The Dublin core metadata initiative. D-lib magazine, 6(12), 1082-9873. http://mirror.dlib.org/dlib/december00/weibel/12weibel.html
4. Гладун А.Я., Рогушина Ю.В. Data Mining: пошук знань в даних. – К.:ТОВ "ВД "АДЕФ-Україна", 2016. – 452 с.

ANALYSIS OF SELF-HEALING OF INFORMATION SYSTEMS IN REAL TIME

Igor Ruban[0000-0002-4738-3286] and Valentyn Lebediev[0000-0002-0095-7481]
Kharkiv National University of Radio Electronics, Kharkiv, Ukraine

In modern conditions of functioning of human life services, an important factor is the round-the-clock automatic monitoring of various systems and a quick mechanism for their recovery, which can constantly maintain the availability of various functions and services offered to the client. Modern information systems are by far the most effective, safe, reliable, easily upgradeable and self-healing. The main goal of self-healing is to create an automated system capable of self-healing without human intervention. Such a system has various predefined actions and procedures that are suitable for recovering the system from various expected failure states. Self-healing property - control of a set of environmental factors faced by the system.
Keywords: self-healing, information systems, real time systems, software reliability, structural reconfiguration.

Introduction

As modern software systems and applications become more versatile and functional, the ability to manage incompatible resources and serve disparate user requirements becomes increasingly necessary. In addition, as the complexity of information systems increases, troubleshooting system failures and recovering from malicious attacks become more complex, time consuming, expensive, and error prone. These factors stimulate research into the concept of self-healing systems. This report analyzes the state of the art in self-healing systems to assess effectiveness through a systematic literature review. Self-healing systems try to "heal" themselves in the sense of recovering from mistakes and restoring the normative level of performance by analogy with how a human biological system heals a wound. Such systems use models, external or internal, to monitor the behavior of the system and use the received inputs to adapt to the runtime environment. This report analyzes the provision of self-healing information systems in real time.

Analysis of self-healing

Today, self-healing is increasingly used in embedded real-time systems, which are used in security-critical environments to mitigate threats. These systems implement a self-healing process through reconfiguration, that is, the exchange of system components during operation, which is aimed at stopping or eliminating failures [1]. This response is severely constrained in real time, as responding too late does not produce the desired result. Therefore, it is necessary to analyze the propagation of failures over time, and also take into account how the propagation of failures changes with reconfiguration. Modern approaches do not analyze the

propagation time of failures and changes in structural reconfiguration during the propagation of failures. The challenge is to improve the approach to hazard analysis by expanding the propagation models of failures due to propagation time and taking into account the behavior of the system during real-time reconfiguration. This allows you to analyze how reconfiguration with a certain duration changes the propagation of failures of the real-time system and, thus, whether it can prevent the danger and implement a self-healing process. The report examines the real-time embedded systems, which are often used in security-critical contexts. Even if the system does not contain any design errors, dangerous situations can be caused by random errors that occur, for example, due to deterioration of physical components. Therefore, these systems must be analyzed for potential hazards with subsequent self-healing. Self-healing systems try to reduce the likelihood of hazards by stopping the propagation of failure. Due to real time, the hazard analysis of such systems must also take into account the temporal properties of the system, i.e. the propagation time of failures as well as the duration of reconfiguration. It also takes into account that failures propagate during the reconfiguration. Modern approaches to automated hazard analysis [2] do not take into account the propagation time of failures and changes in structural reconfiguration during the propagation of failures. Thus, they are not suitable for self-healing systems. The report discusses a component-based approach to hazard analysis that specifically targets real-time reconfigurable systems that use self-healing. Heterogeneity, portability, complexity, and new application areas are creating new software reliability problems that cannot be solved cost-effectively only with classical approaches to software development [3]. Self-healing systems can successfully solve these problems, thereby increasing software reliability while reducing maintenance costs. Self-healing systems should be able to automatically detect failures at runtime, detect failures, and find a way to get the system back to acceptable behavior. The report discusses the problems underlying the construction of self-healing systems, with special emphasis on functional failures, and presents a set of methods for creating software systems that can automatically eliminate such failures. Methods of automatic receipt of assertions for efficient detection of functional failures and determination of sequences of actions, alternative to the sequence of failures, in order to return the system to an operating state are considered. Assessing the performance of self-healing systems is challenging due to their interaction with an often very dynamic environment. In a specific case, evaluating the performance of self-healing systems, approaches to self-healing and tuning their parameters are based on the characteristics of the occurrence of failures and the resulting interactions with self-healing actions [4]. A classification of the different types of input for such systems is presented and the limitations of each type of input are analyzed. The main conclusion is that the inputs used are often not complex in terms of the characteristics considered for the occurrence of failures. To further explore the impact of the identified constraints, experiments are presented showing that incorrect assumptions about failure characteristics can lead to large performance prediction errors, disadvantageous design decisions regarding alternative self-healing approaches, and

disadvantageous deployment. In addition, experiments show that using multiple alternative input characteristics can help reduce the risk of premature bad decisions during system design. Systems designed for self-healing are able to recover in response to changing environmental or operating conditions. Thus, the goal is to avoid catastrophic failure by quickly taking corrective action. A self-healing mechanism is proposed that monitors, diagnoses and heals one's own internal problems, using self-awareness as contextual information [5]. A self-management system that encapsulates a self-healing mechanism associated with improved reliability, addresses to perform self-diagnosis and self-healing. To confirm the effectiveness of the self-healing mechanism, practical experiments are carried out with a prototype system.

Conclusions

Thus, in the course of studying self-healing systems, it can be concluded that there are various features characterizing self-healing mechanisms that create opportunities for building a model and developing a method for ensuring self-healing systems in real time. In the implementation of the mechanisms of autonomy and self-healing of the system, human functions are limited by the level of Supervision. In this case, the subsystems that generally implement the concept of autonomic calculations, must have at the input of the analysis system a full vector of values of indicators that fully reflect the basic processes of the systems. The dimensionality of this vector is determined by the complexity of the systems, so at present it is necessary to develop a comprehensive criterion for calculating the evaluation of indicators, in order to minimize control effects under harsh conditions to ensure sustainable operation. Thanks to the technology of autonomic computing, today it is possible to implement the functions of self-recovery of information systems, which may be required in hardware and software failures, for self-defense in a computer network, for self-management of resources, to optimize the system as a whole.

References

1. Priesterjahn C., Steenken D., Tichy M. Component-based timed hazard analysis of self-healing systems //Proceedings of the 8th workshop on Assurances for self-adaptive systems. – 2011. – С. 34-43.
2. Ansari Z. System and method for active diagnosis and self healing of software systems : пат. 7293201 США. – 2007.
3. Gorla A. et al. Achieving cost-effective software reliability through self-healing //Computing and Informatics. – 2012. – Т. 29. – №. 1. – С. 93-115.
4. Ghahremani S., Giese H. Evaluation of Self-Healing Systems: An Analysis of the State-of-the-Art and Required Improvements //Computers. – 2020. – Т. 9. – №. 1. – С. 16.
5. Park J. et al. Self-management system based on self-healing mechanism //Asia-Pacific Network Operations and Management Symposium. – Springer, Berlin, Heidelberg, 2006. – С. 372-382.

METHOD FOR EVALUATING THE EFFECTIVENESS OF METHODS FOR EMBEDDING DIGITAL WATERMARKS

Nataliia Bolohova[1[0000-0001-8927-0055]], Vitalii Martovytskyi[1[0000-0003-2349-0578]], Vladyslav Diachenko[1[0000-0003-2725-8784]], Oleksiy Kolomiitsev[2[0000-0001-8228-8404]] and Volodymyr Fedorchenko[1[0000-0001-7359-1460]]

[1] Kharkiv National University of Radioelectronics, 14, Nauki Avenue, 61000, Kharkiv, Ukraine

[2] National Technical University «Kharkiv Polytechnic Institute», 2, Kyrpychova str., 61002, Kharkiv, Ukraine

In recent years, we have seen a significant increase in traffic moving across various networks and channels. The development of technology and global network leads to an increase in the amount of multimedia traffic. To authenticate and avoid abuse, data should be protected with watermarks. This paper discusses various robust and invisible watermarking methods in the spatial domain and the transform domain. The basic concepts of digital watermarks, important characteristics and areas of application of watermarks are considered in detail. The paper also presents the most important criteria for assessing the digital watermark effectiveness. Based on the analysis of the current state of the digital watermarking methods, robustness, imperceptibility, security and payload have been determined as the main factors in most scientific works. Moreover, researchers use different methods to improve / balance these factors to create an effective watermarking system. Our research identified the main factors and new techniques used in modern research. And the assessment of watermark method effectiveness was proposed.

Keywords: Digital watermark, LSB, Copyright protection.

Introduction

In recent years, we have seen a significant increase in traffic moving across various networks and channels. The development of technology and global network leads to an increase in the amount of multimedia traffic [1]. To authenticate and avoid abuse, data should be protected with watermarks. A digital watermark prevents illegal copying and distribution of multimedia content by hiding unremarkable ownership data [2, 3]. A digital watermark (DWm) is a technology used in information security to solve the problem of copyright protection for certain digital information. With this, a special mark is applied to the digital graphic images, which can remain visible or invisible to a person.

The watermark embedding process can be determined based on domain and different groups. According to the domain, the methods of embedding digital watermarks could be divided both in the domain space and during domain transformation, for example, methods based on SVD [4]. Initially, approaches to the spatial domain, in which the process of embedding watermarks can be carried out by directly changing the pixels in an image, were used. It has such advantages

as low computational complexity and ease of implementation. The most commonly used techniques in this area are the least significant bit (LSB) and the spread spectrum and correlation. However, techniques such as discrete cosine transforms (DCT), discrete wavelet transforms (DWT), discrete Fourier transforms (DFT), singular value decomposition (SVD), and Karhunen-Loew transforms (KLT) are examples of transform domain techniques. In the context of DWM visibility, there are two different categories of digital watermark: visible and invisible. In addition, there are various types of invisible watermarks that are both robust and fragile. A complete classification of digital watermarks is presented in [4].

The system for embedding and extracting watermarks performs the tasks of embedding and extracting a digital watermark from a container image. A precoder is a device designed to convert a hidden watermark to a form suitable for embedding in a container. Digital watermark embedding device is intended for embedding a hidden digital watermark in the container image. The system combines two types of information so that they can be seen by two fundamentally different detectors. One of the detectors is a digital watermark extraction system, and another is a human. Pre-processing is often performed using a key to improve the secrecy of the data that is embedded. The following stage is the digital watermark "embedding" into the container, for example, by modifying the least significant bits of the coefficients. This process is possible due to the peculiarities of the human perception system. It is well known that images have great psycho-visual redundancy. The human eye is similar to a low-pass filter that skips fine details [5].

Analysis of modern methods of digital watermarking

After analysing the current state of research on watermark methods and significant watermark characteristics, it is possible to form the following assessment of watermark method effectiveness:

$$EF = R \bullet \alpha_r + SR \bullet \alpha_{sr} + ER \bullet \alpha_{er} + SC \bullet \alpha_{sc} + DT \bullet \alpha_{dt}, \quad (1)$$

where $R, R \in [0,1]$ – a security assessment of the method for digital watermark embedding;

$RS, RS \in [0,1]$ – an assessment of digital watermark invisibility on an image;

$ER, ER \in [0,1]$ – the probability of an error of the first and second kind;

$SC, SC \in [0,1]$ – an assessment of digital watermark fragility;

DT – is the number of embedded digital watermarks;

$\alpha_r, \alpha_{sr}, \alpha_{er}, \alpha_{sc}, \alpha_{dt}$ – significance coefficients of the corresponding criteria of the digital watermark method. Such coefficients are needed, since there is no universal method for digital watermark embedding, therefore, due to such coefficients, it is possible to adjust the significance of each criteria and thereby influence the final method efficiency for a specific task facing a digital watermark.

As an example, we describe the effectiveness criteria of the digital watermark method for images.

Robustness of the method for digital watermarking can be clarified in a statistical sense by accepting the following assumptions:

The digital watermark method W can be defined as a set of some functions F and G that describe the process of embedding and extracting a digital watermark on a multiple data, such that each element is a pattern required for the digital watermark method:

$$E = (E_i, i = 1, 2, ..., N) \quad (2)$$

For simplicity, we will assume that the input data set includes a pair of values for a container image Im and a digital watermark Wm:

$$E_i = \{\mathrm{Im}_i, Wm_i\}, \quad (3)$$

The method has two stages, embedding $F(E_i) = \mathrm{Im}_i^*$ and extracting $G(\mathrm{Im}_i^*) = Wm_i$. Since robustness is the algorithm ability to resist attacks, we add an attack function $At_j \in At$, where At is the set of admissible attacks on the digital watermark:

$$At = (At_j, j = 1, 2, ..., M). \quad (4)$$

Using the function $At_j(\mathrm{Im}_i^*) = \mathrm{Im}_i^{*\prime}$ will distort the digital watermark container. Then, for some values of E_i the obtained value from $G(\mathrm{Im}_i^{*\prime})$ may be within acceptable limits $\triangle i$:

$$\left| G(\mathrm{Im}_i^{*\prime}) - G(\mathrm{Im}_i^*) \right| \le \triangle i \quad (5)$$

For all other E_i, that form subset of $E_l \in E$, execution of $G(\mathrm{Im}_i^{*\prime})$ does not provide an acceptable result, that is:

$$\left| G(\mathrm{Im}_i^{*\prime}) - G(\mathrm{Im}_i^*) \right| > \triangle i \quad (6)$$

All such cases are called false.

Each value of E_i represents a possible combination of values that can be input to functions F and G. Number N of possible E is very large, but finite. Combination of actions,

$$F \to \forall At_j, At_j \in At \to G, \quad (7)$$

that results in correct reading of the digital watermark from the container or a false positive. Thus, the probability P that, after an attack on the container with digital watermark $At_j(\mathrm{Im}_i^*) = \mathrm{Im}_i^{*\prime}$ removing the digital watermark from the container will lead to an erroneous result (6) is equal to the probability that the set of input data E_i used in the j attack belongs to the set E_l. Suppose that $n_{l,j}$ is the number of different input data sets contained in E_l for the j attack, then $Q_j = n_{l,j} / N$ is a probability that the execution of the sequence of functions (7) on the data set E_i, randomly selected from E among the equally probable, will result in false digital watermark extraction. In this case, $P_j = 1 - Q_j = 1 - n_{l,j} / N$ is the probability that in the j attack on the input set E_i, randomly selected from E, will lead to the correct digital watermark extraction, equation (5). Since various attacks are independent events, the probability that all attacks will not lead to false extraction of the digital watermark, equation (6), is equal to the product probability of each attack:

$$R = \prod_{1}^{j} P_j \tag{8}$$

Robustness of the digital watermark method will be assessed by this product probability.

To assess imperceptibility, it is possible to use various metrics to estimate the difference between two images, such as the signal-to-noise metric. But such a metric will be of a higher order than all other terms, which can lead to a false assessment of the method effectiveness. Therefore, the following metric was proposed for assessing imperceptibility:

$$SR = \frac{h_x + h_y}{2};$$

$$h_x = \frac{1}{Wh} \cdot \sum_{i=1}^{Wh-1} \left[\frac{1}{\sigma_x^{i-1} \sigma_x^{i-1}} \cdot \sum_{j=0}^{Ht-1} (Di_{i-1,j} - m_x^{i-1}) \cdot (Di_{i,j} - m_x^{i}) \right]^2; \tag{9}$$

$$h_y = \frac{1}{Ht} \cdot \sum_{i=1}^{Ht-1} \left[\frac{1}{\sigma_x^{i-1} \sigma_x^{i-1}} \cdot \sum_{j=0}^{Wh-1} (Di_{i,j-1} - m_y^{i-1}) \cdot (Di_{i,j} - m_y^{i}) \right]^2;$$

where SR - an indicator for estimating the changes made when embedding digital watermark distortions;

h_x, h_y – average value of the square of linear correlation coefficient of the changes made when embedding digital watermark image distortions vertically and horizontally, respectively;

m_x^i, m_y^i – mathematical expectation of the distortion amount in corresponding column or row of the pixel matrix;

σ_x^i, σ_y^i – standard deviation of the distortion amount in corresponding column or row of the pixel matrix;

$Di_{x,y}$ – amount of brightness distortion in the corresponding pixel;

Wh – image width in pixels;

Ht – image height in pixels;

$Pixel'$ – pixel matrix of the watermarked image;

$Pixel$ – pixel matrix of the original image;

Y – pixel brightness determination operator;

ER – sum of errors of first and second kind when digital watermark extracting from the container.

To assess the fragility of a digital watermark, the following equations will be used. Suppose that there is some distortion function $H(k, imige)$, where k is the number of distortion pixels, and $imige$ - is an image with an embedded digital watermark. Then the fragility estimate will be as follows

$$F(E_i) = \mathrm{Im}_i^*;$$

$$G(\mathrm{Im}_i^*) = Wm_i;$$

$$H(k, \mathrm{Im}_i^*) = \mathrm{Im}_i'^*$$

$$G(\mathrm{Im}_i'^*) = Wm_i';$$

$$count = \sum_{x=0}^{W} \sum_{y=0}^{H} \begin{cases} 1, if\ Wm_{i[x,y]}' \neq Wm_{i[x,y]} \\ 0, if\ Wm_{i[x,y]}' = Wm_{i[x,y]} \end{cases}; \quad (10)$$

$$SC = count / (W \bullet H);$$

$$Di_{x,y} = \left| Y(Pixel'_{x,y}) - Y(Pixel_{x,y}) \right|;$$

$$m_x^i = \frac{1}{Ht} \bullet \sum_{i=0}^{Ht-1} Di_{i,j};$$

$$m_y^i = \frac{1}{Wh} \bullet \sum_{i=0}^{Wh-1} Di_{i,j};$$

$$\sigma_x^i = \sqrt{\sum_{i=0}^{Ht-1} (Di_{i,j} - m_x^i)^2};$$

$$\sigma_y^i = \sqrt{\sum_{i=0}^{Wh-1} (Di_{i,j} - m_y^i)^2}$$

where $count$ - number of digital watermark pixels that do not match the original digital watermark after distorting the k pixels of the digital watermark container. W, H - image width and height in pixels.

Conclusion

This paper discusses various robust and invisible watermarking methods in the spatial domain and the transform domain. The basic concepts of digital watermarks, important characteristics and areas of application of watermarks are considered in detail.

The paper also presents the most important criteria for assessing the digital watermark effectiveness. Based on the analysis of the current state of the digital watermarking methods, robustness, imperceptibility, security and payload have been determined as the main factors in most scientific works. Moreover, researchers use different methods to improve / balance these factors to create an effective watermarking system. Our research identified the main factors and new techniques used in modern research. And the assessment of watermark method effectiveness was proposed.

References

1. Merlac, V., Smatkov, S., Kuchuk, N., Nechausov, A.: Resourses Distribution Method of University e-learning on the Hypercovergent platform. In: Conf. Proc. of 2018 IEEE 9th Int. Conf. on Dependable Systems, Service and Technologies. DESSERT'2018, pp. 136-140. Kyiv, May 24-27, 2018.
2. Van Schyndel R. G., Tirkel A. Z., Osborne C. F. A digital watermark //Proceedings of 1st international conference on image processing. – IEEE, 1994. – Т. 2. – С. 86-90.
3. Islam M., Laskar R. H. Geometric distortion correction based robust watermarking scheme in LWT-SVD domain with digital watermark extraction using SVM //Multimedia Tools and Applications. – 2018. – Т. 77. – №. 11. – С. 14407-14434.
4. Loukhaoukha K. On the security of digital watermarking scheme based on SVD and tiny-GA //Journal of Information Hiding and Multimedia Signal Processing. – 2012. – Т. 3. – №. 2. – С. 135-141.
5. Ruban, I., Bolohova, N., Martovytskyi, V., Lukova-Chuiko, N., Lebediev, V. Method of sustainable detection of augmented reality markers by changing deconvolution. International Journal of Advanced Trends in Computer Science and Engineering, 2020, 9(2), p. 1113-1120.

STRUCTURAL CONSTRUCTIONS OF COMPUTER ENGINEERING ELEMENTS IN K-VALUED LOGIC

Volodymyr Yartsev [1] and Dmitry Gololobov [2]
[1] State University of Telecommunications, Kyiv, Ukraine
jvp57@ukr.net
[2] National Aviation University, Kyiv, Ukraine
gololobov.dma@meta.ua

Abstract. Methods are proposed for implementing the logical functions of digital devices without memory and with memory using three-valued logic obtained using direct and inverse multiplexer circuits as a logical basis. This simplifies the development of digital logic circuits with a large number of input arguments, reduces the number of tiers of the tree structure of the circuit implementation. Using a multiplexer to perform operations in three-valued logic allows reducing the number of logical elements in the structure of a digital device, reducing the delay time when generating a control signal, and using existing microcircuits with transistor-transistor logic. As an example, structural synthesis of a finite state machine recognizing a given combination is given.
Keywords: digital devices, logical basis, finite-state machine, logic function, multiplexer, ternary adder, ternary trigger.

The possibility of using direct and reverse multiplexer circuits as a logical basis for constructing in three-valued logic structural diagrams of digital devices without memory and with memory of any complexity is shown. As an example of using a multiplexer to implement 3-valued logic, a structural synthesis of a finite state machine is given, which, from a sequence of input signals, detects a given combination and generates an output control signal. Using the proposed method for constructing block diagrams of digital devices based on k-valued logic, it will be possible to create control signal conditioners for constructing various models of discrete-analog representation of data on semiconductor indicators. This will increase the capacity, processing speed and display of large amounts of information while reducing hardware and energy costs for their technical implementation.

MINIMIZATION OF AVERAGE DELAY IN HYPERCONVERGED NETWORK ENVIRONMENT

Nina Kuchuk[1(0000-0002-0784-1465)], Heorhii Kuchuk[1(0000-0002-2862-438X)],
Andriy Kovalenko[2(0000-0002-2817-9036)], Oleksii Liashenko[2(0000-0002-0146-3934)],
Roman Yaroshevych[2(0000-0002-7949-1513)],
[1]National Technical University «KhPI», Kharkiv, Ukraine
nina_kuchuk@ukr.net, kuchuk56@ukr.net
[2]Kharkiv National University of Radio Electronics, Kharkiv, Ukraine
andriy_kovalenko@yahoo.com, oleksii.liashenko@nure.ua,
roman.yaroshevych@nure.ua

Hyperconverged networks can reduce operating costs and suggest the possibility of rapid horizontal scaling. At the same time, due to the centralization of management and standardization of hardware and software modules, efficiency indicators are reduced, which leads to a deterioration in QoS indicators. The task of minimizing the average delay of message transmission in hyperconverged network environment is currently relevant and focused on improving QoS indicators. The purpose of the paper is to develop such a method for the synthesis of a hyperconverged network, in which the processing time of requests is reduced, but the total costs of data transmission do not exceed the specified values. In addition, mathematical model of message transmission in hyperconverged network environment is described. Based on the model, a method for message transmission time minimization has been developed. The optimization problem is formulated in the following form: determine the optimal values of the information flow density that minimizes the average delay. To solve it, the method of indefinite Lagrange multipliers is used. To determine the undefined Lagrange multiplier, the limiting value of the data transfer costs was used. The proposed analytical expressions allow, at a given cost of transferring a unit of information, selecting both the number of elements of the buffer memory of hyperconverged network modules and the optimal value of the information flow density that provides the minimum average delay in message transmission in hyperconverged networks.
Keywords: information structure, resource allocation, hyperconverged network, delay.

Introduction

Motivation

Today, hyperconverged networks are becoming more and more widespread as the backbone networks for departmental and small regional computer systems. Companies that operate computer systems are attracted by two factors: lower operating costs and the ability to quick horizontal scaling. At the same time, due to the centralization of management and standardization of hardware and software

modules, efficiency indicators are being reduced, which leads to a deterioration in QoS indicators. However, if, when synthesizing or reconfiguring the architecture of a hyperconverged network, the specifics of this network are taken into account, then it is possible to affect the timing, for example, to reduce the delay time of requests when accessing elementary modules, to reduce the time to receive responses to transaction requests from users of the computer system, that is to minimize the average delay in message transmission.

Analysis of related works

The peculiarities of the functioning of hyperconverged networks are considered in [1, 2]. Various methods for improving QoS performance are widely presented in modern scientific literature, for example, in [3-8]. However, these papers do not take into account the peculiarities of hyperconverged networks. A number of papers take into account the specifics of the functioning of the hyperconverged network. So, in [9, 10], the methods of computing resource distribution are considered. Papers [11, 12] are focused on the specifics of certain tasks of computer system users. However, these papers did not consider the problem of minimizing the average delay in messages transmission in a hyperconverged network environment. In addition, the existing models do not consider the possibility of taking into account cost constraints during the operation of the system.

Goals

The problem of minimizing the average delay of message transmission in a hyperconverged network environment is currently relevant and is focused on improving QoS indicators. The purpose of the paper is to develop such a method for the synthesis of a hyperconverged network, which reduces the processing time of requests, but the total cost of data transmission does not exceed the specified values. The application of this method will increase the efficiency of using the computational resources of the hyperconverged network and will allow meeting the QoS requirements.

Method for message transmission time minimization in hyperconverged network environment

To achieve an efficient allocation of resources, it is proposed to use, as a limiting condition, the cost of transmission of the certain amount of information per single unit of bandwidth. Then the total amount of transmitted information will determine the total income from the use of communications. The total cost of the network can be determined at the final design stage, calculated, for example, as the sum of appropriate bandwidths. This approach will allow determining the payback period of the network, taking into account the costs of its implementation and income from its operation.

The optimization problem is formulated in the following form: determine the optimal values of the information flow density that minimizes the average delay

$$\bar{T} = \frac{1}{\gamma}\sum_{i=1}^{I} \rho_i \frac{\Sigma'_{m_i}}{\Sigma_{m_i}} \to \min , \qquad (1)$$

with a D_{req} limitation on the cost of transmission of the total amount of information per single unit of communication link bandwidth

$$D = \nu \sum_{i=1}^{I} \rho_i \le D_{req} . \qquad (2)$$

where $\Sigma_{m_i} = \frac{1 - \rho_i^{m_i+2}}{1-\rho_i}$; γ is aggregated network traffic, i.e. total number of requests entering the network per second; I is the number of elementary modules of the network under synthesis; ρ_i is utilization factor of the i-th elementary module of the network; m_i is the number of places in the queue for the i-th elementary module of the network; Σ' denotes the derivative of $\frac{\partial \Sigma}{\partial \rho}$; ν is normalizing factor.

The solution of the optimization problem (1)-(2) by the method of indefinite Lagrange multipliers determined the optimal value of the specific flow in the branches

$$\rho_{opt} = \frac{D_{req}}{\nu \cdot I}, \qquad (3)$$

providing the minimum value of the average delay:

$$\bar{T}^{\min} = \frac{1}{\gamma} \cdot \frac{D_{req}}{\nu} \cdot \left(\frac{\Sigma'_m}{\Sigma_m} \right)_{opt}, \quad 0 < \rho_{opt} < 1. \qquad (4)$$

Discussion and Conclusion

The result obtained in expression (4) shows the cost of data transfer. According to expression (3), we can conclude that in order to minimize costs, hyperconverged network should tend to isotropy in the sense of the constancy of the flow density values of the transmitted information to all elementary modules:

$$\rho_{opt} = \frac{D_{req}}{kn} < 1 \quad - \text{ for any elementary module.}$$

The curves of the dependence of average delay minimum values, which are calculated according to (4) at various values of buffers number in the switching nodes, on the optimal values of the specific flow are presented in the upper part of Fig. 1. Appropriate delay in a non-buffered system provides a lower bound for delay in a wide variety of multiple access systems with buffering and flow control.

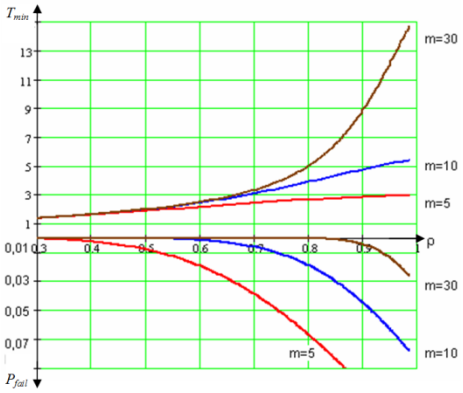


Fig. 9. Dependence of the average delay on the optimal values of the specific flow

The analysis of the curves allows concluding that the minimum average delay can be achieved with a higher information flow density transmitted over the network with a decrease in number of buffer memory elements in switching nodes in three times. In this case, the probability of failure increases to a completely acceptable value. Thus, the proposed analytical expressions allow, at a given cost of transferring a unit of information, selecting both the number of elements of the buffer memory of hyperconverged network modules and the optimal value of the information flow density that provides the minimum average delay in message transmission in hyperconverged network.

References

1. Merlac, V., Smatkov, S., Kuchuk, N., Nechausov, A.: Resourses Distribution Method of University e-learning on the Hypercovergent platform. In: Conf. Proc. of 2018 IEEE 9th Int. Conf. on Dependable Systems, Service and Technologies, DESSERT'2018, Kyiv, May 24-27, pp. 136-140, doi: http://dx.doi.org/ 10.1109/DESSERT.2018.8409114 (2018).
2. Kianpisheh, S., Glitho, R.H.: Cost-efficient server provisioning for deadline-constrained VNFs Chains: A parallel VNF processing approach. In: Proceeding of 2019 16th IEEE Annual Consumer Communications & Networking Conference, DOI: 10.1109/CCNC.2019.8651799. https://doi.org/10.1109/CCNC.2019.8651799 (2019).
3. Petrescu, Marius, Mădălina, Cuc, Ionica, Oncioiu: Analytical Modelling of Share Price Value Using Computational Intelligence Methods. In: Int. journal of computers communications & control, vol. 15, no. 4, pp. 1-12, doi: https://doi.org/10.15837/ijccc.2020.4.3910 (2020).
4. Lee S. R.: Dispersion-managed links formed of SMFs and DCFs with irregular dispersion coefficients and span lengths. In: Journal of Information Communication Convergence Engineering, vol. 16, no. 2, pp. 67-71, DOI:

10.6109/jicce.2018.16.2.67 (2018).
5. Ruban, I.V., Churyumov, G.I., Tokarev, V.V., Tkachov, V.M.: Provision of survivability of reconfigurable mobile system on exposure to high-power electromagnetic radiation. In: Selected Papers of the XVII International Scientific and Practical Conference on Information Technologies and Security (ITS 2017). CEUR Workshop Processing, pp. 105–111, (2017).
6. Chen, Y., Ran, R., Oh, S.: Performance analysis for pilot decontamination of massive MIMO using alternative projection. In Proceedings of IEEE International Conference on Ubiquitous and Future Networks, Vienna, Austria, pp. 297–300. DOI: 10.1109/ICUFN.2016.7537036 (2016).
7. Mukhin, V., Kuchuk, N., Kosenko, N., Kuchuk, H., Kosenko, V.: Decomposition Method for Synthesizing the Computer System Architecture. In: Advances in Intelligent Systems and Computing", AISC, vol. 938, pp 289-300, DOI: doi.org/10.1007/978-3-030-16621-2_27 (2020).
8. Semyonov, S.G., Gavrilenko, S.Y., Chelak, V.V.: Information processing on the computer system state using probabilistic automata. In: Proceedings - 2017 2nd International Ural Conference on Measurements, UralCon 2017, 2017, 2017-November, pp. 11-14 (2017).
9. Franti, P.: Efficiency of random swap clustering. In: Journal of Big Data, vol. 5, is. 13, pp. 1–29, DOI: doi.org/10.1186/s40537-018-0122-y (2018).
10. Ye, Q., Zhuang, W.: Distributed and adaptive medium access control for internet-of-things-enabled mobile networks. In: IEEE Internet of Things Journal, vol. 4, no. 2, pp. 446-460, doi: 10.1109/JIOT.2016.2566659. DOI: doi.org/10.1109/JIOT.2016.2566659 (2017).
11. Tkachov V., Hunko M., Volotka V.: Scenarios for Implementation of Nested Virtualization Technology in Task of Improving Cloud Firewall Fault Tolerance. In: 2019 IEEE International Scientific-Practical Conference Problems of Infocommunications, Science and Technology (PIC S&T), Kyiv, Ukraine, 2019, pp. 759-763, DOI: 10.1109/PICST47496.2019.9061473 (2019).
12. Lima A.A., de Barros F.K., Yoshizumi V.H., Spatti D.H., Dajer M.E.: Optimized artificial neural network for biosignals classification using genetic algorithm: In Journal of Control, Automation and Electrical Systems, 30(3), pp. 371–379 (2019).

ДЕЦЕНТРАЛІЗАЦІЯ ПРОБЛЕМНО-ОРІЄНТОВАНОЇ ПЛАТФОРМИ ТРАНСФЕРУ ЗНАНЬ

В.В. Циганок, М.М. Савченко, С.В.Каденко, О.В.Андрійчук
Інститут проблем реєстрації інформації НАН України, Київ, Україна
tsyganok@ipri.kiev.ua, zitros.lab@gmail.com, seriga2009@gmail.com,
andriichuk@ipri.kiev.ua

Проблема якомога більш повного використання наявних знань є нині актуальною проблемою глобального характеру. Ефективність застосування знань у різних галузях є одним із найважливіших чинників, що сприяють загальному прогресу. Відповідно до досліджень, що проводились американською компанією DelphiGroup на початку нинішнього тисячоліття, досить значна частка знань (більше ніж 40%), які використовуються певною організацією є неформалізованими і не зареєстровані на носіях інформації. Ці знання базуються лише на навичках, досвіді та інтуїції певних спеціалістів-експертів.

Не підлягає сумніву, що при прийнятті обґрунтованих рішень в різних сферах діяльності слід спиратись на всі наявні знання, як доступні на поточний момент, зареєстровані на певних носіях, так і експертні. Особливо це стосується слабко структурованих предметних областей, де рівень невизначеності та неповноти інформації, зазвичай, дуже високий.

Інтеграція знань різної природи, отриманих із різних джерел, що змінюються у часі, з експертними та формалізованими знаннями, їх обробка та зберігання у відповідних базах у вигляді моделей предметних областей може стати першим етапом вирішення вищезгаданої проблеми. Наступна передача (трансфер) накопичених актуальних знань для подальшого їх використання при підтримці прийняття рішень може бути завершаючим етапом вирішення.

Функціонально платформу трансферу знань доцільно розділити на:

1. підсистему збору і актуалізації знань (платформа підтримки роботи експертів, інженерів знань (організаторів експертиз), осіб, що приймають рішення (ОПР) з наступним функціоналом:
   - реєстрація та облік роботи користувачів платформи;
   - оцінювання компетентності у певному питанні інженерів знань та експертів;
   - формування баз знань інженерами знань, і тому числі з проведенням експертиз;
2. підсистема надання знань і ППР, до функцій якої належать:
   - вибір оптимального рішення для ОПР;
   - оцінювання варіантів рішень;
   - стратегічне планування.

Передбачається, що інженер знань, який є основним актором у системі володіючи певними знаннями в предметній області, користуючись інструментарієм автоматизованого екстрагування знань та організації

групових експертиз створює моделі предметних областей у вигляді відповідних баз знань. У подальшому, на основі створеної моделі предметної області генеруються рекомендації для осіб, що приймають рішення (ОПР). Функціонально, в платформі трансферу знань за генерацію рекомендацій відповідає система підтримки прийняття рішень (СППР), як складова платформи[1]. Ще однією складовою платформи може виступати професійна соціальна мережа, що об'єднує спеціалістів-експертів різних напрямків діяльності, як джерел отримання знань. Відносна компетентність цих спеціалістів щодо розглянутих питань, рівень якої може динамічно змінюватись у ході набуття досвіду роботи з платформою, є невід'ємним фактором, який враховується при узагальненні надаваної ними інформації та при отриманні ними винагороди за експертизу.

Необхідність децентралізації платформи трансферу знань

Під децентралізацією розуміється використання криптографічно захищеної децентралізованої платформи даних, які здебільшого працюють на основі технології блокчейн [2]. Децентралізація будь-якої системи є важливим, ледь чи не необхідним елементом, якщо існує необхідність забезпечення безвідмовності її роботи та абсолютної захищеності від несанкціонованого втручання і цілеспрямованих атак.

Децентралізація системи, тобто перетворення системи з повністю централізованої у частково або повністю децентралізовану, може виконуватись задля надання системі наступних переваг [3]:

- Гарантія надійності збереження даних в довічному, глобальному розподіленому реєстрі даних. Такий реєстр не може зберігати великі об'єми даних, але в зберіганні великої кількості даних зазвичай немає потреби.
- Неможливість підробки збережених даних способом, який не був передбачений при створенні такого реєстру даних, в тому числі підміну, некоректне додавання та видалення даних з реєстру.
- Виконання лише строго передбачених обчислень та операцій над даними, що зберігаються в децентралізованому реєстрі.

Розглядаючи децентралізацію систем, що будуються на децентралізованих платформах даних, варто відокремлювати наступні три типи децентралізації: повна,часткова децентралізація та відсутність децентралізації. Рівень "децентралізації" також визначається не тільки способом застосування децентралізованих платформ даних по відношенню до системи, що з нею інтегрується, а і децентралізацією самої децентралізованої платформи даних.

При розробці розподіленої системи збору експертних знань, за допомогою часткової децентралізації можна надати СППР наступних важливих властивостей:

- Гарантовану надійність зберігання експертних знань, що включає довічне безвідмовне зберігання невеликої (але достатньої) кількості даних для проведення обчислень в СППР, повну захищеність від

несанкціонованих змін та повну доступність даних. Дана властивість є особливо затребуваною для експертних систем, адже нерідко дані одного проекту, після його завершення, необхідно зберігати роками з моменту завершення проекту.

– Можливість проведення публічного аудиту системи, на основі достовірних вхідних даних, що зберігаються в децентралізованому реєстрі. Оскільки дані, будучи збережені в децентралізованому реєстрі є криптографічно захищеними, немає жодних підстав сумніватись у їх автентичності (очевидність даного факту також досягається завдяки тому, що будь-яка децентралізована платформа передбачає запис всієї історії змін даних).
– Як результат, підвищення довіри до функціонування платформи трансферу знань в цілому, що також забезпечує довіру до результатів обробки отриманих даних СППР.

Альтернативою децентралізації підсистеми збору знань системи підтримки прийняття рішень є використання електронно-цифрового підпису та декількох централізованих реєстрів даних (традиційних баз даних). Варто зазначити, що навіть використовуючи ці альтернативи, система збору знань все одно не матиме тих властивостей, які були перераховані вище.

Запропонована модель системи підтримки прийняття рішень з повною децентралізацією системи збору експертних знань

З проведеного аналізу, в системі збору експертних данихзагальний розмір вхідних експертних оцінок, що записуються в децентралізований реєстр є невеликим (менше 500 кілобайт на один проект). Враховуючи те, що результати роботиСППР можуть використовуватись та переглядатися протягом тривалого періоду (не один рік), для такої системи має сенс записувати всю вхідну інформацію в децентралізований реєстр. Іншими словами, провести повну децентралізацію системи збору експертних даних, що буде означати часткову децентралізацію СППР, оскільки система збору експертних даних є її підсистемою.

Однак, децентралізація системи завжди дещо ускладнює роботу користувача з нею, оскільки тепер для взаємодії з системою необхідно мати не тільки децентралізований акаунт (також відомий як "гаманець" в криптографічних децентралізованих платформах даних), а і певну кількість криптовалюти для оплати транзакцій. Дана проблема класифікується як проблема адаптації технології і є однією з причин чому багато систем та продуктів не можуть інтегруватися з децентралізованими технологіями [4]. Однак, за допомогою розробленого методу делегування транзакцій [5], дана взаємодія спрощується лише до створення акаунта в децентралізованій системі, без втрати будь-яких властивостей децентралізації.

Після децентралізації системи збору експертних даних, вона не буде мати принципових відмінностей від попередньої системи, окрім як до неї додатково будуть додані декілька додаткових підготовчих кроків для

експертів. Таким чином, підготовка до проведення децентралізованого експертного оцінювання міститиме додатково такі кроки:

– Підготовка спеціалізованої децентралізованої програми для проекту, що досліджується за допомогою СППР. Дана програма може бути шаблонізована, тобто розроблена один раз, і лише незначним чином модифікована для кожного нового застосування.
– Створення акаунта-делегата та забезпечення його криптовалютою, необхідною для проведення децентралізованих транзакцій. Даний крок можна виконати лише один раз для всіх подальших досліджень.

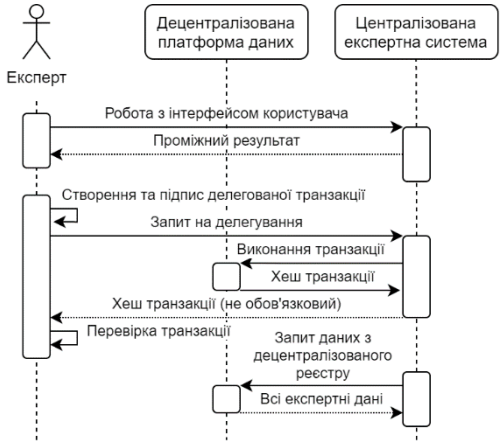
Для експертів, що працюють з системою, пропонується декілька додаткових нескладних кроків: перед експертним оцінюванням, кожний експерт генерує свій власний децентралізований акаунт, і за допомогою графічного інтерфейсу користувача реєструє його в децентралізованій програмі (окремо для кожного проекту-експертизи або один раз), тим самим затверджуючи свою ідентичність, створюючи відповідність експерта до певного децентралізованого акаунту та будь-яких додаткових даних, як, наприклад, його e-mail адреса або телефон. Ніхто крім самого експерта не може записувати оцінки від його імені після такої реєстрації в децентралізованій системі.

Послідовність підготовки та роботи є наступною:

1. Адміністратор розробляє (або клонує уже існуючу) децентралізовану програму, що спеціалізується на конкретній експертній системі (або є шаблонною). Функцією даної програми є верифікований запис експертних даних в децентралізований реєстр за допомогою делегата, для подальшого зчитування цих даних системою підтримки прийняття рішень.
2. Адміністратор реєструє дану програму в децентралізованій платформі даних, тим самим роблячи її незмінною. Отримується унікальна адреса програми, яка також є незмінною.
3. Отримана адреса децентралізованої програми записується в централізоване сховище (або програмний код) експертної системи для подальшої взаємодії з нею.
4. Додатково, адміністратор генерує акаунт-делегат, і поповнює його баланс в децентралізованій системі. Акаунт-делегат записується до системи підтримки делегованих транзакцій, яка може бути частиною системи підтримки прийняття рішень.
5. Перед початком роботи з системою, експерт створює свій децентралізований акаунт (наприклад, за допомогою розширення для браузеру Metamask), та реєструє його в системі шляхом створення електронно-цифрового підпису певних даних (наприклад, своєї email адреси).
6. Після проведення роботи з системою, експерт виконує делеговану транзакцію для запису ним згенерованих даних в децентралізовану програму. Особливістю даної транзакції є її непідробність, тобто, ніхто крім самого експерта не може змінити ці дані, і сам експерт може

впевнитися в тому, що в децентралізовану мережу були записані правильні дані (хоч і в цьому немає потреби), які ніколи не будуть змінюватись, якщо це не передбачено самою децентралізованою програмою.

При проведенні самого експертного оцінювання СППР, дані зчитуються з децентралізованого реєстру та конвертуються у формат зручний для обробки. В подальшому, дані з децентралізованого реєстру можуть бути використані також, наприклад, для аудиту певного проекту. На рисунку зображено взаємодію експерта та централізованої системи по збору експертних даних з децентралізованою програмою до та після проведення експертизи (робота експерта та зчитування даних з децентралізованого реєстру).



Результати децентралізації системи збору експертних знань

Після децентралізації, тобто при проведені запропонованих змін у порівнянні з раніше централізованою системою збору експертних знань, вона набула наступних властивостей:

1. Вхідні дані, отримані від експерта, тепер є захищеними та підтримуються всесвітньою децентралізованою платформою даних, що гарантує їх практично довічне зберігання та неможливість підробки.
2. Оскільки дані є криптографічно захищеними, і записані одного разу дані є незмінними, вони вважаються достовірними (за умови правильності децентралізованої програми, яку можна легко перевірити в будь-який момент). Це дозволяє підвищити рівень довіри до системи підтримки прийняття рішень в цілому, та проводити постфактум аудит системи, якщо для цього виникне необхідність.

Вище наведені властивості вдалося додати до системи лише з наступними мінімальними змінами до наявної СППР та підсистеми збору експертних даних:

1. Необхідність для експертів встановлення додаткового (відносно простого) програмного забезпечення у вигляді або мобільного додатку, або плагіну для браузера. В свою чергу, даний підхід може також повністю замінити крок реєстрації в системі збору експертних даних за допомогою логіну та паролю.
2. Додавання двох нових кроків для експерта: електронний підпис даних до (реєстрація) та після (виконання транзакції запису даних в децентралізовану програму) роботи з системою.

Список літератури

1. Tsyganok V.V., Borokhvostov I.V., Roik P.D. Problem-oriented knowledge transfer platform for decision making support in socio-technical systems / CEUR Workshop Proceedings, Vol. 2067 Selected Papers of the XVII International Scientific and Practical Conference on Information Technologies and Security (ITS 2017); Kyiv, Ukraine, November 30, 2017. P.112-117.
2. Скічко Д. Безпека технології блокчейн для децентралізованих систем / Д. Скічко, Т. О. Гріненко, О. П. Нарєжній // Global Cyber Security Forum : матеріали Першого міжнародного науково-практичного форуму, 14 – 16 листопада 2019 р. – Харьков : ХНУРЭ, 2019. – С. 98–99.
3. Savchenko M., Tsyganok V., Andriichuk O. A Cost-Effective Approach to Securing Systems through Partial Decentralization. / Information & Security: An International Journal 47, no. 1 (2020): 109-121.
4. Chod, J., Trichakis, N., Tsoukalas, G., Aspegren, H. And Weber, M., 2020. On the financing benefits of supply chain transparency and blockchain adoption. Management Science.
5. Savchenko M., Tsyganok V., Andriichuk O. An Approach to Transaction Delegationin Self-protected Decentralized Data Platforms / CEUR Workshop Proceedings, Vol. 2577 Selected Papers of the XIX International Scientific and Practical Conference on Information Technologies and Security (ITS 2019); Kyiv, Ukraine, November 28, 2019. P.169-188.

ВІЗУАЛІЗАЦІЯ ГРУПОВОЇ ГРИ ПЕРЕСЛІДУВАННЯ НА ПЛОЩИНІ

Барановська Л.В.[0000-0003-0024-8180], Гирявець Д.М.[0000-0003-1310-0943],
Барановська Г.Г. 3[0000-0002-6878-6973]
КПІ ім. Ігоря Сікорського, Київ, Україна
lesia@baranovsky.org

У даній роботі розглянуто диференціальні ігри групового переслідування. Наведено схему методу розв'язуючих функцій А.О. Чикрія. Знайдено достатні умови для закінчення гри і одержано рівняння для чисельного знаходження розв'язуючих функцій. Сформульовано достатні умови для закінчення гри для процесу з простими матрицями. Для задачі простого переслідування реалізовано візуалізацію траєкторії рухів переслідувачів і втікача на площині. Для цього був створений програмний продукт на мові програмування Python. Даний програмний продукт є прототипом системи моделювання «переслідувачі-втікач», яку можна буде застосовувати для вибору керування в задачах переслідування. У розробці було розглянуто два випадки вибору керування втікача. В першому, керування втікача будується за наступним алгоритмом: втікач визначає найближчого переслідувача в плані евклідової норми; втікач будує своє керування на промені, який виходить з положення найближчого переслідувача та перетинає положення втікача, у напрямку протилежному переслідувачу з максимальною швидкістю. В другому випадку, керування втікача задає користувач.
Ключові слова: конфліктно-керовані процеси, диференціальні ігри, ігри переслідування, групова задача переслідування.

Схема методу

У теорії динамічних ігор (конфліктно-керованих процесів, диференціальних ігор) розроблено ряд методів, які забезпечують гарантований результат. До таких методів належать, зокрема, перший прямий метод Л.С. Понтрягіна, метод екстремального прицілювання М.М. Красовського і метод розв'язуючих функцій А.О. Чикрія [1]. У даній роботі застосований буде саме останній метод, який дає теоретичне обґрунтування класичного правила паралельного переслідування і методу зближення за променем. Схема методу для диференціально-різницевих ігор розроблено в роботах [2, 3, 4]. У роботах [5–8] розглянуто задачу зближення для групи переслідувачів та одного втікача. Схему методу розв'язуючих функцій для диференціально-різницевих систем нейтрального типу розроблено у роботі [9]. Для об'єктів з різною інерційністю у роботі [10] запропоновано модифікацію методу. Ігрові проблеми зближення при відмові керуючих пристроїв розглянуто у роботах [11, 12].

Розглянемо рух керованого об'єкта, який описується системою диференціальних рівнянь:

$$\dot{z}_l = A_l z_l + \varphi_l(u_l, v),\ z_l \in R^{n_l},\ l = \overline{1,\vartheta},\ u_l \in U_l,\ v \in V \qquad (1)$$

де A_l – квадратна матриця порядку $n_l, n_1 + \cdots\ n_\vartheta = n$, U_l та V – непорожні компакти з деяких скінченновимірних просторів, вектор-функція $\varphi_l(u_l, v): U_l \times V \to R^{n_l}$ неперервна за сукупністю змінних. Нехай $z(0) = z^\circ$– початковий стан процесу (1). Термінальна множина є циліндричною і має вигляд:

$$M^* = \cup_{l=1,\dots,\vartheta}\{M_l^\circ + M_l\} = \cup_{l=1,\dots,\vartheta} M_l^* \qquad (2)$$

де M_l° - лінійний підпростір з R^{n_l}, M_l – випуклий компакт з ортогонального доповнення L_l до M_l в просторі R^{n_l}.

Гра (1), (2) вважається закінченою, якщо для деякого $l = \overline{1,\vartheta}$ виконується умова $z_l \in M_l$. Переслідувачі використовуть квазістратегії, а втікач - програмне керування. Нехай π_l – оператор ортогонального проектування з R^{n_l} на L_l. Розглянемо багатозначні відображення $W_l(t, v) = \pi_l e^{A_l t}\varphi_l(U_l, v),\ W_l(t) = \cap_{v\in V} W_l(t, v), t \geq 0, v \in V.$

Умова Понтрягіна. Відображення $W_l(t) \neq \emptyset$ для всіх $l = \overline{1,\vartheta}, t \geq 0.$

Виберемо борелівський селектор $\gamma_l(t)$ відображення $W_l(t)$. Зафіксуємо його та покладемо $\xi_l\big(t, z_l, \gamma_l(\cdot)\big) = \pi_l e^{A_l t} z_l + \int_0^t \gamma_l(\tau)d\tau$.Кожному переслідувачу поставимо у відповідність розв'язуючу функцію

$$\alpha_l\big(t, \tau, z_l, v, \gamma_l(\cdot)\big) =$$
$$= \sup\{\alpha_l \geq 0: [W_l(t-\tau, v) - \gamma_l(t-\tau)] \cap \alpha_l\left(M_l - \xi_l\big(t, z_l, \gamma_l(\cdot)\big)\right) \neq \emptyset\} \qquad (3)$$

Якщо $\xi_l\big(t, z, \gamma_l(\cdot)\big) \in M_l$, то покладемо $\alpha_l\big(t, \tau, z, v, \gamma_l(\cdot)\big) = +\infty, 0 \leq \tau \leq t, v \in V$. В інших випадках $\alpha_l\big(t, \tau, z, v, \gamma_l(\cdot)\big)$ обмежена при будь-яких $\tau \in [0, t],\ v \in V$. Візьмемо $\gamma(\cdot) = column(\gamma_1(\cdot), \dots, \gamma_\vartheta(\cdot)$ і нехай$\Gamma_\vartheta = \{\gamma(\cdot): \gamma_l(t) \in W_l(t), t \geq 0, l = \overline{1,\vartheta}\}, T_\vartheta\big(z, \gamma(\cdot)\big) = \min\{t \geq 0: \inf_{v(\cdot)\in\Omega_V} \max_{l=1,\dots,\vartheta} \int_0^t \alpha_l\big(t, \tau, z_l, v(\tau), \gamma_l(\cdot)\big)\, d\tau \geq 1\}.$

Теорема 1 [1]. Нехай для конфліктно-керованого процесу (1), (2), виконується умова Понтрягіна, z°- початковий стан процесу (1), і для деякого борелівський селектор $\gamma^\circ(t) \in \Gamma_\vartheta$ виконується нерівність $T_\vartheta\big(z^\circ, \gamma^\circ(\cdot)\big) < +\infty$. Тоді траєкторія процесу (1) хоча б одного l може бути приведена на термінальну множину M_l^* з початкового стану z° в момент $T_\vartheta\left(z^\circ, \gamma^\circ(\cdot)\right)$.

Приклад задачі для процесу з простими матрицями

Розглянемо задачу групового переслідування з ϑ переслідувачами та одним втікачем:

$$\dot{z}_l = a_l z_l + u_l - v, \qquad a_l < 0, z_l \in R^k, || u_l|| \leq 1, || v|| \leq 1, l = \overline{1,\vartheta} \qquad (4)$$

Гра вважається закінченою, якщо для деякого l $x_l = y$. Термінальна множина $M^* = \cup_{l=1,\dots,\vartheta}\{M_l^\circ\} = \cup_{l=1,\dots,\vartheta}\{z_l: z_l = 0\}$. $W_l(t) = \{0\}, l = \overline{1,\vartheta},$ отже селектори $\gamma_l(t) = 0$ функції $\xi_l\big(t, z_l, \gamma_l(\cdot)\big) = e^{a_l t} z_l, l = \overline{1,\vartheta},$ розв'язуючі

функції переслідувачів $z_l \neq 0 \quad \alpha_l(t,\tau,z_l,v,0) = \max\{\alpha_l \geq 0: -\alpha_l e^{a_l t} z_l \in e^{a_l(t-\tau)}(S-v)\} = e^{-a_l\tau}\alpha_l(z_l,v)$, де $\alpha_l(z_l,v)$ має вигляд

$$\alpha_l(z_l,v) = \frac{(z_l,v)+((z_l,v)^2+\|z_l\|^2(1-\|v\|^2))^{1/2}}{\|z_l\|^2}, l = \overline{1,\vartheta}. \quad (5)$$

Час завершення групового переслідування:

$T_\vartheta(z) = \min\left\{t \geq 0: \inf_{v(\cdot)\in\Omega_V} \max_{l=1,\dots,\vartheta} \int_0^t e^{-a_l\tau}\alpha_l\left(z_l, v(\tau)\right) d\tau = 1\right\}$.

Позначимо $\delta(z) = \min_{\|v\|\leq 1} \max_{l=1,\dots,\vartheta} \alpha_l(z_l,v)$, отримали обмеження зверху

$$T_\vartheta(z) \leq \frac{1}{a_*}\ln\left(1+\frac{a_*\vartheta}{\delta(z)}\right), \quad (6)$$

де $a_* = -\max\limits_{l=1,\dots,\vartheta} a_l$ [1].

Теорема 2 [1]. Нехай z°- початковий стан (4). Тоді, якщо $0 \in int\, co\{z_l^\circ\}$, то задача групового переслідуваня вирішувана в момент $T_\vartheta(z^\circ)$, для якого справедлива оцінка (6). При цьому, якщо $t_* = t_*(v(\cdot))$ – момент переключення $t_* \leq T_\vartheta(z^\circ)$, який є нулем контрольної функції $1 - \max_{l=1,\dots,\vartheta} \int_0^t e^{-a_l\tau}\alpha_l\left(z_l, v(\tau)\right) d\tau$, то керування переслідувачів, які реалізують час $T_\vartheta(z^\circ)$, на інтервалі $[0,t_*]$ має вигляд $u_l(\tau) = v(\tau) - \alpha_l\left(z_l^\circ, v(\tau)\right) z_l^\circ, l = \overline{1,\vartheta}$,а на інтервалі $(t_*, T_\vartheta(z^\circ)] \quad u_l(\tau) = v(\tau)$ для індексів l, які задовільняють рівності $\int_0^t e^{-a_l\tau}\alpha_l\left(z_l^\circ, v(\tau)\right) d\tau = \max\limits_{l=1,\dots,\vartheta} \int_0^t e^{-a_l\tau}\alpha_l\left(z_l^\circ, v(\tau)\right) d\tau$, для інших індексів керування довільні на інтервалі $\left(t_*, T_\vartheta(z^\circ)\right]$. Якщо $0 \notin int\, co\{z_l^\circ\}$, то вирішувана задача втечі, яка реалізується на будь-якому постійному керуванні втікача, на якому досягається мінімум функції $\max_{l=1,\dots,\vartheta} \alpha_l\left(z^\circ, v\right)$.

Візуалізація простого переслідування на площині

Нехай є ϑ переслідувачів та один втікач. Вважаємо, що положення учасників є геометричні точки, тобто не враховуємо їх розміри. Закони рухів переслідувачів мають вигляд $\dot{x}_l = u_l, \quad x_l \in R^k, k \geq 1, \|u_l\| \leq a_l, l = \overline{1,\vartheta}$. Закон руху втікача $\dot{y} = v, \quad y \in R^k, \|v\| \leq b$. Нехай $a_l \geq b, l = \overline{1,\vartheta}$ та існує певна дискретизація часу.

Для представлення траєкторій рухів переслідувачів та втікача у грі для випадку площини, був створений програмний продукт на мові програмування Python. Для програми вхідними даними є наступні відомості: кількість переслідувачів; початкові координати всіх учасників; максимальні значення швидкості для всіх учасників. Для того, щоб алгоритм працював потрібен вибір керування втікача. У розробці було розглянуто два випадки вибору керування втікача. В першому, керування втікача будується за наступним алгоритмом: 1) втікач визначає найближчого переслідувача в плані евклідової норми; 2) втікач будує своє керування на промені, який

виходить з положення найближчого переслідувача та перетинає положення втікача, у напрямку протилежному переслідувачу з максимальною швидкістю. В другому випадку, керування втікача задає користувач.

Література

1. Chikrii, A.: Conflict-Controlled Processes. Springer Science &Business Media (2013).
2. Baranovska, L.V.: Quasi-Linear Differential-Deference Game of Approach. In: Sadovnichiy, V., Zgurovsky, V. (eds.) Modern Mathematics and Mechanics. Understanding Complex Systems, pp. 505–524. Springer, Cham (2019).
3. Baranovska, L.V.: On Quasilinear Differential-Difference Games of Approach. Journal of Automation and Information Sciences 49(8), 53–67 (2017).
4. Baranovska, Lesia V.: Pursuit differential-difference games with pure time-lag. Discrete and Continuous Dynamical Systems – Series B 24(3), 1024–1031 (2019).
5. Baranovskaya, G.G., Baranovskaya, L.V.: Group Pursuit in Quasilinear Differential-Difference Games. Journal of Automation and Information Sciences 29(1), 55–62 (1997).
6. Baranovskaya, L.V., Chikrij, A.A., Chikrij, Al.A.: Inverse Minkowski functionals in a nonstationary problem of group. Izvestiya Akademii Nauk. Teoriya i Sistemy Upravleniya (1), 109–114 (1997).
7. Baranovskaya, L.V., Chikrij, A.A., Chikrij, Al.A.: Inverse Minkowski functionals in a nonstationary problem of group. Journal of Computer and Systems Sciences International 36(1), 101–106 (1997).
8. Барановская, Л.В., Барановская, Г.Г.: О дифференциально-разностной игре группового преследования. Доповіді Національної академії наук України (3), 12–15 (1997).
9. Baranovskaya, L.V.: A method of resolving functions for one class of pursuit problems. Eastern-European Journal of Enterprise Technologies, vol.2, 4(74), 4–8 (2015).
10. Baranovska, L.V.: Method of resolving functions for the differential-difference pursuit game for different-inertia objects. In: Sadovnichiy, V., Zgurovsky, V. (eds.) Advances in Dynamical Systems and Control, vol.69, pp. 159–176 (2016).
11. Chikrij, A.A., Baranovskaya, L.V., Chikrij, Al.A.: The game problem of approach under the condition of failure of controlling devices. Problemy Upravleniya i Informatiki (Avtomatika) (4), 5–13 (1997).
12. Baranovskaya, L.V., Chikrii, Al.A.: Game Problems for a Class of Hereditaly Systems. Journal of Automation and Information Sciences 29(2-3), pp. 87–97 (1997).

GENERAL CASE OF WAVELET TRANSFORM WITH REDUCIBLE RATIONAL DILATION FACTOR

Volodymyr Malchykov[0000-0002-1710-9171]
Igor Sikorsky Kyiv Polytechnic Institute, Kyiv, Ukraine
mavr2k@gmail.com

Dataset of different nature can be effectively analyzed with the help of wavelet analysis. There are two types of discrete wavelet transforms – dyadic and non-dyadic. The last one can give more accurate detection and separation of features that are present in analyzed data. A lot of methods use the irreducible fractions as a dilation factor for rational wavelet transform. In this paper the general case of reducible rational dilation factor will be considered. The procedure for building filters will be shown as well as the perfect reconstruction condition. Also, an approach for selecting the best reducible rational dilation factor will be proposed.
Keywords: Wavelet Transform, Non-Dyadic Wavelet, Dilation Factor.

Introduction

Wavelet transform (WT) is wildly used in order to analyze datasets of various natures. WT has proved its efficiency in analysis of medical signals [1], processing of multimedia data [2], speech and image recognition, etc.

According to the value of dilation factor discrete WT are classified into dyadic, where dilation factor equals 2, and non-dyadic in other cases. Dyadic WT are often used, but non-dyadic wavelet transform can be more suitable for precise localization of signal singularities and similar tasks.

Various authors proposed their own approaches to non-dyadic WT. Their properties and main features are shortly described in previous work [3].

Rational multiresolution analysis [4] looks like the simplest, but the most effective method among all others.

Problem Formulation

Usually the irreducible dilation factor is used in rational wavelet transforms. But in some cases using the reducible dilation factor can improve the quality of wavelet analysis.

The conditions for building filters for the dilation factor that equals 6/4 was shown earlier [3, 5]. In this work these conditions will be generalized to the arbitrary reducible fraction.

Also, the criteria for selecting the best value of rational dilation factor from a set of proportional fractions will be introduced. Proposed approach will be illustrated with the help of geomagnetic data.

References

1. Hussain, L., Aziz, W.: Time-Frequency Wavelet Based Coherence Analysis of EEG in EC and EO during Resting State. International Journal of Information Engineering and Electronic Business 7(5), 55–61, (2015).
2. Rhif, M., Ben Abbes, A., Farah, I.R., Martinez, B., Sang, Y.: Wavelet Transform Application for/in Non-Stationary Time-Series Analysis: A Review. Applied Science 9, 1345, (2019).
3. Chertov, O., Malchykov, V.: Rational wavelet transform with reducible dilation factor. In: Selected Papers of the XIX International Scientific and Practical Conference “Information Technologies and Security” (ITS-2019), pp. 146–158. (2020).
4. Baussard, A., Nicolier, F., Truchetet, F.: Rational multiresolution analysis and fast wavelet transform: application to wavelet shrinkage denoising. Signal Processing 84(10), 1735–1747 (2004).
5. Chertov O., Malchykov V.: Perfect Reconstruction Condition for Rational Wavelet Transform with Reducible Rational Dilation Factor. In: Shkarlet S., Morozov A., Palagin A. (eds) Mathematical Modeling and Simulation of Systems (MODS'2020). Advances in Intelligent Systems and Computing, vol. 1265, pp. 248-254. (2020).

PROCESS APPROACH IN THE TASKS OF ENSURING INFORMATION SECURITY OF COMPUTER NETWORKS

Dmytro Holubnychyi[1][0000-0001-8927-0055], Vitalii Martovytskyi[2][0000-0003-2349-0578], Oleksandr Sievierinov[2][0000-0002-6327-6405], Oleksandr Permiakov[3][0000-0002-2206-3761] and Vyacheslav Tretiak[4][0000-0003-2599-8834]

[1] Simon Kuznets Kharkiv National University of Economics, Kharkiv, Ukraine

[2] Kharkiv National University of Radioelectronics, Kharkiv, Ukraine

[3] National Defence University of Ukraine named after Ivan Cherniakhovskyi, Kyiv, Ukraine

[4] Ivan Kozhedub Kharkiv National Air Force University, Kharkiv, Ukraine

The management system is one of the key concepts of the organization theory, closely related to the purposes, functions, management process, activity of managers and distribution of power between them in performance of certain purposes. Within this system, the entire management process takes place, which involves managers of all levels, categories and professional specialization. The management system of the organization is built so that all the processes taking place in it, carried out in a timely manner and quality. Therefore, the heads of organizations and specialists pay special attention to it, with the aim of continuous improvement, development of both the system as a whole and its individual components. It is obvious that the study and improvement of the management system, both within an individual organization and the state, society as a whole, contributes to the rapid achievement of goals and objectives.

Of course, a modern enterprise, regardless of its size and type of activity, deals with a large amount of information, which is simply impossible to process without the help of computers and computer networks. Timely processing and obtaining the necessary information, ensuring its safety and confidentiality is the key to successful business.

Today to increase the efficiency of the enterprise, computer networks are used, which are the result of the evolution of computer technology. All these and many other benefits of a computer network make it clear that it is necessary for today's business.

Keywords: information system, process, administration, security.

INTRODUCTION

The purpose of the information system in any organization is to ensure continuous operation and minimize damage from events that threaten information security. This requires minimizing the effects of threats. The information system is a component of the management system in the organization, which provides an opportunity to adequately respond to external and internal influences. This will allow the organization be adapted to changing conditions, i.e. make it self-regulating.

Regarding management techniques, it should be noted that the harmony and efficiency of the management system largely depend on the document management system of the organization. The number of accounting and planning errors and the efficiency of responding to a certain impact proportionally depend on it [1].

Information security management allows to use information collectively, while ensuring its protection and protection of computing resources. Information security consists of three main components: confidentiality; integrity; accessibility.

- confidentiality: protection of confidential information from unauthorized disclosure or interception;
- integrity: ensuring the accuracy and completeness of information and computer programs;
- accessibility: ensuring the availability of information and vital services to users when needed.

Confidentiality, integrity, and accessibility can be affected by a variety of processes, but we will only consider issues related to the administration of computer systems and networks. Also will be formulate requirements for the security of personnel, including their physical security.

REQUIREMENTS FOR INFORMATION SECURITY MANAGEMENT

To reflect the developed approaches we use the notation of the IDEF0 standard [2, 3]. The main purpose of the model is to reflect the processes of information security system management in the organization. To initiate and control the process of information security we need to create an appropriate management structure.

The level of detail and formality of the procedures required for the administration and operation of computer systems and networks varies significantly depending on the size of the organization and the type of equipment, as well as the nature and vulnerability of software applications (fig.1).

Using the decomposition of the information security system management process, we define three main processes (fig.2).

Decomposition administration process of computer systems and networks defines such processes: definition of procedures; definition of software requirements; determination of requirements for computer systems planning; protection of in-machine information resources; computer network security management.

Definition of procedures. The decomposition of the procedure for defining procedures includes such processes as: definition of operational procedures and responsibilities; documentation of operating procedures; defining procedures for responding to events [4].

Defining software requirements. Decomposition of the process of determining procedures includes such processes as: classification of software development and work programs; malware protection.

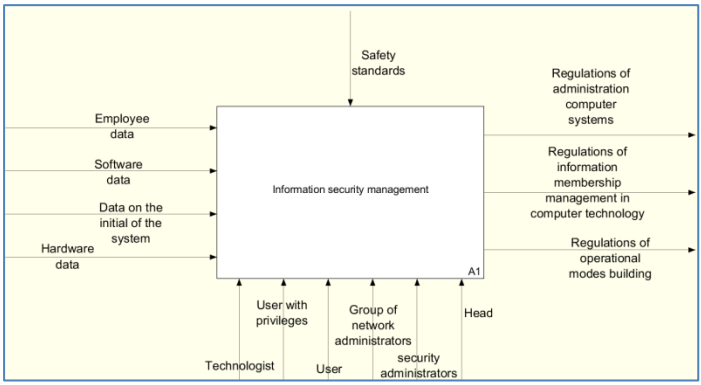


Fig. 1. Context diagram of the information security system management process

1. Administration of computer systems and computer networks (level A11);
2. Ensuring the safety of staff (level A12);
3. Ensuring physical security (level A13);

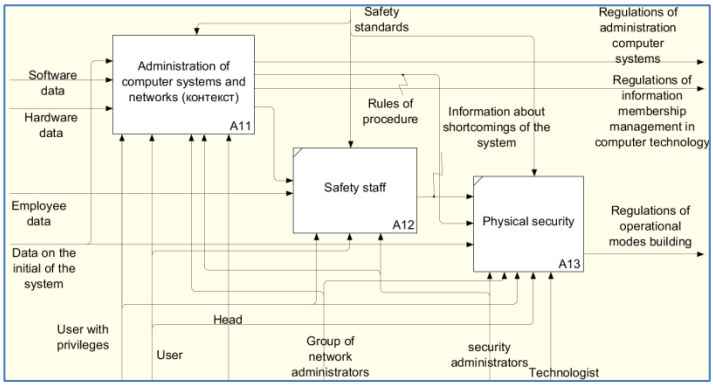


Fig. 2. Diagram of functional decomposition of the information security system management process

Defining requirements for computer systems planning. Decomposition of the process of determining the requirements for the planning of computer systems includes such processes as: load planning of computer systems; acceptance of computer systems; registration of failures in computer systems; managing the process of making changes to operating systems; maintenance of computer systems [5].

Protection of in-machine information resources. Decomposition of the protection process of in-machine information resources includes such processes as: backup of data; keeping a log of events; protection against industrial espionage.

Security management of computer networks. Decomposition of the security management process of computer networks includes such processes as: prevention of damage to information resources; protection of system documentation; data and program exchange; protection of electronic data exchange; email protection; e-office system protection.

CONCLUSION

Thus, the result of the process approach to the administration of computer systems and networks should be considered as follows: defining the regulatory model of network and system administrator at the macro level, taking into account the processes of information security in the organization; developed a set of interconnected models that completely determine the microstructure of the information security system. This set of models essentially constitutes organizational and methodological support for the design of organizational information security system

References

1. Semyonov, S.G., Gavrilenko, S.Y., Chelak, V.V.: Information processing on the computer system state using probabilistic automata. In: Proceedings - 2017 2nd International Ural Conference on Measurements, UralCon 2017, 2017, 2017-November, pp. 11-14 (2017).
2. Kim, Soung-Hie, and Ki-Jin Jang. "Designing performance analysis and IDEF0 for enterprise modelling in BPR." International Journal of production economics 76.2 (2002): 121-133.
3. Waissi, Gary R., et al. "Automation of strategy using IDEF0—A proof of concept." Operations Research Perspectives 2 (2015): 106-113.
4. I. Ruban, V. Martovytskyi and N. Lukova-Chuiko, "Designing a monitoring model for cluster super-computer", Eastern-European Journal of Enterprise Technologies, vol. 6, no. 84, pp. 32-37, 2016.
5. Ruban, I.V., Churyumov, G.I., Tokarev, V.V., Tkachov, V.M.: Provision of survivability of reconfigurable mobile system on exposure to high-power electromagnetic radiation. In: Selected Papers of the XVII International Scientific and Practical Conference on Information Technologies and Security (ITS 2017). CEUR Workshop Processing, pp. 105–111 (2017).

АДАПТИВНИЙ ПІДХІД ДО ПОБУДОВИ МОДЕЛЕЙ РИЗИКІВ ФІНАНСОВИХ СИСТЕМ

Наталія Кузнєцова

Інститут прикладного системного аналізу Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського», Київ, Україна

natalia-kpi@ukr.net

У доповіді викладені основні питання розробки адаптивного підходу менеджменту ризиків фінансових систем. Запропоновано створення узагальненої інформаційної системи підтримки прийняття рішень, наведено її архітектура, визначено основні складові. Принцип адаптивного менеджменту ризиків передбачає реалізацію двох контурів адаптації: внутрішнього і зовнішнього, що забезпечує динамічне оцінювання фінансових ризиків в реальному часі через ймовірнісну та вартісну складову. Використання двох контурів адаптації дозволяє одразу частково подолати ризик, а також скорегувати рішення у випадку, якщо воно виявилось недостатньо ефективним.

Ключові слова: адаптивний менеджмент ризиків, інформаційна система підтримки прийняття рішень, скорингові моделі, структурно-параметрична адаптація.

Вступ

Коронакриза 2020 року стала значним стимулом змін для майже всіх систем, які функціонують у сучасному світі. Розрив зв'язків між системами, країнами, континентами, зупинка промисловості, компаній і заводів, транспорту, закриття кордонів, обмеження свободи пересування і логістики виявили той самий тісний взаємозв'язок, вплив і симбіоз систем, який навіть не простежувався на перший погляд. Всі системи змушені були реагувати на нові обставини і виклики, шукати компроміси, адаптуватись до змін, ідентифікувати для себе нові шляхи розвитку, враховуючи нові обставини. Фактично, тепер ми спостерігаємо саме адаптацію і напрацювання нових стратегій і рішень, динамічне реагування і врахування нових факторів, прийняття рішень в умовах невизначеності. Насправді ефективність функціонування системи у таких реаліях буде визначатись тим, наскільки ефективно менеджмент проаналізував можливі шляхи розвитку ситуації і напрацював ефективні альтернативні управлінські рішення у разі несприятливого розвитку ситуації, тобто наскільки система підготована, налаштована,і здатна адаптуватись.

Було розроблено метод, який базується на принципі адаптивного менеджменту, і спрямований на урахування нових даних, критеріїв та навіть появу нових ризиків, уточнюючи існуючі моделі за рахунок адаптації структури і параметрів [1, 2], або обирати альтернативний метод і

створювати нову модель, яка враховує як історичні дані, так і нові дані, пов'язані з критичними змінами зовнішніх умов. Це дозволить знизити ризики прийняття помилкових управлінських рішень та надасть ефективний інструмент для прийняття рішень в реальних системах при зміні зовнішніх умов.

Адаптивний менеджменту ризиків фінансових систем

Ризики фінансових систем є по суті ризиками несистемними та в явному вигляді не визначеними, можуть бути як наслідком дії інших ризиків, так і їх комбінацією. Тому принцип їх оцінювання має дозволяти не лише враховувати зміну рівня та ступеню ризику системи, а й прогнозувати, моделювати та знижувати нові ризики, які можуть виявитись у процесі функціонування системи. Характерні особливості ризиків фінансових систем полягають у тому, що вони вимірюються у грошовому еквіваленті, пов'язані з фінансовими процесами, які можуть змінюватись від стаціонарних до нестаціонарних, від гомоскедастичних до гетероскедастичних, суттєво залежать від фінансових ринків, економічних складових, політичних рішень, та навіть не можуть бути чітко визначені і класифіковані по групах [3].

Принцип адаптивного менеджменту ризиків полягає у можливості оцінювання та адаптації ризиків в процесі функціонування системи за рахунок використання адаптивних методів та моделей, уточнення кращої моделі, її параметрів та структури, застосування адаптивних оцінок прогнозів, використання множини комплексних критеріїв для оцінювання якості структури і параметрів моделей, введення нових критеріїв оцінки якості рішення та з передбаченою можливістю введення нових статистичних даних. Важливою можливістю для адаптації є застосування комбінованих та інтегрованих методів і моделей, які дозволяють подолати існуючі обмеження для використання того чи іншого методу і отримати вищі оцінки якості прогнозів при моделюванні фінансових ризиків.

Концепція адаптації моделі для прогнозування динамічних процесів передбачає ієрархічний підхід до процедур моделювання і прогнозування з урахуванням можливої структурної, параметричної та статистичної невизначеності, адаптації математичних моделей до можливих змін у процесі навчання і при використанні альтернативних методів оцінки параметрів з метою моделювання і покращення прогнозних оцінок. Принцип адаптивного менеджменту ризиків передбачає реалізацію двох контурів адаптації: внутрішнього і зовнішнього, що забезпечує динамічне оцінювання фінансових ризиків в реальному часі через ймовірнісну та вартісну складову. Використання двох контурів адаптації дозволяє одразу частково подолати ризик (оскільки зменшується невизначеність, яка спричиняє появу ризиків, через урахування нових статистичних даних та збурень), а також скорегувати рішення у випадку, якщо воно виявилось недостатньо ефективним. Розроблена схема структурно-параметричної адаптації є універсальною і може бути застосована для аналізу та прогнозування

ризиків різних типів. Апробація запропонованого підходу здійснювалась на поширених фінансових ризиках, зокрема кредитних, а також для прогнозування цін акцій на фінансових ринках. Єдине обмеження, яке накладалося, пов'язане скоріше з класом задач (оцінювання ризиків), які розв'язувались у процесі виконання досліджень. Необхідно було забезпечити використання таких методів, які дозволяють отримати коректні оцінки за факторами, що характеризують ризик: ймовірність і втрати.

Інформаційна система підтримки прийняття рішень на основі методу адаптивного менеджменту ризиків

Було розроблено інформаційну систему, яка передбачає потужний функціонал і застосування значної групи методів для динамічного оцінювання ризиків, розробки інтегрованих моделей та застосування спеціальних засобів для адаптивного менеджменту ризиків.

Розроблена ІСППР має надати особі, що приймає рішення, повний спектр механізмів для моделювання, оцінювання та прогнозування фінансових ризиків та критеріїв для оцінювання якості даних та моделей, з можливістю адаптивного реагування для запобігання і зниження фінансових ризиків з урахуванням нових свідчень, даних та викликів у процесі моделювання [4]. ІСППР генерує множину альтернатив для надання рекомендацій особі, що приймає рішення, має містити множину функцій та процедур для реалізації достатньої повноти рішень. Розширена архітектура багатофункціональної ІСППР представлена в узагальненому вигляді на Рис. 1.

Архітектура ІСППР доволі складна, але може бути представлена у вигляді сукупності послідовно поєднаних модулів та додатків на основі системної методології, зокрема: модуль опрацювання невизначеностей та неповноти даних та знань; модуль розробки скорингової карти як міри ризику; модуль опрацювання невизначеностей та неповноти даних та знань; модуль розробки скорингової карти як міри ризику; модуль динамічного оцінювання фінансових; модуль оцінювання якості даних та рішень і модуль адаптації.

Висновки

Здатність систем до адаптації при появі нових даних, суттєвих змінах зовнішніх умов, врахуванні нових вимог, визначає не лише їх стійкість до появи нових ризиків, а й майбутнє існування таких систем і здатність їх до виживання навіть після прояву системних, економічних, соціальних, екологічних криз, динамічного оновлення та модифікації, а отже й адаптації, до нових умов їх функціонування. Принцип адаптивного менеджменту ризиків, запропонований на основі реалізації адаптивного підходу та методу структурно-параметричної адаптації математичних моделей, передбачає реалізацію двох контурів адаптації: внутрішнього і зовнішнього, що забезпечує динамічне оцінювання фінансових ризиків в реальному часі через ймовірнісну та вартісну складову. Розроблена схема структурно-

параметричної адаптації є універсальною і може бути застосована для аналізу та прогнозування ризиків різних типів. Апробація запропонованого підходу здійснювалась на поширених фінансових ризиках, зокрема кредитних, а також для прогнозування цін акцій на фінансових ринках.

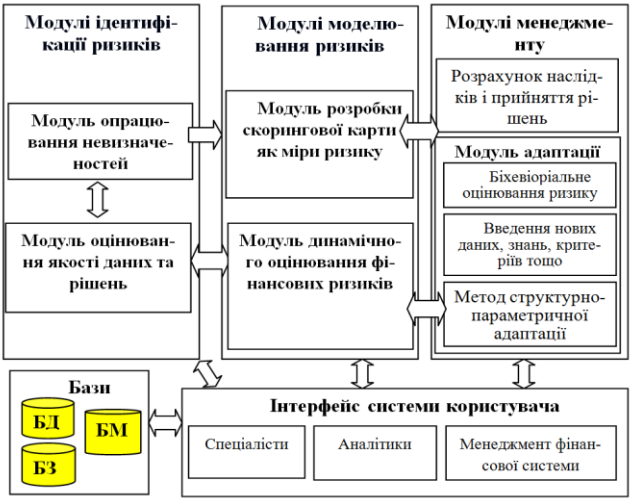


Рис. 1. Розширена структура інформаційної системи підтримки прийняття рішень для задач менеджменту фінансових ризиків.

Література

1. Растригин Л. А. Адаптация сложных систем. Методы и приложения. Рига: Зинатне, 1981. 375 с.
2. Кузнєцова Н. В., Бідюк П. І. Структурно-параметрична адаптація ймовірнісно-статистичних моделей для оцінювання фінансових ризиків KPI Science News. 2018. №3. С. 23–34.
3. ING Homepage. https://www.ing.com/Home.htm.
4. Кузнєцова Н. В., Бідюк П.І. Узагальнення структури інформаційної технології фінансового ризик-менеджменту. KPI Science News. 2019. №3. С. 33–43.

EMPIRICAL STUDY OF NEW METRICS FOR THE INTERNET ROUTE HIJACK RISK ASSESSMENT

Vitalii Y. Zubok
Pukhov Institute for Modelling in Energy Engineering
National Academy of Sciences of Ukraine
15, General Naumov Str. Kiev, 03164, Ukraine
vitaly.zubok@gmail.com

One of the most significant problems of the Internet conbnectivity is deriving from the Border Gateway Protocol (BGP) weaknesses - lack of verification of input routing data. It leads to so known route leaks and route hijacks. None of proposed and partially implemented upgrades and add-on measures described in MANRS can not deliver reliable defense against those types of attacks. That's why it is necessary to learn how to manage risks arising from cyberattacks on global routing.

Assessing the risks of route hijack requires quantitative measurement of the impact of an attack on the routing distortion, and therefore, the breach of information security. In previous papers we used the knowledge of the features of the Internet topology to find the relationship between topology and global routing vulnerability. In this paper we offer new node metrics for representation of both components of information security risk - possible lossess and likelyhood of losses. First metrics we called 'significance' and tied it to importrance of node in routes distribution, with impact of number and weight of announced prefixes. The second metrics we called 'trust' and it reflects likelyhood of hijacking a route on a particular node. At the final, we demonstrate some empirical results of how these metrices can model the effective network topology regarding to relaxating risks of route hijack.

Keywords: The Internet, Global Routing, Route Hijack, Trust Metrics, Cybersecurity.

Introduction

The exterior gateway protocol BGP 4 has been developed to deliver this feature, along with policies and procedures of inter-domain routing. Developed for the network of hundreds nodes which rely on information from each other, after decades BGP-4 is still the same with tens thousands nodes and its crucial lack of routing data integrity. Nowadays there are over 80000 nodes called Autonomous Systems (ASes) interconnected in some way and thus building the telecommunication network – the Internet [1]. Such big figure of transit nodes and even much bigger number of links moves us from the theory of graphs to the theory of complex networks, where the study of the general properties of tolpology is preferred to the study of specific connections between nodes [2], [3]. This is the start point of route forges, route hijacks and other frauds with global impact each [4].

There are proposed proactive mechanisms such as Resource Public Key Infrastructure (RPKI) [5]. It's part of the Internet Routing Registry system. This service provides a collective method to allow one network to filter another networks routes. Method begins with cryptographic signing the route origin. A Route Origin Authorisation (ROA) is a cryptographically signed object that states which AS is authorised to originate a certain prefix. However such techniques are fully effective only in global deployment, and operators are reluctant to deploy them because of the associated technical and financial costs.

In the face of the impossibility of reliable protection against damage associated with an attack, it is necessary to learn how to manage risks arising from cyberattacks on global routing. For this purpose we must use well-studied topological percularities of the Internet to find methods of routing attacks mitigation by aforehead improvement of the connections between Internet nodes.

Analysis Of Recent Research And Publications

Anti-hijack protection consists of two steps: detection and mitigation. RPKI mechanism with route origin validation is not sufficient to fully mitigate AS hijacking. An analysis of the mechanisms of the attack, depending on its objectives and options for its implementation is described in detail in [6]. Detection is mainly provided by third-party services such as BGPMon. They notify the network administrator of suspicious events related to their prefixes based on routing information. They track worldwide routes by tracing and keep track of route announcements in BGP. In the event of an incident, the affected networks begin to mitigate the consequences of the event, for example by announcing more specific prefixes to their networks or by requesting other ASs to filter out false announcements. There are some other studies which offer mechanisms for route attack detection such as ARTEMIS [7] and Peerlock [8].

However, due to the combination of technological and practical deployment issues, existing reactive approaches are largely inadequate, while only few minutes of unauthorized traffic diversion can result in heavy financial losses due to unavailability of service or security breaches. At the same time, the response time to incidents is slow in any case, as current practice requires the need to manually check alerts coming from monitoring systems and third-party services. The duration of widely known incidents ranged from several hours to months [4].

Assuming the information above, we made formal description of the Internet routing [9] and continuing to study the relationship between topology and routing vulnerability to obtain a method for quantifying information risk using a formal global routing model.

Introducing the New Node Metrics

From the explanation of the hijack mechanism in [4],[6],[7], follows that forged route leads to information risk only in two cases: (a) if it changes the route of IP packets to malicious node; (b) if it changes final destination of IP packets. For intruder, it is easier to manipulate the path length if the path is longer. In the long way in the middle there are more nodes through which one can announce a forged

route. Therefore, the probability *P* of interception between nodes *u,v* increases for distant nodes and decreases for close ones [4],[10]: $P(v,u) \sim d(v,u)$.

And also information losses increase with increasing number of affected nodes. $d(v,u)$ affects whether destination node *u* receives false of legitimate route. So does the risk, and we reasonably assume that risk is proportional to distance :

$$R_v \sim \sum_{i=1}^{|V|} d(v,u);\ u \in V\ . \tag{1}$$

The expression (1) is relative quantity of route hijack risk for node *v* regarding target group of network nodes *V*. One cannot predict whether destination node *v* receives false or legitimate route since there no ways to see the BGP processes inside *v* in real time. But one can make personal probability estimate. Let's call it "trust". The matter of trust is probability that node *v* receives and uses (and further propagates) legitimate route originated by *u*. The value of trust *T* is a ratio of average distance between *v* and other nodes, and the distance between *v* and particular *u*:

$$T_u^v = \frac{\sum_{i}^{|V|} d(u,i)}{d(u,v)(|V|-1)}\ ;\ \{i,u,v\} \in V\ ;\ u \neq v\ ;\ u \neq i \tag{2}$$

The risk depends on two components – loss and likelihood and the last one is much similar to probability. So we got a new mertrics for Internet nodes related to route protection.

If we express likelihood via trust, let's express losses using number of nodes impacted by false routes due to route hijack. The more shortest paths $\pi(p_v)$ go through node *v* or prefixes originated by it, the more is impact of this node to routes distribution. This parameter is calculatable by BGP routing tables. Let's call it "significance" S_v^u. Significance should characterize node *v* towards number of IP addresses which might potentially use routes received through *v*. It is impossible to know for sure, we we offer a simplified estimation based on quantity and weight of network prefixes announced via *v*. By "weight" w_π we mean amount of IP addresses covered by network prefix, defined by prefix length $l(\pi)$: $w_\pi = 2^{24-l(\pi)}$. According to this, the weight of the prefix length of 24 bits and covering 256 addresses (which is the least prefix to appear in global routing) is 1, and, for example, weight of 16-bit prefix covering 32768 addresses is 256.

In addition, it should be noted that for each network received forged route via *v*, the *v* also has a certain trust metric. Thus, the degree of influence of the route received from the provider's node will have the greatest impact, because the distance to the provider is the smallest. With taking this into account, we offer to account each prefix π with relaxing coefficient $(1+\delta)^{-1}$ where δ is a metric distance between prefix origin and target node *v*. For example, prefixes originated

directly by *v* will have δ=0, and will be accounted with $(1+\delta)^{-1}$=1. Thus the significance metrics will have this form:

$$S_v^u = \sum_{\pi} w_{\pi}(1+\delta_{\pi})^{-1} = \sum_{\pi} 2^{24-l(\pi)}(1+\delta_{\pi})^{-1} \quad (4)$$

where δ_{π} is metric distance between prefix origin and node *v*.

Two metrics create model risk-oriented node distribution in a 3-dimentional space (*R*,*T*,*S*):

$$R_v^u = T_v^u S_v^u \quad (5)$$

where R_v^u is a risk of hijacking routes originated by *u* on particular node *v*, T_v^u is a trust metrics of *v* evaluated by *u*, S_v^u is significance metrics of *v* evaluated by *u*. And there is an integral risk R^u of route hijacks through a set of Internet nodes *V*:

$$R^u = \sum_{i \neq u}^{|V|-1} R_i^u \quad (6)$$

Empirical Study: Risk Assessment and Mitigation

For experiment we processed real BGP routing tables of several autonomous systems, and got the real node risk distributions. Here we AS8258, AS6939 (Hurricane Electrics), AS15645 (Ukrainian Internet Exchange). For each study we found most significant nodes and measured trust to each of them from the point of AS8258. Then, we calculated the resk of route hijack for each node according to (5) and integral risk according to (6).

Figure 5 represents as *R1* data row the nodes (AS identifiers) ordered by risk at initial topology. The integral risk of this model is $R^u = 1098206$.

Then we modified source routing data pretending that AS8258 has direct links to 3 most risky nodes, recalculated metrics and risk, and received a result on Figure 5 R2 data row .

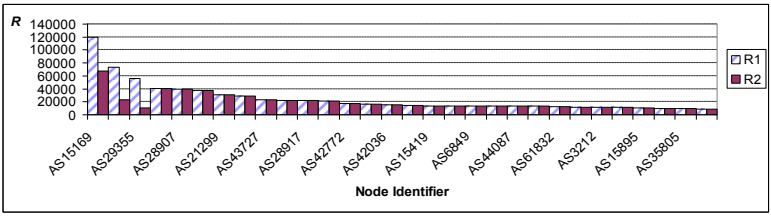


Fig.5. Comparison of risk for resulting topology (R2) to initial topology (R1).

The integral risk of this model is $R^u = 950756$. So, with modeling new topology using trust metrics we have achieved the risk reduced by approximately 15%.

Conclusion

An important step towards assessing the risk posed by attacks on global Internet routing is to predict the impact of the attack, namely to assess the scale of the attack (distribution routes, impact area, number of "damaged" routes). There is the relationship between the topology of the Internet and routing vulnerability.

We formulated and proposed an approach for assessing and mitigationg risk of route hijack using a two new metrics for Internet nodes derived from topology learning – trust and significance. Both metrics together are two components of information security risk related to attacks on global routing. Empirical studies approves the hypotesis that measuring the risk opens the way for developing ways of improvement of AS links topology towards more information security by mitigating the possible risks of attacks on global routing.

References

1. Internet Mapping and Annotation. Center for Applied Internet Data Analysis [Online]. Available: https://www.caida.org/research/topology/ internet_mapping/. Accessed on: June 28, 2020.
2. Newman M. The structure and function of complex networks. *SIAM Review*,2003,Vol.45: 167–256.
3. Faloutsos M., Faloutsos P., and Faloutsos C. On Power Law Relationships of the Internet Topology. *Comput. Commun. Rev*.,1999,№29:251-263.
4. Zubok, V. Retrospective Analysis Cyber Incidents Related to Attacs on Global Routing // *Modelyuvannya ta informaciyni technologii (Modeling and Information Technologies)*: Coll. of Scientific Papers. 2019,№86:41-49. DOI:10.5281/zenodo.3610642.
5. RIPE NCC's Implementation of Resource Public Key Infrastructure (RPKI) Certificate Tree Validation [Online]. Available: https://tools.ietf.org/html/rfc8488. Accessed on: May 25, 2020.
6. Zubok, V.: Metric Approach to Risk Evaluation of Cyberattacks on Global Routing: Selected Papers of the XVIII International Scientific and Practical Conference "Information Technologies and Security" (ITS 2018). Vol-2318 urn:nbn:de:0074-2318-4.
7. P. Sermpezis. ARTEMIS: Neutralizing BGP Hijacking within a Minute / P. Sermpezis, V. Kotronis, et al. // arXiv:1801.01085v4 [cs.NI] 27 Jun 2018.
8. T. McDaniel. Peerlock: Flexsealing BGP / T. McDaniel, J.M. Smith, M. Schuchard // arXiv:2006.06576v3 [cs.NI] 17 Jul 2020.
9. V.Zubok. Building Formal Model of the Internet Routing for Risk Evaluation of Cyberattacks on Global Routing. *CEUR workshop Processing*. 2020, Vol.2577:292-301. [Online]. Avaliable: http://ceur-ws.org/Vol-2577/. Accessed on Aug 12, 2020.

МОДЕЛЮВАННЯ СИСТЕМИ ІНФОРМАЦІЙНОЇ БЕЗПЕКИ ТА АВТОМАТИЗОВАНА ОЦІНКА ІНТЕГРАЛЬНОЇ ЯКОСТІ ВПЛИВУ КОНТРОЛІВ НА ФУНКЦІОНАЛЬНУ СТІЙКІСТЬ ОРГАНІЗАЦІЙНОЇ СИСТЕМИ

Бабенко Тетяна Василівна, Гнатієнко Григорій Миколайович,
Вялкова Віра Іванівна,
Київський національний університет
babenkot@ua.fm, G.Gna5@ukr.net, veravialkova@gmail.com

Запропоновано математичну модель задачі впровадження системи контролів. Розглянуто поняття критичності контролів, а також різні аспекти функціональної стійкості та її співвідношення з надійністю, живучістю, відмовостійкістю. Суттєву увагу приділено урахуванню суб'єктивної складової у задачах визначення якості впровадження контролів та оцінювання інтегрального показника захищеності інформаційної системи. Приділено увагу розгляду зернистості при побудові функції належності нечіткій множині. Розглянуто проблему оцінювання інтегральної якості впровадження контролів та розв'язанню оптимізаційної задачі підвищення рівня якості захищеності інформаційної системи.

Ключові слова: система контролів, інформаційна безпека, функціональна стійкість, прийняття рішень, функція належності нечіткій множині.

Вступ

Побудова ідеальної надійної системи захисту інформації, що обробляється з використанням інформаційно-комунікаційних систем є принципово неможливою задачею. В сучасних умовах заходи і засоби захисту інформації, що використовуються, дозволяють лише суттєво знизити ймовірність негативних наслідків від порушення основних властивостей інформації або збитків від них, але не дають можливості уникнути їх повністю. Тому є сенс розглядати процес забезпечення інформаційної безпеки на деякому прийнятному для організації рівні, що відповідає реально існуючим загрозам.

Постановка задачі

Нехай побудовано систему управління інформаційною безпекою, для якої, відповідно до «кращої практики» [1-3], визначено та впроваджено систему контролів. Множину індексів контролів будемо позначати $i \in I = \{1,...,n\}$.

При цьому кожен контроль характеризується рівнем його впровадження у систему $a_i, i \in I$, та якістю його застосування чи робастності $b_i, i \in I$.

Нехай відомі також, оцінені або визначені експертно, взаємозв'язки між контролями $v_{ij}, i, j \in I$, які характеризують рівень впливу контролю $a_i, i \in I$, на контроль $a_j, j \in I$.

Задача полягає у моделюванні характеристик системи управління інформаційною безпекою (СУІБ), яка створюється, а також арифметизації (метризації, оцифровці) якості контролів та визначенні інтегральної оцінки рівня захищеності інформації. Кінцевою метою такого моделювання є забезпечення функціональної стійкості системи [4]. Для задачі забезпечення інформаційного захисту функціональна стійкість системи полягає у визначенні такої конфігурації контролів та у виборі такого граничного рівня якості контролів, які дозволяють забезпечити допустимий рівень захисту.

Математична модель

Множину контролів та взаємозв'язків між ними будемо моделювати графами або матрицями суміжності чи інцидентності. Зазначимо, що рівень впровадження контролю може характеризуватися деякими дискретними значеннями: бальними оцінками, вербальними виразами, кластеризованими показниками тощо. А якість контролю є функціонально залежною від рівня його впровадження і виражається деякою заданою чи визначеною емпірично функцією – в аналітичному або табличному вираженні $b_i = f(a_i), i \in I$.

На основі аналізу контролів, за допомогою залучення групи експертів, можна побудувати граф взаємозв'язків контролів, який у загальному вигляді є багатодольним. Вершинами графа є контролі з множиною індексів $i \in I$, кожен з яких характеризується рівнем впровадження контролю у систему $a_i, i \in I$, та якістю функціонування $b_i, i \in I$. Взаємозв'язки між контролями є дугами графа $v_{ij}, i, j \in I$. У разі відсутності дуги між деякими вершинами графа, який будується, тобто $\exists: v_{ij} = 0, i, j \in I$, вплив контролю з індексом $i, i \in I$, на контроль з індексом $j, j \in I$, відсутній.

Будемо вважати, що на початковому етапі моделювання та оцінки СУІБ визначено, що рівень впровадження контролів у систему складає $a_i^0, i \in I$, а якість функціонування кожного з них визначена чи виміряна як $b_i^0, i \in I$. Моделювання можливих станів системи полягає у тому, що гіпотетично або практично змінюються початкові рівні впровадження деяких контролів і, відповідно до введених евристик, визначається, яким чином ці зміни вплинуть на якість функціонування взаємопов'язаних з ними контролів та СУІБ в цілому.

При цьому рівень впливу між контролями виражається у зворотному зв'язку: цей зв'язок може бути позитивним або негативним. Позитивний зворотний зв'язок $v_{ij}^+, i, j \in I$, полягає у тому, що у разі досяжності вершини

$i \in I$ графа, навіть при відсутності контролю $a_i = 0, i \in I$, системою забезпечується деякий рівень якості у цій вершині, тобто $b_i > 0, i \in I$. Конкретне числове значення рівня якості контролю у цьому випадку визначається експертним шляхом, експериментально, емпірично чи статистично. Рівень негативного зворотного зв'язку $v_{ij}^{-}, i, j \in I$ при зниженні рівня контролю $a_i^t < a_i^{t-1}, i \in I$, тягне за собою зниження якості контролю не тільки цієї вершини $b_i^t < b_i^{t-1}, i \in I$, але й пов'язаних з нею вершин графа: $b_j^t < b_j^{t-1}, \forall j : v_{ij} > 0, i, j \in I$, де t – такт оцінки якості системи: $t = 0,1,2,...$.

Оцінювання інтегрального рівня контролю

На сьогодні існує група показників, які застосовуються для визначення загального стану захищеності системи. Однією з поширених задач експертного оцінювання є вибір у заздалегідь фіксованому класі відношень деякого результуючого (групового, колективного, компромісного) відношення. При цьому на основі кількох суперечливих показників здійснюється «згортання» (агрегування, інтегрування, узагальнення тощо) показників у деякий єдиний інтегральний показник. Побудувати згортку (узагальнений, агрегуючий, інтегральний, інтегративний критерій якості об`єкта) – це означає доповнити частковий порядок на множині об`єктів до повного. Ця процедура може бути здійснена багатьма способами і з необхідністю включає елемент суб`єктивності.

На першому етапі експерти будують модель ідеальної системи *контролю*, яка відповідає стандарту [5], у вигляді графа з нормативними вершинами та дугами, модель якого описано вище.

На другому етапі експерт або група експертів, які здійснюють аудит реальної системи контролів і встановлюють чи оцінюють наявність контролів, рівень їхнього впровадження у системі і наповнюють змістом граф, який моделює реальну СУІБ. При цьому можуть враховуватися коефіцієнти відносної компетентності експертів [6] тощо. На основі експертно визначених чи обчислених за іншою методикою рівнів контролів $a_i, i \in I$, з урахуванням системи, яка відповідає стандарту [5], визначаються залежні від цієї інформації рівні якості функціонування контролів: $b_i, i \in I$.

На третьому етапі, за участі експертів, здійснюється кластеризація рівнів якості функціонування СУІБ для побудови інтегральної функції належності, яка відображує розподілення якості контролів за рівнями якості і створює функцію належності на основі врахування частотності значень. Інтегральне значення рівня якості впровадження системи контролів, яке свідчить про ступінь функціональної стійкості системи, може бути обчислене, наприклад, за методикою, розробленою авторами [7].

Для визначення інтегральної оцінки будуємо матрицю частот різних рівнів якості виконання функцій $V = \left(v_{ij}\right),\ \ i = 1,...,100,\ \ j < n.$ Кожен рядок цієї матриці відображує оцінений рівень якості виконання функції від 0% до 100%, а стовпчик – кількість функцій з зазначеним рівнем якості виконання.

Для визначення інтегрального рівня якості функціонування складної системи здійснюється класифікацію функцій за рівнем якості та повноти їх виконання. Після цього будується функція належності нечіткій множині значень інтегральної якості впровадження контролів [7, 8].

Інтегральну оцінку якості функціонування системи інформаційної безпеки будемо визначати за допомогою адитивного критерію. При цьому застосовуємо ще низку евристик, які дозволяють обґрунтувати адекватність обчислення єдиного інтегрального значення критерію.

Якість функціонування системи інформаційної безпеки значною мірою залежить від якості функціонування елементів системи. Визначення інтегрального рівня якості функціонування складної слабкоструктурованої системи на основі аналізу взаємозамінності її підсистем та визначення оптимальних варіантів підвищення якості виконання функцій потребує створення відповідної математичної моделі.

Оптимізація інтегральної якості захисту системи

Для підвищення загального (результуючого, інтегрального, агрегованого, інтегративного) рівня якості впровадження системи контролів експерт або група експертів пропонують варіанти поліпшення якості системи шляхом підвищення рівня впровадження деяких контролів та оцінки вартості впровадження вищого рівня окремих контролів. Це пов'язано з обмеженими ресурсами, які організація може виділити на підвищення якості системи управління інформаційною безпекою.

Задача вибору компромісного варіанту забезпечення якісного контролю є багатоаспектною проблемою і може бути формалізована у класі залач багатокритеріальної оптимізації або шляхом застосування ідеї системної оптимізації [9]. Системна оптимізація для задачі побудови моделі інформаційної безпеки полягає у визначення особою, яка приймає рішення, допустимого рівня захисту та у оптимізації лише тих контролів, які є критичними для забезпечення рівня захисту системи в цілому. При цьому слід зважати, що визначення напрямів та вибір варіантів оптимізації інтегрального рівня інформаційної безпеки організаційної системи є багатокритеріальною задачею. Крім забезпечення бажаного рівня впровадження контролів практично кожна організація має зважати, зокрема, на свої фінансові можливості.

У зв'язку з обчислювальною складністю задачі прямого перебору варіантів оптимізації системи контролів, експерти можуть запропонувати близько десяти таких варіантів поліпшення якості.

На основі запропонованих експертами варіантів підвищення рівня впровадження окремих додаткових контролів здійснюється перерахунок

нових станів системи. Тобто розв'язується оптимізаційна двокритерійна задача щодо підвищення інтегральної якості системи захисту та мінімізації вартості поліпшення стану окремих контролів.

Висновки

Запропоновано модель оцінки інтегральної якості системи управління інформаційною безпекою та шляхи цілеспрямованого поліпшення якості її функціонування. Наведена модель може бути адаптована до потреб конкретної організації, а також застосована у інших предметних областях. Модель є відкритою до удосконалення і легко може бути орієнтована на оперування з нечіткими даними.

Джерела

1. Диогенес Ю., Озкайя Э. Д44 Кибербезопасность: стратегии атак и обороны / пер. с анг. Д. А. Беликова. – М.: ДМК Пресс, 2020. – 326 с.: ил.
2. Building a HIPAA-Compliant Cybersecurity Program: Using NIST 800-30 and CSF to Secure Protected Health Information https://doi.org/10.1007/978-1-4842-3060-2
3. ДСТУ ISO/IEC 27002:2015 Інформаційні технології. Методи захисту. Звід практик щодо заходів інформаційної безпеки (ISO/IEC 27002:2013; Cor 1:2014; IDT)
4. Кравченко Ю.В. Сучасний стан та шляхи розвитку теорії функціональної стійкості / Ю. В. Кравченко, С. А. Микусь // Моделювання та інформаційні технології : збірник наукових праць ІПМЕ ім. Г.Є. Пухова. – 2013. – Вип. 68. С. 60-68.
5. ДСТУ ISO/IEC 27001:2015.
6. Hnatiienko H., Snytyuk V. A posteriori determination of expert competence under uncertainty / Selected Papers of the XIX International Scientific and Practical Conference "Information Technologies and Security" (ITS 2019), pp. 82–99 (2019).
7. Hnatiienko H., Vialkova V. Model-Based Analysis Of The Estimation Of Integral Level Of Secuhity Of The Information System // Scientific and Practical Cyber Security Journal (SPCSJ). Vol.2, No.4, December, 2018. Pp. 95-103.
8. Гнатієнко Г.М., Снитюк В.Є. Експертні технології прийняття рішень: Монографія. – К.: ТОВ «Маклаут», 2008. – 444с.
9. Глушков В.М. Основы безбумажной информатики. М.: Наука. Главная редакция физико-математической литературы, 1982. — 552 с.

FEATURES OF DECISION SUPPORT BY EXPERTS OF THE SITUATIONAL CENTER UNDER CONDITIONS OF UNCERTAINTY OF INPUT INFORMATION IN EMERGENCY SITUATIONS

Igor Ruban[1][0000-0002-4738-3286], Vadym Tiutiunyk[2][0000-0001-5394-6367], Olha Tiutiunyk[3][0000-0002-3330-8920]
[1] Kharkiv National University of Radio Electronics, Kharkov, Ukraine
[2] National University of Civil Defence of Ukraine, Kharkov, Ukraine
[3] Simon Kuznets Kharkiv National University of Economics, Kharkov, Ukraine,
tutunik_v@ukr.net

Considering the uncertainty of the parameters affecting the conditions for the normal functioning of the territory of the state, it is proposed to create an effective information and analytical subsystem to manage the processes of prevention and elimination of emergencies when integrated into the existing civil protection system vertically from the object to the state levels, various functional elements of the territorial monitoring system situations and systems of situational centers. In conditions of uncertainty of the input information for experts of the system of situational centers, there is a methodology for substantiating optimal anti-crisis solutions to provide an appropriate level of life safety of the state in emergency situations of various nature.

Keywords: information uncertainty, civil protection system, emergency monitoring system, situational centers system, decision support.

Introduction

The creation of an effective information and analytical subsystem for managing the processes of preventing and eliminating emergency situations is proposed to be comprehensively included in the civil protection system vertically from the object to the state levels of various functional elements of the territorial subsystem for monitoring emergency situations and the components of the subsystem of situational centers, which are rigidly interconnected in informational and executive levels to make appropriate anti-crisis decisions when solving various functional tasks of monitoring, prevention and elimination of emergencies of natural, man-made, social and military nature [1].

One of the urgent directions of creating an information-analytical subsystem to manage the processes of prevention and elimination of emergencies in the civil protection system is the development of a justification methodology, in conditions of uncertainty of the input information for experts of the system of situational centers, optimal anti-crisis solutions to provide an appropriate level of safety of the state in emergency situations, situations of natural, technogenic, social and military nature.

Unresolved issues

An obligatory stage in the functioning of the system of situational centers is decision making. At the same time, not only incorrect, but also ineffective decisions lead to losses or irrational use of financial, time, labor, energy and other resources when managing the processes of prevention and elimination of emergency situations. In this regard, the problem of developing a scientifically grounded methodology to make effective decisions is one of the urgent scientific problems.

In these conditions, it is extremely important to develop formal, normative methods and models for a comprehensive solution to the problem of decision-making in conditions of multi-criteria and uncertainty.

In this direction, main, fundamental results were obtained [2–11], however, the only solution to the problem is far from completion and the continuation of research in this direction is undoubtedly relevant both in theoretical and applied aspects for the development of a justification methodology, in conditions of uncertainty in the input information for experts of the system of situational centers, optimal anti-crisis solutions to provide the necessary level of life of the state in emergency situations of natural, man-made, social and military nature.

Main part

The purpose of this study is to develop the scientific and technical foundations for creating an information-analytical subsystem to manage the processes of preventing and localizing the consequences of emergencies of the civil protection system by developing a methodology for substantiating optimal anti-crisis solutions to provide an appropriate level of safety of the state life in emergency situations of various character, in conditions of uncertainty of input information for experts of the system of situational centers.

The situational center operating in the civil protection system shall, in accordance with the data in Fig. 1, provide: 1) analysis of the information received from the monitoring subsystem; 2) modeling the development of emergency situations on the territory of the city, region, state; 3) development and adoption of managerial decisions to prevent and eliminate emergencies, as well as to minimize their consequences.

The functioning of the scheme shown in fig. 1 in the conditions of completeness of the input information and the presence of one partial criterion for assessing the set of feasible decisions does not present difficulties in substantiating optimal anti-crisis solutions. On the other hand, modern problem situations are characterized by incompleteness of knowledge (uncertainty) of the input data and multiplying particular evaluation criteria.

Thus, the traditional approach, based on the decomposition of the problem into two conditionally independent tasks – in-criterial optimization in deterministic, that is, without considering uncertainty, setting and making a decision under uncertainty for a scalar objective function in modern conditions, does not meet the requirements of practice by accuracy and efficiency.

This is due to the fact that the problem of multicriteria optimization is incorrect, since it allows one to determine the solution only up to the area of compromise solutions, and its regularization for determining a single solution, based on the calculation of a generalized multivariate scalar estimate, is based on poorly structured, subjective expert assessments, the determination of which leads to large errors.

On the other hand, methods of decision-making in conditions of uncertainty on a scalar estimate and the expected effect, without considering its multicriteria, are also inadequate. Therefore, there is a need to develop a methodology for comprehensive solutions to the problem of decision-making, considering the multi-criteria and incomplete uncertainty of the original data.

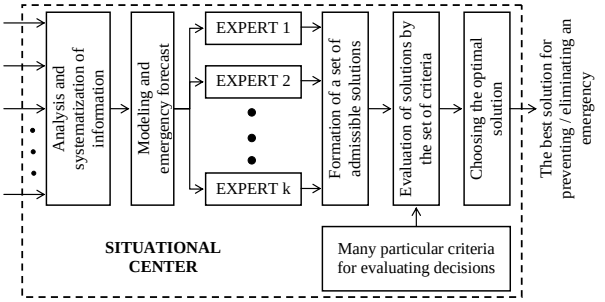


Fig. 1. Functional scheme for substantiating optimal anti-crisis solutions to ensure an appropriate level of life safety of the state in emergencies of a different nature, in conditions of uncertainty of initial information for experts of the system of situational centers of the civil protection system.

In general, the admissible set of solutions contains subsets of consistent X^S and contradictory (compromise) X^C solutions. A feature of the latter is the impossibility of improving any particular criterion $k_j(x)$, $j=\overline{1,n}$ without deteriorating the quality of at least one particular criterion. In this case, by definition, an effective solution x° necessarily belongs to the area of compromise. This means that the problem of multiobjective optimization

$$x^\circ = \arg \underset{x \in X}{extr} < k_j(x) >, \ \forall j = \overline{1,n}, \qquad (1)$$

has no solution, i.e. is incorrect according to Adamar, since in the general case it does not provide the definition of the only optimal solution from the set of compromises X^C.

Thus, the problem of multiobjective optimization arises. The main idea of the methods for solving a multicriteria decision-making problem (MDMP) is to develop a certain regularizing procedure that allows choosing a single solution

from the area of compromises X^C. There are two possible approaches to the implementation of such a task: heuristic, when the decision-maker (DM) makes a choice based on their experience, and formal, based on some formal rules (compromise schemes).

Conclusion

1. It is shown that the basis of the civil protection system shall be a classical control loop, providing: collection, processing and analysis of information; modeling of the development of the situation at the object of management and the development of emergency situations on the territory of the city, region, state; development and two-overthrow of managerial decisions to prevent and eliminate emergencies, as well as to minimize their consequences; implementation of decisions on prevention and elimination of emergency situations, as well as minimization of their consequences.

2. It is proposed to create an effective information and analytical sub-system for managing the processes of prevention and elimination of emergencies by integrated inclusion in the existing civil protection system vertically, from the object to the state levels, of various functional elements of the territorial system for monitoring emergency situations and components of the system of situational centers, rigidly connected among themselves at the information and executive levels for making appropriate anti-crash decisions, for solving various functional tasks of monitoring, preventing and eliminating emergencies of a natural, man-made, social and military nature.

3. It has been determined that the functioning of the civil protection system, and the information and analytical subsystem for managing the processes of handling and liquidating emergencies (which consists of functional elements of the territorial system for monitoring emergencies and the system of situational centers) is visible, takes place in conditions of probabilistic dynamics of the level dangers of vital functions of the country's regions. This dynamics is due to the uncertainty of the parameters affecting the conditions of normal functioning of the territory of Ukraine. In this regard, the problem arises of making optimal anti-crisis decisions in conditions of uncertainty regarding the provision of an appropriate level of safety for the life of the state.

It is shown that the procedure for making managerial decisions is complicated by the fact that the necessary conditions for the effectiveness of decisions are their timeliness, completeness and optimality. Therefore, an increase in the efficiency of decisions made is associated with the need to solve the problem of multi-criteria optimization in conditions of uncertainty, which requires the development of formal, normative methods and models for a comprehensive solution to the problem of decision-making in conditions of multi-criteria and uncertainty when managing the processes of prevention and elimination of emergency situations to provide effective functioning protection system.

4. In order to solve the problem of multicriteria optimization in conditions of uncertainty, in this study, firstly, methods for obtaining input information about the advantages of a decision-maker are formalized, based on both the traditional

heuristic procedures of expert evaluation, and on their formal methods of comparator identification. It is shown that regardless of the method of obtaining the input information and the form of its presentation, the most adequate is the interval assessment of the preferences of the decision-maker. Secondly, it is synthesized a model of a multicriteria scalar assessment of the usefulness of the assumed alternative solutions.

References

1. Tiutiunyk V., Kalugin V., Pysklakova O., Levterov A., Zakharchenko Ju. Development of Civil Defense Systems and Ecological Safety. IEEE Problems of Infocommunications. Science and Technology, pp. 295–299 (2019).
2. Tiutiunyk V., Ruban I., Tiutiunyk O. Cluster analysis of the regions of Ukraine by the number of the arisen emergencies. IEEE Problems of Infocommunications. Science and Technology (2020).
3. Fahimnia B., Tang C.S., Davarzani H., Sarkis J. Quantitative models for managing supply chain risks: A review. European Journal of Operational Research, No.247, pp.1–15 (2015).
4. Haugen S., Vinnem J.E. Perspectives on risk and the unforeseen. Reliability Engineering and System Safety, No.137, pp.1–5 (2015).
5. Khan, F. , Rathnayaka, S. , & Ahmed, S. Methods and models in process safety and risk management: Past, present and future. Process Safety and Environmental Protection, No.98 , pp.116–147 (2015).
6. Aven T. Risk assessment and risk management: Review of recent advances on their foundation. European Journal of Operational Research, Vol.253, No.1, pp.1–13 (2016).
7. Gordon R. Strategic transformational organizational leadership. Encyclopedia of Strategic Leadership and Management, pp.1667-1684 (2017).
8. Gavrylenko S., Babenko O., Ignatova E. Development of the disable software reporting system on the basis of the neural network, Journal of physics, Series 998(2018)012009, pp.1–8. (2018).
9. Marko Jajčinović, Marko Toth. Models of decision making – advantages and drawbacks in crisis management. Ann. Disaster Risk Sci., No 2, pp.129-138 (2018).
10. Kuchuk G., Kovalenko A., Komari I.E., Svyrydov A., Kharchenko V. Improving big data centers energy efficiency: Traffic based model and method. Studies in Systems, Decision and Control. Springer Nature Switzerland AG, Vol.171, pp.161–183 (2019).
11. Jonatan Gehandler, Ulrika Millgård. Principles and Policies for Recycling Decisions and Risk Management. Recycling, Vol.5, No.21, pp.1–18 (2020).

IMPROVING QUESTION ANSWERING SYSTEM FOR UKRAINIAN LANGUAGE

Dmytro Dashenkov[1][0000-0001-9797-1863] and Oleksii Turuta[1][0000-0002-0970-8617]
[1]Kharkiv National University of Radio Electronics, Kharkiv, Ukraine
first.last@nure.ua

In this publication, we explore the possibilities for solving the natural language processing task of automated question answering for multiple languages, including Ukrainian and English. We discuss the differences between Ukrainian and English in terms of the number of datasets available. Also, we discuss the ways of compiling new data sets, including parsing school handbooks and various tests with known correct answers. Finally, we take a look at transfer learning as a method which allows us to reduce the need in the task-specific data by focusing our attention on training models for more general and simpler NLP tasks, and reusing the resulting models for more complicated tasks, such as automated question answering.

Keywords: Natural Language Processing, Question Answering, Data Mining, Ukrainian Language, Transfer Learning.

Introduction

Natural language processing (NLP) is a segment of machine learning which works with human language. NLP works with many tasks. The most common of them are:

- Machine tagging — generating keywords based on a text.
- Text summarization — generating a short summary of a longer text with the purpose of preserving essential meaning of the original.
- Machine translation — translating text in one language into another language.
- Dialog, including question answering — generating responses to questions of a human. In case of a dialog, the task implies preservation of context between several question-answer or command-answer interactions.
- Named entity recognition — classification of entities referenced by a proper noun, such as a person's name, name of a company, title of a book, etc.
- Sentiment analysis — assessment of the emotion of a person who authored a text fragment. For example, used to identify positive and negative reviews of a certain product or service.

Tasks such as machine tagging, text summarization, question answering, and sentiment analysis are grouped into the category of reading comprehension tasks — those which emulate human understanding of a text.

This publication concentrates on reading comprehension tasks, and, in particular, question answering.

Methods for tackling reading comprehension tasks vary in complexity greatly. Tasks such as sentiment analysis may be approached with algorithms as primitive

as naïve Bayes [1]. At the same time, text summarization and question answering require much more sophisticated approaches.

A dominant machine learning model architecture which is used for complex reading comprehension tasks is called a transformer [2]. Transformer architecture uses the encoder-decoder approach, which means that the neural network encodes the input text into a vector of high dimensionality and then produces the output based on the encoded representation of the input.

Noticeable examples of models built on the transformer architecture are BERT (Bidirectional Encoder Representations from Transformers) [3] and its derivative models, such as RoBERTa (Robustly Optimized BERT Pretraining Approach) [4], mBERT (Multilingual BERT) [5], and ALBERT (A Lite BERT) [6]. BERT, along with its optimizations, is designed to be appended with extra output layers which allow the model to solve multiple tasks, such as question answering. The initial task, however, is one of filling masks — substituting gaps in a text with the probable skipped words.

Other noticeable examples of transformers are the models of the GPT (Generative pre-trained transformer) family, GPT-2 [7] and GPT-3 [8]. Unlike BERT, GPT outputs a sequence of words, one by one. Each generated word is added to the initial input, so that the words following it are semantically and grammatically bound to it. GPT models, especially GPT-2 and GPT-3 are designed to work with large numbers of parameters, up to 1.5 billion for GPT-2 17 billion for GPT-3 [7, 8].

Challenges for QA systems in Ukrainian language

Many of the NLP tasks listed in section 1 face similar problems. First and foremost, NLP models are capable of many things, but those things cannot match human understanding. A model is trained to see patterns in text, but it cannot genuinely comprehend the meaning of the text.

This problem edges into the field of philosophy and the definition of done on solving it is not clear.

However, more solvable, practical problems are:

4. Multilingual processing without a form of translation or with translation which does not impair original meaning of the text. Any form of translation, human or machine, often takes some specifics of the underlying text intention. Even more so with machine translation, which, as mentioned, cannot truly understand the input text.
5. Multi-document reasoning. A person may read several documents (snippets of text) in a day. For the person, the thoughts laid out in those documents may intertwine and form a coherent picture of the domain. For a machine, connecting the documents into a single “world view”, is a non-trivial exercise.
6. Data collection. For many languages, such as English, data collection is not complicated. There is plenty of English datasets for NLP, such as SQuAD [9], WikiQA [10], SearchQA [11]. However, when it comes to other languages, especially those with less native speakers, the problem becomes harder. Some datasets do collect data for multiple languages. Many models, such as mBERT

[5], use datasets based on Wikipedia articles. Such datasets often compile a certain number of Wikipedias with the most articles [12].

SOTA for Ukrainian language processing

While many multilingual language models, such as mBERT [5], support Ukrainian, some models may show less performance on Ukrainian compared to English. Standalone models trained for Ukrainian specifically exist. One of them is RoBERTa trained on a Ukrainian dataset [13].

However, dedicated labeled datasets for Ukrainian are harder to come by. One such dataset, or rather a collection of datasets, are compiled by the is the lang-uk researched community [14]. These datasets are targeted for the word embedding and named entity recognition tasks.

Among the recent works, there are a few developments in Ukrainian NLP in general. One of them is the work in modeling Ukrainian language [15], that is, describing its structure formally for processing goals. Another noteworthy paper presents a corpus of Ukrainian text of different genres, annotated by with regard to the region or diaspora of origin [16]. Yet another paper presents a corpus of Ukrainian and other languages targeting multilingual learning [17]. Finally, another paper explores opportunities of fine-tuning BERT [3] on machine-translated data [18].

Goal

At the current time, we are planning on training an ensemble of models specifically targeting the question answering task in multiple languages. The goal is to deliver a model configuration which would overperform existing generic models in multilingual question answering, particularly in Ukrainian and English.

This task requires a two-stage approach:

1. Collecting a question-answer dataset in the target languages.
2. Constructing an ensemble of models and training and fine-tuning them.

English, a one of our target languages, is a sufficient language. There are many datasets designed for the question answering task for English. Some of them are SQuAD [9], WikiQA [10], SearchQA [11], DeepMind Q&A [19], MS Macro QA [20], etc. These datasets have a similar structure. Each entry is a question and an answer or a number of acceptable answers, sometimes accompanied by context — a text snippet which contains the answer to the question. However, Ukrainian, on the other hand, is a much less studies language. There are currently no big public datasets for Ukrainian built for the question answering task. This gap is the reason motivating us to pursue the stated task.

Methods of collecting data

Collecting text data for NLP can be approached by two means. First is manual labor — a group of human assistants who produce the data by creating question-answer pairs from a given text corpus. Such an approach required financing and project management currently unavailable to us.

The second approach is mining datasets from already existing sources. This approach required much less financing and is highly automatable.

For us, sources for creating existing data are the Ukrainian traffic rules and driving exam questions, school handbooks in literature, independent and government tests in Ukrainian and foreign literature, etc. Each of these sources separately provide scarce data, however, combined, they may account for a sufficiently big dataset, which could allow us to fine-tune models trained on more simple tasks in Ukrainian, such as mask filling, for the question answering task — an approach known as transfer learning.

Metrics and benchmarks

A standard metric for NLP tasks is known as the F-score [21]. F-score requires the model output while training to be binary. For this reason, the models output while training can be boiled down to the answer to the question, "Is the provided answer correct for the stated question?". The F-score measures the ratio between the number of true positive answers and the sum of the number of true positive and half of the number of false answers:

$$F_1 = \frac{tp}{tp+\frac{1}{2}(fp+fn)}, \qquad (2)$$

here *tp* — number of true positive answers, *fp* — number of false positive answers, *fn* — number of false negative answers.

We will use F-score for training the model and evaluating its efficiency.

Also, for evaluating the model against other models, we will use a number of benchmarks — fixed tasks and datasets designed specifically for stacking up NLP models one against another. Such benchmarks are SuperGLUE [22] — for general reading comprehension tasks, BLEU [23] — evaluation of multilingual models, and XTREME [24] — reading comprehension tasks on multilingual data.

Results

For the purpose of evaluating existing models for processing Ukrainian, we have constructed a small dataset for the mask filling task. The dataset amounts to around 30 statements taken from the Ukrainian Constitution and other laws. All the statements have a common structure. Each of them is a definition of a term. To evaluate mask filling performance of different models, we've replaced a key noun from the definition, that is the noun which is the closest semantically to the term being defined.

On that dataset, we've run two models, mBERT [5] and RoBERTa trained on Ukrainian [13]. The results showed that, although mBERT did a fine job finding a fitting grammatically noun, RoBERTa overperformed it by the factor of 2, with a recall rate of 27.6% for mBERT, and 55.2% for RoBERTa. Also, an important difference in the output of the models is the tokenization step. RoBERTa is capable of merging neighboring tokens, creating an opportunity to "hint" the model on which the ending of the word should be.

It is important to note that, although underperforming, mBERT was capable of recalling some words which RoBERTa was not. This gives us hope in compiling these models into an ensemble in order to gain yet better performance.

Conclusion

The field of NLP is reach and still largely unexplored when it comes to multilingual models and non-English language processing in general. Although there is some work done for Ukrainian language, it most certainly can be improved. With the automatic question answering task in mind, we are seeking to expand the capabilities of the state of art NLP algorithms in Ukrainian an English alike by both combining already trained models and training new models.

References

1. Aggarwal C.C., Zhai C. A Survey of Text Classification Algorithms. In: Aggarwal C., Zhai C. (eds) Mining Text Data. Springer, Boston, MA (2012)
2. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I.. Attention is all you need. *ArXiv:1706.03762* (2017) *[Cs]*. http://arxiv.org/abs/1706.03762
3. Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. "Bert: Pre-Training of Deep Bidirectional Transformers for Language Understanding." arXiv:1810.04805 [cs] (May 24, 2019). Accessed December 2, 2020. http://arxiv.org/abs/1810.04805.
4. Liu, Yinhan, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. "RoBERTa: A Robustly Optimized BERT Pretraining Approach." arXiv:1907.11692 [cs] (July 26, 2019). Accessed December 2, 2020. http://arxiv.org/abs/1907.11692.
5. Wu, Shijie, and Mark Dredze. "Are All Languages Created Equal in Multilingual BERT?" arXiv:2005.09093 [cs] (September 30, 2020). Accessed December 2, 2020. http://arxiv.org/abs/2005.09093.
6. Lan, Zhenzhong, Mingda Chen, Sebastian Goodman, Kevin Gimpel, Piyush Sharma, and Radu Soricut. "ALBERT: A Lite BERT for Self-Supervised Learning of Language Representations." arXiv:1909.11942 [cs] (February 8, 2020). Accessed December 2, 2020. http://arxiv.org/abs/1909.11942.
7. Radford, A., Jeffrey Wu, R. Child, David Luan, Dario Amodei and Ilya Sutskever. "Language Models are Unsupervised Multitask Learners." (2019).
8. Brown, Tom B., Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, et al. "Language Models Are Few-Shot Learners." *arXiv:2005.14165 [cs]* (July 22, 2020). Accessed December 2, 2020. http://arxiv.org/abs/2005.14165.
9. Rajpurkar, Pranav, Robin Jia, and Percy Liang. "Know What You Don't Know: Unanswerable Questions for SQuAD." arXiv:1806.03822 [cs] (June 11, 2018). Accessed December 2, 2020. http://arxiv.org/abs/1806.03822.

10. Microsoft Research WikiQA Corpus, https://www.microsoft.com/en-us/download/confirmation.aspx?id=52419, last accessed 2020/12/02.
11. Dunn, Matthew, Levent Sagun, Mike Higgins, V. Ugur Guney, Volkan Cirik, and Kyunghyun Cho. "SearchQA: A New Q&A Dataset Augmented with Context from a Search Engine." arXiv:1704.05179 [cs] (June 11, 2017). Accessed December 2, 2020. http://arxiv.org/abs/1704.05179.
12. List of Wikipedias. https://en.wikipedia.org/wiki/List_of_Wikipedias, last accessed 2020/12/01.
13. We Trained the Ukrainian Language Model. https://youscan.io/blog/ukrainian-language-model/, last accessed 2020/12/01.
14. Homepage: lang-uk. https://lang.org.ua/en/models/, last accessed 2020/12/02.
15. Khaburska A., and Tytyk I.: Toward Language Modelling for the Ukrainian Language. In: CEUR Workshop proceedings (2019).
16. Shvedova M.: The General Regionally Annotated Corpus of Ukrainian (GRAC, uacorpus.org): Architecture and Functionality. In: CEUR Workshop proceedings (2020).
17. Grabar N., Hamon T.: Creation of a multilingual aligned corpus with Ukrainian as the target language and its exploration. In: Colins AI (2017).
18. Tiutiunnyk S., Dyomkin V.: Context-Based Question-Answering System for the Ukrainian language. In: CEUR Workshop proceedings (2019).
19. DMQA. https://cs.nyu.edu/~kcho/DMQA/, last accessed 2020/12/02.
20. Bajaj, Payal, Daniel Campos, Nick Craswell, Li Deng, Jianfeng Gao, Xiaodong Liu, Rangan Majumder, et al. "MS MARCO: A Human Generated MAchine Reading COmprehension Dataset." arXiv:1611.09268 [cs] (October 31, 2018). Accessed December 2, 2020. http://arxiv.org/abs/1611.09268.
21. Sokolova, Marina & Japkowicz, Nathalie & Szpakowicz, Stan. Beyond Accuracy, F-Score and ROC: A Family of Discriminant Measures for Performance Evaluation. AI 2006: Advances in Artificial Intelligence, Lecture Notes in Computer Science. Vol. 4304. 1015-1021. (2006)
22. SuperGLUE Benchmark. https://super.gluebenchmark.com/, last accessed 2020/12/01.
23. bleu.dvi. https://www.aclweb.org/anthology/P02-1040.pdf, last accessed 2020/12/01.
24. XTREME. https://sites.research.google/xtreme, last accessed 2020/12/01.

ОНТОЛОГІЧНИЙ ПІДХІД ДО АНАЛІЗУ МЕТАДАНИХ В ДОМЕНІ ІНФОРМАЦІЙНОЇ БЕЗПЕКИ

Гладун Анатолій Ясонович[0000-0002-4133-8169],
Хала Катерина Олександрівна[0000-0002-9477-970X]
Міжнародний науково-навчальний центр інформаційних технологій та систем НАН та МОН Україні, Київ, Україна
glanat@yahoo.com, cecerongreat@ukr.net

Зі зростанням і частим ускладненням загроз кібербезпеки, стає очевидним, що одним із найважливіших ресурсів для боротьби з кібератаками є оброблення великого обсягу даних у кіберсередовищі. Для оброблення величезної кількості даних та для прийняття рішень постає потреба у автоматизації задач пошуку, відбору та інтерпретації Великих Даних для вирішення оперативних задач інформаційної безпеки. Однак традиційні технології аналітики Великих Даних мають обмежені можливості і потребують нового підходу – застосування знань для керування життєвим циклом Великих Даних. Аналітика великих даних доповнена семантичними технологіями, може покращити кіберзахист, та дозволяє обробляти і інтерпретувати великі обсяги інформації в кіберсередовищі. Для аналізу метаданих Великих Даних автори пропонують попередню обробку метаданих на рівні семантики. Детальний опис знань про домен інформаційної безпеки має ієрархічну структуру, яка складається з декількох рівнів. Для побудови тезаурусу задачі запропоновано використати стандарти відкриті інформаційні ресурси, словники, енциклопедії. Розробка ієрархії онтологій формалізує взаємозв'язки між елементами даних, які в майбутньому будуть використані для машинного навчання та алгоритмів штучного інтелекту для адаптації до змін у середовищі, що у свою чергу підвищить ефективність аналітики великих даних для домену кібербезпеки.

Ключові слова: Аналітика великих даних, Інформаційна безпека, Кібербезпека, Онтологія, Тезаурус, Неструктуровані дані, Метадані.

Вступ

Великі Дані сьогодні ефективно використовують для прийняття рішень в системах інформаційного захисту та кібербезпеки. Аналітика Великих Даних дає змогу приймати більш обґрунтовані рішення, забезпечувати регламентоване виконання та рекомендації для удосконалення політик, настанов, процедур, інструментів та інших аспектів мережних процесів. Застосування методів семантичного моделювання в аналітиці Великих Даних необхідна для відбору та поєднання гетерогенних джерел Великих Даних, розпізнавання закономірностей мережних атак та інших кіберзагроз, що має відбуватися швидко для запровадження протидії [1].

Роль метаданих для інтерпретації Великих Даних

Аналітика Великих Даних у інформаційній безпеці потребує для вирішення поставлених задач зовнішніх блоків Великих Даних. Ці дані використовують з метою прогнозування та зупинки кібератак. Попередження атак та розвідка щодо загроз стають важливими для убезпечення інформаційних систем та технологій.

Для того, щоб набір даних можна було вважати великими даними, він повинен володіти однією або декількома такими характеристикам: об'єм; швидкість; різноманіття; достовірність; цінність [2]. Об'єм – це обсяг наборів даних, тобто кількість даних, що генерується; швидкість (швидкість формування та передавання даних) охоплює структуру, поведінку та перестановки наборів даних; різноманіття (тип структурованих та неструктурованих даних) охоплює інструменти та методи, які використовують для оброблення значних або складних наборів даних. Достовірність стосується якості або точності даних, що може стати причиною оброблення даних для усунення помилок і шумів. Цінність визначають, як корисність даних і вона інтуїтивно пов'язана з достовірністю, тому що чим вища точність даних, тим більша їх корисність.

Метадані для Великих Даних є блоками даних, як фізично приєднаними до великих даних, так і знаходитись зовнішньо від Великих Даних. Ці метадані надають інформацію про характеристики та структуру наборів Великих Даних: назва; походження даних, інформація щодо джерела даних; джерело; теги XML із зазначенням автора та дати створення документа; атрибути, що вказують на розмір і форматування, контроль загальної суми; кількість записів набору даних; роздільна здатність зображення; бріф опис даних тощо [3, 4].

Отже, виникає необхідність семантичного аналізу метаданих Великих Даних на основі розроблення методів аналізу природномовних (ПМ) текстів метаданих з використанням онтології Великих Даних, яка формалізує знання та особливості домену і дає змогу для семантичного оброблення при необхідності інших елементів метаданих великих даних.

Онтологічний аналіз для інформаційної безпеки

На сьогодні існує велику кількість онтологій для інформаційної безпеки, що відображають різні окремі аспекти цієї предметної області.

Для відбору та інтерпретації зовнішніх блоків Великих даних використовують онтологію задачі, яку необхідно вирішити у сфері кібербезпеки. Інформаційна система кібербезпеки містить ієрархічну структуру взаємопов'язаних онтологій: онтологія домену, онтологія Великих Даних та онтологія задачі. Для вибору блоків Великих Даних онтологія задачі може бути замінена на тезаурус задачі, який може бути побудований користувачем вручну або створений автоматично.

Для побудови онтології Великих Даних необхідно виділити набір класів та набір екземплярів класів. Також доцільно виділити відношення між

екземплярами атрибутів та їх значеннями. Для опису онтологій великих даних використано таку формальну модель:

$$O_{BD} =< C, R, I, D_t, A > \quad (1)$$

яка містить такі елементи:

$C = C_C \cup C_{In}$ – сукупність концептів онтології, де C_C набір класів, C_{In} набір екземплярів класу;

$R = cr_{er} \cup \{or_i\} \cup or_{ie} \cup \{dr_j\} \cup dr_{er}$ – сукупність відношень між елементи онтології, де cr_{er} – ієрархічні відношення між класами онтології; $\{or_i\}$ – набір властивостей об'єкта, що встановлюють відношення між екземплярами класів; or_{er} – ієрархічні відношення між or_i класів онтології; $\{dr_j\}$ – набір властивостей даних, що встановлюють відношення між екземплярами класів та значеннями з D_t; dr_{er} – ієрархічні відношення між властивостями цих екземплярів класів онтології;

$I = \{I_C \cup I_P\}$ – набір характеристик, які можна використовувати для логічного виведення над онтологією;

D_t – набір типів даних для значень dr_j класів онтології;

A – сукупність правил.

Тезаурус задачі є окремим випадком онтології предметної області, який містить лише онтологічні терміни, але не описує семантику взаємозв'язків між ними з метою аналізу ПМ-текстів. Він може автоматично генеруватися з онтології предметної області та опису задачі на ПМ [6].

В тезаурус задачі для концептів та відношень вводиться вага W, яка вказує на ступінь значимості того чи іншого концепту чи відношення, що підвищує якість оброблення моделі. Формальна модель тезауруса має вигляд:

$$T =< C_t, R_t, Inf > \quad (2)$$

де $C_t \subseteq C$ – кінцевий набір термінів; і $R_t \subseteq R$ – кінцевий набір відношень між цими термінами, Inf – додаткова інформація про терміни (наприклад, вага).

Тезаурус задачі має простішу структуру, оскільки він не охоплює онтологічні відношення і для кожного поняття має додаткову інформацію як ваги $w_i \in W, i = \overline{1,n}$. Тоді формальна модель тезауруса задачі визначається як сукупність упорядкованих пар $T_{task} =< (c_i \in C_t, w_i \in W), \varnothing, Inf >$ з додатковою інформацією в Inf щодо джерела онтології. Алгоритм генерування тезаурусу для ІнРс має такі основні етапи:

1) Формування $Docs = \{doc_i\}, i = \overline{1,n}$, початкового набору $Docs$ із текстових документів doc_i пов'язаних з ІнРс, де кожен документ doc_i з

набору $Docs$ має коефіцієнт ваги W, що дозволяє визначити важливість елементів документа для тезауруса ІнРс.

2) Формування словника ІнРс $D(doc_i)$ для кожного doc_i, який містить усі слова, що зустрічаються в документі. Тоді словник D_{Docs} формується як сума $D(doc_i)$: $D_{Docs} = \bigcup_{i=1}^{n} D(doc_i)$.

3) Генерування тезаурусів ІнРс T_{res}, як проекції набору онтологічних понять C на множину D_{Docs}. $T_{res} \subseteq C$. Цей крок оброблення спрямований на видалення термінів з інших нецікавих для користувача доменів та стоп-слів. Основна проблема на цьому етапі полягає у семантичному відношенні фрагментів ПМ із T_{res} з поняттями множини C онтології домену O_{BD}. Її можна вирішити лінгвістичними методами, що використовують лексичні бази знань для кожної ПМ, що виходить за рамки цієї роботи. Кожне слово з тезаурусу необхідно пов'язувати з одним із онтологічних термінів. У випадку коли відношень бракує, слово розглядають як стоп-слово або елемент позначення, у такому разі його треба відхилити.

Висновки

Перспективи автоматизації створення тезаурусів на основі онтологій залежать від доступності відповідних онтологій доменів та добре структурованих, надійних ІнРс, що характеризують потреби та інтереси користувачів в інформації. Тому ми можемо знайти ІнРс, де такі параметри чітко визначені і можуть бути оброблені без додаткової попередньої обробки. Семантичні Wiki, де взаємозв'язок між поняттями та їх характеристиками визначають за допомогою семантичних властивостей, відповідають таким умовам.

Перелік посилань

1. Erl T., Khattak W., and Buhler P.: Big Data Fundamentals: Concepts, Drivers & Techniques. Prentice Hall, ServiceTech press (2016).
2. P. Buneman, S. Davidson, M. Fernandez, D. Suciu:Adding structure to unstructured data, In 6th International Conference on Database Theory, pp. 336-350. Delphi, Greece (1997).
3. Smith K., Seligman L., Rosenthal A.: Big Metadata: The Need for Principled Metadata Management in Big Data Ecosystems. In Proceedings of the Company DanaC@SIGMOD, p. 46-55. Snowbird, UT, USA (2014).
4. Dey A., Chinchwadkar G., Fekete A., Ramachandran K.: Metadata-as-a-Service. In Proceedings of the 31st IEEE International Conference on Data Engineering Workshops, p.6-9. IEEE, Seoul, South Korea, (2015).

DEFINITION THE LIST OF THREATS OF SECURITY OF WEB SYSTEMS AND APPLICATIONS AT THE DEVELOPMENT STAGE

Denys Saveliev
Pukhov Institute for Modelling in Energy Engineering National Academy of Sciences of Ukraine, Kyiv, Ukraine
desmix07@gmail.com

The report describes the process of forming a model of cybersecurity threats of web systems, the list of threats of which based on the recommendations and experience of reputable organizations in the world and Ukraine. The list of web security threats based on the main recommendations for ensuring system security during design and development, in particular, Open Web Application Security Project (OWASP) recommendations for web-based software developers and recommendations of the Ukrainian State Center for Computer Incident Response CERT-UA on web resource security used for administrators.

Current recommendations for ensuring the security of the web systems and applications development process are given. A list of types of threats that may arise due to non-compliance with the recommendations has been formed. Types of threats are classified based on the stages of system development, as proposed by the non-profit organization The Web Application Security Consortium (WASC) within the WASC Threat Classification project, and is called "Representation of the development phase" - design, implementation and implementation. For each of these types of threats, the possible consequences of their implementation are identified according to the main aspects of information security - violation of confidentiality, integrity and availability of information.

The concept of risk is defined as a combination of the concepts of the probability of occurrence of an undesirable event in the system and the criticality (consequence, impact or severity) of an undesirable event in the event of its occurrence. There are generally accepted levels of criticality of threats, which may vary depending on the specifics of the assessed system, and the levels of probability of threats. The combination of criticality and probability leads to a single value of the risk index, which allows determining the priority of threats to risk management. The risk index is presented in the form of a matrix, where the intersection of levels of criticality and probability of threat realization gives a specific numerical assessment (index) of the threat. The index in turn determines the levels of priority of risks. Obtaining results with a high priority index indicates the need for unambiguous redesign to eliminate or reduce the likelihood of such threats and/or the level of criticality according to the allowable range.

Keywords: Cybersecurity, Threat, Risk, Web, Software, Development, Analysis, Risk index, Software engineering.

The list of threats

Threats to the security of the software system are caused by intentional or unintentional actions of attackers, non-compliance with measures to ensure the security of the system by developers, actions of system users (including administrators and moderators), which can have serious negative consequences.

The detailed description of the process of building a threat model and its main components is given in the article [1] using a list of types of cybersecurity threats, the implementation of which is possible at critical information infrastructure in the field of nuclear energy.

Further formation of the list of security threats to web systems based on the article [2], which contains the main recommendations for ensuring the security of the system during design and development. The basics are the recommendations of the non-profit organization Open Web Application Security Project (OWASP) for web-based software developers who want to maintain an up-to-date level of protection for their products. The recommendations are presented in the form of 10 points that should be considered when creating web applications to avoid potential vulnerabilities in the code [3]. For completeness of the list, the recommendations of the Ukrainian State Computer Response Center CERT-UA on web resource security for administrators are used [4]:

- compliance with current safety requirements;
- use secure frameworks and libraries;
- software update management;
- ensuring secure access to data;
- encryption and data security;
- check input data;
- digital identity protection;
- access control;
- comprehensive information protection;
- monitoring and maintenance of security logs;
- correct handling of errors and exceptions;
- periodically check directories on the web resource server;
- avoid vulnerable web server configurations;
- separation of web applications.

Based on the above recommendations, we can identify a list of types of threats that may arise due to non-compliance, that shows in Table 1.

Implementing web system security threats can have the following consequences:

- violation of information confidentiality;
- violation of the integrity of information;
- violation of information availability.

To classify these threats at the stage of web system development, you can use the stages of the software development life cycle. This classification was proposed by the non-profit organization WASC (The Web Application Security Consortium), which was previously actively involved in the development and promotion of application security standards, in the WASC Threat Classification

project, and is called "Presentation of the development phase". It is offered the distribution of the types of threats by three phases of development - design, implementation and implementation [5].

Table 1 lists the current types of threats, the possible consequences of their implementation and the development phase in which the shortcomings that cause the threat are allowed.

Table 1. List of threats and possible consequences of their implementation

Possible type of threat	Possible consequences of the threat			Development phase		
	Violation of confidentiality	Violation of integrity	Violation of accessibility	Designing	Realization	Implementation
Negative impact on software from development tools	-	+	+	+	+	-
Malicious bookmarks in the code	+	+	+	-	+	-
Abuse of functionality	+	-	-	+	-	-
Use of components with known vulnerabilities	+	+	+	+	+	+
Injections	+	+	+	-	+	-
Cross-site scripting	+	+	+	-	+	-
Execution of commands not provided by the system	+	+	+	+	+	-
Authentication violation	+	+	+	+	+	-
Interception of HTTP cookies	+	-	-	+	+	-
Unauthorized access control	+	+	+	+	+	+
Use of sensitive data	+	+	+	+	+	+
Access server files via URL	+	+	+	-	+	+
Network traffic analysis	+	-	-	+	+	+
Network scanning	+	+	+	-	-	+
Implementing the wrong route / network object	+	+	+	-	-	+
Insufficient journaling and monitoring	+	+	+	-	+	+

Information leakage	+	+	+	+	+	+
Dangerous deserialization	+	+	+	-	+	-
Denial of service	-	+	+	-	+	+
Splitting HTTP requests	+	+	+	-	+	-
External XML objects	+	+	-	-	+	+
Security errors	+	+	+	-	-	+
Indexing web server directories	+	+	+	-	-	+
Impact due to vulnerabilities in other web resources hosted on the same server	+	+	+	-	-	+

Risk levels

Technical standard NASA-GB-8719.13 “NASA Software Safety Guidebook” [6] defines risk as a combination of:

- the probability (quantitative or qualitative) that an undesirable event will occur in the program or system, such as a security breach, security compromise, or system component failure;
- the consequence, impact or severity of the adverse event, if any.

According to this definition, to fully determine the types of risks of a project or program, it is necessary to determine a set of levels of "seriousness" of the threat - its criticality. It is worth using generally accepted organizational definitions, if any, and acceptable. Having a common definition comes in handy when members of a development team assess systemic and programmatic threats, consequences, and management tools.

The combination of criticality and probability leads to a single value of the security risk index, which will determine the priority of threats and manage risks. The most likely and critical threats require clear control in the analysis and development process, while unlikely or non-critical threats may receive little or no attention, except for existing effective design, development and implementation practices used by the project team.

Conclusion

Determining the list of security threats to the developed system is one of the key stages of the risk assessment process, the quality of which in turn ensures the effectiveness of security analysis and risk control.

Current recommendations for ensuring the security of the web systems and applications development process are given. The list of actual types of threats is determined, proceeding from non-fulfillment of the given recommendations and consequences of their realization. For each type of threat, the phase of the project development life cycle is determined, during which the security analysis requires threat control.

References

1. *Saveliev D.:* A model of cybersecurity threats of critical information infrastructure objects in the nuclear power industry. // Modelyuvannya ta informaciyni technologii, vol. 83, pages 42-48.
2. *Saveliev D.:* Web-application security – complex task from design to exploitation. // Modelyuvannya ta informaciyni technologii, vol.88, pages 104-109.
3. *Anton K., Manico J., Bird J.* 10 Critical Security Areas That Software Developers Must Be Aware Of // OWASP Top Ten Proactive Controls Project. 2018. – URL: https://owasp.org/www-pdf-archive/OWASP_Top_10_Proactive_Controls_V3.pdf.
4. CERT-UA recommendations on web resource security: https://cert.gov.ua/recommendations/19, last accessed on 2020/04/20.
5. Thread Classification Development View // The Web Application Security Consortium: http://projects.webappsec.org/w/page/13246969/Threat Classification Development View, last accessed on 2020/04/23.
6. NASA-GB-9719.13: NASA Software Safety Guidebook. NASA Technical Standard. – Washington D.C. National Aeronautics and Space Administration, 2004.

МЕТОДИКА ВИОКРЕМЛЕННЯ КЛЮЧОВИХ СЛІВ І СЛОВОСПОЛУЧЕНЬ ТА ПОБУДОВИ НАПРАВЛЕНИХ ЗВАЖЕНИХ МЕРЕЖ ТЕРМІНІВ ІЗ ЗАСТУВАННЯМ PART-OF-SPEECH TAGGING

Д.В. Ланде [0000-0003-3945-1178] О.О. Дмитренко [0000-0001-8501-5313]
Інститут проблем реєстрації інформації НАН України, Київ, Україна
dwlande@gmail.com, dmytrenko.o@gmail.com

У цій роботі запропонований новий метод виокремлення ключових слів і словосполучень з тематичних інформаційних потоків та новий метод встановлення напрямків зв'язків між вузлами у ненаправлених мережах термінів із застосуванням більш широкої обробки природної мови, що базується на розбитті на частини мови (Part-of-speech tagging). Представлено ідею встановлення вагових значень зв'язків між вузлами у направленій мережі термінів. Також представлена цілісна методика комп'ютерної обробки текстових корпусів та побудови направлених зважених мереж термінів (ключових слів та словосполучень), виокремлених за допомогою попереднього процесу класифікації слів за частинами мови та відповідним маркуванням – Part-of-Speech tagging, та подальшого статистичного зважування. Апробацію запропонованої методики було проведено на прикладі алегоричної повісті-казки "Маленький принц" (англ. "The Little Prince") Антуана де Сент-Екзюпері. Застосовуючи запропонований метод було виокремлено ключові терміни та побудовано направлену зважену мережу зі слів та словосполучень, які відповідають окремим ключовим поняттям у досліджуваному творі.

Ключові слова: текстовий корпус, обробка природньої мови, Part-of-speech (PoS) tagging, термінологічна онтологія, мережа термінів.

Постановка проблеми

Ця стаття присвячена вирішенню актуального науково-практичного завдання, що стосується концептуалізації та подальшої формалізації у вигляді мережі термінів неструктурованих текстових даних, що містяться у тематичних інформаційних потоках розподілених в мережі Інтернет.
Враховуючи той факт, що багато задач, що виникають під час роботи з текстовими інформаційними потоками, лежать на перетині між математичними науками та лінгвістикою, то це відкриває широкі можливості для застосування потужного математичного апарату та лінгвістичної теорії.

Метою цієї роботи є запропонувати новий метод визначення напрямків зв'язків між вузлами ненаправленої мережі, побудованої із ключових слів та словосполучень тематичного текстового масиву, щоб будувати термінологічні онтології у вигляді направлених мереж термінів для того, щоб у подальшому робити конструктивні висновки щодо мережевої структури та її параметрів, та на основі цього приймати ефективні рішення у

відповідно розглянутих проблемних предметних галузях, з якими змістовно пов'язані тексти.

Методика

Побудова направленої мережі термінів здійснюється в межах кожного окремого речення текстового корпусу.

У цій роботі для автоматичного розбиття на токени та розмічування тексту й присвоєння тегів кожному слову застосовуються відповідно окремі функції "word_tokenize" та "pos_tag" спеціалізованої надбудови – модуля NLTK (Natural Language Toolkit), що розроблений на мові програмування Python [1].

Також у цій роботі окрім стандартних наборів стоп-слів, що доступні за посиланнями [2], [3] пропонується використовувати список стоп-слів, що сформований експертами в межах досліджуваної предметної галузі.

Запропонований у цій роботі метод визначення ключових слів та словосполучень, а також напрямків зв'язків базується на використанні результатів отриманих за допомогою процесу класифікації слів за частинами мови та відповідним маркуванням – розмічування частин мови (Part-of-Speech tagging) [4]. Виходячи із практичних досліджень [1] можна помітити, що найбільш вживаними членами речення у англійській мові є артиклі (DT – determiner), іменники (NN – sing or mass noun, NNS – plural noun), займенники (PR – personal pronoun), дієслова (VB – verb base form), означення (JJ – adjectives) та прислівники (RB – adverb). Загалом ключовими словами являються окремі іменники, що зазвичай стосуються людей, місць, речей чи концептів, та іменники у парі з означеннями – словосполучення виду "JJ NN". Також в цій роботі вважається, що важливими можуть бути словосполучення виду "NN_1 NN_2", "JJ_1 JJ_2", "JJ_1 JJ_2 NN", "JJ_1 JJ_2 NN_1 NN_2". Хоча артиклі, прийменники (IN – preposition), сполучники (CC – conjunction, coordinating), окремі дієслова, прислівники та займенники являються стоп-словами, проте словосполучення виду "VV_1 to VV_2" "NN_1 IN/CC NN_2", "JJ_1 IN/CC JJ_2", "JJ NN_1 IN/CC NN_2", "JJ_1 IN/CC JJ_2 NN", "JJ_1 JJ_2 NN_1 IN/CC NN_2", "JJ_1 IN/CC JJ_2 NN_1 IN/CC NN_2" можуть бути ключовими. Після формування вищеназваних термінів та упорядкування їх у певному порядку (формується послідовність, де словосполучення з більшою кількістю слів розташовуються перед словосполученнями та словами, які є їх частиною) здійснюється видалення одиничних стоп-слів (окремих артиклів, прийменників, сполучників, деяких дієслів, прислівників та займенників).

Далі за допомогою глобальної частоти терміна – GTF [5], що визначається відношенням загальної кількості появи терміна у всіх документах корпусу до загальної кількості термінів у документах корпусу, здійснюється статистичне зважування слів та словосполучень, що входять у сформовану на попередньому етапі послідовність.

Для кожного слова у порядку його зустрічання у тексті формується так званий кортеж. Кожен елемент кортежу складається з трьох значень: перше – термін (слово або словосполучення); наступне – тег, який присвоюється

слову в залежності від його приналежності до певної частини мови; останній елемент такого набору – числове значення GTF. Важливо зазначити, що GTF обчислюється з урахуванням двох попередніх значень – слова або словосполучення та частини мови, до якої воно належить. Кількість таких однакових кортежів у всьому тексті, що нормована на загальну кількість сформованих термінів, і визначає значення третього елемента.

На наступному кроці пропонується визначити ненаправлені зв'язки між термінами у тексті. Для досягнення цієї мети застосовується алгоритм графа горизонтальної видимості для часових рядів (Horizontal Visibility Graph algorithm – HVG) [6]. Часовим рядом у нашому випадку є послідовність числових значень GTF, що сформована на попередньому етапі. Ідея алгоритму полягає у тому, що два вузли t_i та t_j, які відповідають елементам часового ряду x_i і x_j, знаходяться у горизонтальній видимості тоді й тільки тоді, коли $x_k < \min(x_i; x_j)$ для всіх t_k таких, що $t_i < t_k < t_j$. У нашому випадку послідовність t_i, $i=1,..,n$ – це послідовність слів у межах речення (n – кількість слів, що залишились у реченні після вищеописаної попередньої обробки). HVG дозволяє будувати мережеві структури на основі текстів, в яких окремим словам або словосполученням деяким чином поставлені у відповідність числові вагові значення.

Якщо між вузлами t_i до t_j часового ряду існує ненаправлений зв'язок, встановлений за вищеописаним алгоритмом, то:

- напрямок зв'язку пропонується встановлювати від вузла t_i до t_j, якщо у реченні слово (не словосполучення), якому відповідає вузол t_i зустрічається раніше ніж термін (слово або словосполучення), якому відповідає вузол t_j;
- напрямок зв'язку пропонується встановлювати від вузла t_j до t_i, якщо у реченні словосполучення (не слово), якому відповідає вузол t_j зустрічається раніше ніж термін, якому відповідає вузол t_i.

Беручи до уваги принцип формування послідовності із термінів, що описаний вище, та запропоновані правила встановлення зв'язків можна помітити, що слова та словосполучення будуть входити у відповідні словосполучення, мають більшу кількість слів. Тобто значна частина словосполучень з більшою кількістю слів є розширенням відповідних їм словосполучень та слів. Подібний принцип побудови направлених мереж зі слів, побудова мереж природніх ієрархій термінів, запропонований у роботі [7], де направлена мережа зі слів та словосполучень будується за принципом входження терміна у відповідне йому словосполучення.

Вагові значення зв'язків між вузлами направленої мережі визначаються за запропонованим у роботі [8] принципом, який полягає у тому, що вузли, які відповідають однаковим термінам побудованої на попередньому етапі направленої мережі, об'єднуються ("склеюються"), а кількість однаково-направлених зв'язків між відповідними вузлами і визначає вагове значення зв'язку між цими вузлами.

Результати досліджень

Запропонована методика обробки текстових корпусів та побудови направлених зважених мереж термінів була апробована на прикладі найвідомішого твору Антуана де Сент-Екзюпері алегоричної повісті-казки “Маленький принц” (англ. “The Little Prince”).

Відповідно до методики, що запропонована вище, було здійснено обробку обраного текстового документу й виокремлено ключові терміни та побудовано направлену зважену мережу зі слів та словосполучень, які відповідають окремим ключовим поняттям у досліджуваному творі (рис. 1). Для побудованої мережі було видалено всі зв’язки, які мають вагове значення рівне 1 та вузли, вихідна та вхідна степінь яких дорівнює нулю.

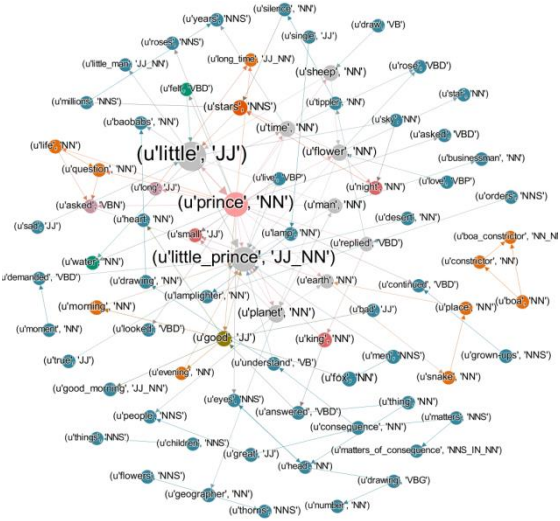


Рис. 10. Направлена зважена мережа термінів побудована для тексту “The Little Prince” (мітки вузлів містять термін та відповідний йому тег).

Висновки

У цій роботі було запропоновано новий метод виокремлення ключових термінів та новий метод встановлення напрямків зв’язків із застосуванням більш широкої обробки природної мови, що базується на розбитті на частини мови (Part-of-speech tagging). Також представлена цілісна методика, що дозволяє будувати направлені зважені мережі з ключових слів та словосполучень текстового корпусу.

Апробацію запропонованої методики було проведено на прикладі алегоричної повісті-казки “Маленький принц” (англ. “The Little Prince”) Антуана де Сент-Екзюпері. Проаналізувавши результати дослідження було виявлено найбільш вагомі зв’язки між відповідними вузлами у мережі

термінів, що відповідають окремим ключовим поняттям у досліджуваному творі. У межах запропонованої онтологічної моделі ключовими виявились терміни “little”, “prince” і “little_prince”, що відповідають назві твору, а найбільш вагомими, як і очікувалось, зв’язки між цим ж термінами “little → little_prince” та “little → prince”.

Літературні джерела

1. Steven Bird, Ewan Klein, Edward Loper. Natural Language Processing with Python. O'Reilly Media (2009). ISBN 0-596-51649-5
2. Google Code Archive: Stop-words. https://code.google.com/archive/p/stop-words/downloads/. Accessed 24 Oct 2020
3. Text Fixer: Common English Words List. http://www.textfixer.com/ tutorials/commonenglishwords.php. Accessed 24 Oct 2020
4. Extract Custom Keywords using NLTK POS tagger in python. https://thinkinfi.com/extract-custom-keywords-using-nltk-pos-tagger-in-python/. Accessed 24 Oct 2020
5. Lande, D., Dmytrenko, O., Radziievska, O.: Determining the Directions of Links in Undirected Networks of Terms. In: CEUR Workshop Proceedings (ceur-ws.org). Vol-2577 urn:nbn:de:0074-2318-4. Selected Papers of the XIX International Scientific and Practical Conference "Information Technologies and Security" (ITS 2019), vol. 2577, 132-145. (2019). ISSN 1613-0073 [http://ceur-ws.org/Vol-2577/paper11.pdf]
6. Luque, B., Lacasa, L., Ballesteros, F., & Luque, J.: Horizontal visibility graphs: Exact results for random time series. Physical Review E, 80(4), (2009). DOI: doi.org/10.1103/PhysRevE.80.046103.
7. Lande, D. V., Snarskii, A. A., Yagunova, E. V., & Pronoza, E. V.: The use of horizontal visibility graphs to identify the words that define the informational structure of a text. In: 2014 12th Mexican International Conference on Artificial Intelligence, pp. 209-215 (2014).
8. Lande D.V., Dmytrenko O.O.: Creating the Directed Weighted Network of Terms Based on Analysis of Text Corpora. 2020 IEEE 2nd International Conference on System Analysis & Intelligent Computing (SAIC) (Kyiv, 5-9 Oct. 2020). DOI: doi.org/10.1109/SAIC51296.2020.9239182.

ІНТЕГРОВАНА ТЕХНОЛОГІЯ ТА FRAMEWORK ДЛЯ МОДЕЛЮВАННЯ СЦЕНАРІЇВ АНАЛІТИЧНОЇ ДІЯЛЬНОСТІ

В. Сенченко
Інститут проблем реєстрації інформації НАН України,
svrukr@gmail.com

Keywords: Аналітична діяльність, framework, Data Mining

Постановка проблеми

Аналітична діяльність (АнД) це сукупність дій на основі концепцій, методів, засобів, нормативно-методичних матеріалів для збору, накопичення, обробки та аналізу даних з метою обґрунтування та прийняття рішень, або генерації нових знань [1,2]. В сучасних системах організаційного управління, мета управління в яких, в більшої степені, досягається через професійну діяльність персоналу, фактор АнД в управлінні та прийнятті рішень суттєво підвищується. Це пояснюється тим, що об'єктом АнД досить часто стають складні взаємопов'язані процеси прийняття рішень, кожен з яких може характеризуватися багатомірністю параметрів, необхідністю обробки великих обсягів даних для пошуку скритих закономірностей та знань. Особливістю задач, що виникають при розв'язанні таких проблем, є відсутність чітких алгоритмів (сценаріїв) при надмірної складності технології та методів обробки даних (очищення, кластерізації, находження асоціації, Data Mining, Text Mining, Web Mining, Deep Learning тощо), які, зазвичай, незрозумілі та не сприймаються більшістю аналітиків.

В контексті зазначених проблем, під сценарієм АнД розуміється певна послідовність дій, яку має виконувати аналітик або група аналітиків при вирішенні завдань з метою підготовки та прийняття збалансованих управлінських рішень [6]. Безумовно, формування сценаріїв АнД, які залучають складні технології обробки даних та зрозуміли для широкого кола фахівців, є дуже актуальною проблемою.

Мета роботи

Метою роботи є подальший розвиток теоретичних основ, методів аналізу та інструментальних засобів моделювання (Framework для інтеграції різнорідних інструментів моделювання аналітичної діяльності) та перевірки ефективності сценаріїв АнД, а також їх практичного застосування scripts при створенні аналітичних додатків в системах організаційного управління.

Аналіз сучасних методології формалізованого опису та побудови сценаріїв АнД на базі комп'ютерних технологій – Схемно-рекурсивний метод [4], Матричний метод опису та формування сценарію [5], Метод «сценарних областей» [7], включаючи й методи сценарного аналізу каскадних події [6] та інших, дозволяє зробити такий загальний висновок, жоден з методів не охоплю вісь процес моделювання зверху до низу. Для формалізації багаторівневих сценаріїв сучасних систем організаційного управління

найбільш ефективними є гібридні методології, які спираються на графічні методи моделювання сценаріїв у сполученні з семантичною їх інтерпретацією на базі онтологічної технології. Запропонований підхід дозволяє представляти та моделювати АнД «зверху до низу» починаючи з моделі інформаційної взаємодії різних за фахом і повноваженнями експертів та закінчуючи детальної проробкою сценаріїв обробки даних та отримання результатів аналізу для різних сфер діяльності.

Виходячи з запропонованого підходу проведено огляд існуючих мов високого рівня та інструментальних програмних засобів формального опису моделей сценаріїв сучасної аналітики, включаючи й графове представлення, а також визначення їх функціональних переваг, сценарних характеристик та недоліків для різних структур організаційного управління, у тому числі й для багаторівневих структур. За результатами аналізу запропоновано створення Web- portal для моделювання сценаріїв аналітичної діяльності в системах організаційного управління. Framework для інтеграції різнорідних інструментів моделювання аналітичної діяльності складається з наступних програмних платформ та мов високого рівня (рис. 1.):

1) *Засоби формування, управління контентом та правами доступу до ресурсів Web-portal Simulation of analytics scenarios (комп'ютерного моделювання сценаріїв АнД).*
2) *Графічний редактор для моделювання сценаріїв інформаційної взаємодії верхнього рівня – BizAgi Process Modeler Version 3.6. (open-source Modeler) [3], який підтримує стандарт ISO/IEC 19510 (November 2013) – Business Process є Model and Notation, Information technology. Окрім візуального відображення, сценарії може бути експортовано в базу даних у форматі XML або XPDL та конвертовано у OWL-модель онтології для подальшого аналізу.*
3) *Студія та графічний редактор для моделювання сценаріїв та процедур обробки широкого спектру даних аналітики (структурованих, слабоструктурованих, неструктурованих включаючи методи DataMining, TextMining та WebMinin). Це open-source RapidMiner Studio та Visual Workflow Designer for Data Scientists програмна платформа Technical University of Dortmund – більш 500 віджетів [9]. Результатом проектування в RapidMiner є виконувана модель сценарію обробки та аналізу даних та scripts моделі сценарію в форматі XML-коду.*
4) *Графічний редактор та фреймворк для моделювання онтологій предметної області – Protégé 5 – free, open-source ontology editor and framework for building intelligent systems [8].*
5) *Машина логічного виводу на прецедентах для сфер діяльності, заснованих на накопиченому досвіді – jCOLIBRI 1.0 in a nutshell. A software tool for designing CBR systems.*
6) *Мова програмування для опису та виводу знань при моделюванні сценаріїв пошуку знань та асоціацій в Web-просторі – OWL2 Web*

Ontology Language Document Overview (Second Edition) W3C Recommendation 11 December 2012.

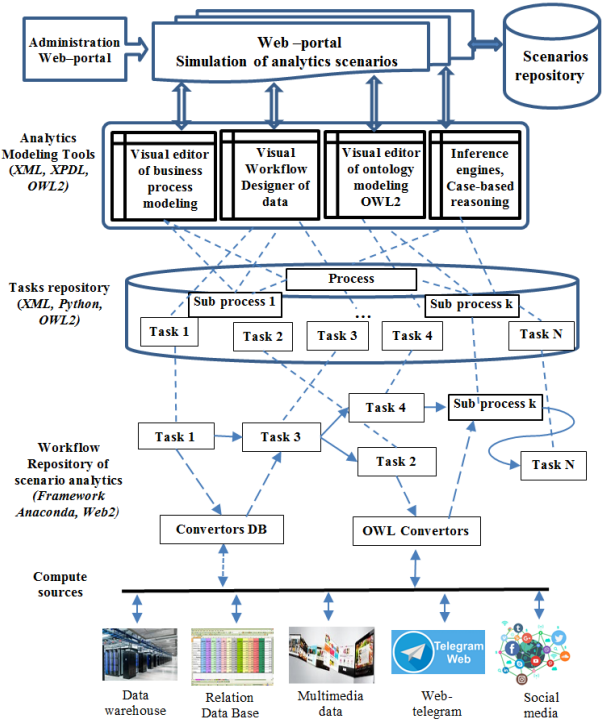


Рис. 1. Framework для інтеграції різнорідних інструментів моделювання аналітичної діяльності

7) *Об'єктно-орієнтована мова високого рівня Python 3.6.4. для безпосереднього формування інтерфейсів кінцевого користувача та роботи з бібліотеками NumPy для роботи з багатовимірними масивами. Бібліотека WSGI — інтерфейс шлюзу з веб-сервером (Python Web Server Gateway Interface), пакети для доступу до різних СУБД: PostgreSQL, Oracle, Sybase, Firebird (Interbase), Informix, Microsoft SQL Server, MySQL та sqlite.*
8) *Фреймворк та система управління середовищем реалізації моделей сценаріїв (Workflow Repository of scenario analytics) Anaconda (відкрита платформа) [10]. Anaconda підтримує мови програмування Python, R та бібліотеки таких як NumPy, SciPy, Astropy , які містять більш 1500 програмних модулів*

для широкомасштабній обробки даних, аналітики, застосування методів машинного навчання. різних програмних продуктів.

Запропонований підхід дає низку переваг – починаючі з суттєвого спрощення самого процесу моделювання та верифікації сценаріїв АнД у парадигмі «зверху до низу», і, закінчуючи, реалізацією моделі в інтегрованому середовищі (Workflow Repository of scenario analytics) з отриманням кодів (scripts) виконуваних задач, функції та операції, а також формуванням інтерфейсу кінцевого користувача.

Література

1. Додонов А.Г., Сенченко В.Р., Коваль А.В., Аналитика и знания в компьютерных системах. Монография.. Институт проблем регистрации информации НАН Украины. Национальный технический университет Украины «Киевский политехнический институт имени Игоря Сикорского», Киев, 2020, 315 С.
2. Курносов Ю.В., 2016 Философия аналитики «Русская аналитическая школа», 2016, режим доступу URL: https://www.twirpx.com/file/ 2257341/
3. Офіційний сайт: Bizagi Modeler, URL: https://www.bizagi.com/en/
4. Прийняття рішень методом аналітичної ієрархі: електронний ресурс URL: https://studopedia.su/6_41843_priynyattya-rishen-metodom-analitichnoi-iierarhii.html
5. Теоретичні основи методу аналітичної ієрархії: електронний ресурс URL https://bibl.com.ua/informatika/4441/index.html?page=4
6. Додонов О.Г., Сенченко В.Р., Коваль О.В., Бойченко А.В.: Моделювання сценаріїв аналітичної діяльності на основі нотації BPMN та OWL Реєстрація, зберігання і оброб. даних. 2020. Т. 22. № 1. С. 31 – 48.
7. Касьянов В.С.: Системный анализ в менеджменте: електронний ресурс URL: https://bstudy.net/750480/ekonomika/ morfologicheskaya_matrits
8. Protégé - Free, open-source ontology editor and framework for building intelligent systems: URL: http://protege.stanford.edu
9. RapidMiner Studio: Visual Workflow Designer for Data Scientists: URL: https://rapidminer.com/products/studio/
10. The World's Most Popular Data Science Platform Anaconda: URL: https://www.anaconda.com/

МОДЕЛЬ ТА МЕТОД РОЗПІЗНАВАННЯ КІБЕРІНЦИДЕНТІВ SIEM-СИСТЕМОЮ НА ОСНОВІ НЕЧІТКОГО ЛОГІЧНОГО ВИВОДУ

Ігор Субач, Олександр Кубрак, Артем Микитюк, Станіслав Коротаєв
Інститут спеціального зв'язку та захисту інформації Національного технічного університету України "Київський політехнічний інститут імені Ігоря Сікорського", Україна
igor_subach@ukr.net

Анотація. Запропоновано модель та метод розпізнавання кіберінцидентів в умовах неповноти на неточності інформації про них на основі застосування теорії нечітких множин та нечіткого логічного виводу. Наведено формальну постановку задачі виявлення кіберінцидентів SIEM-системою та сформульовано основні труднощі, які виникають під час її вирішення. Запропоноване рішення задачі виявлення кіберінцидентів SIEM-системою, як задачі їхньої ідентифікації, шляхом знаходження відображення між множиною ознак кіберінцидентів та множиною їх можливих класів.
Ключові слова: кібербезпека, кіберзахист, кібератака, кіберінцидент, SIEM, теорія нечітких множин, кортежна модель розпізнавання, правило-орієнтований метод

Вступ

Основою побудови ефективної системи кіберзахисту має бути застосування проактивної SIEM-системи – системи управління інформацією та подіями безпеки [1]. Архітектура та функціональна модель проактивної SIEM-системи розглядалася в роботі [1]. Відповідно до задач, які виконує дана система, основу її функціонування складають механізми: нормалізації, фільтрації, класифікації, агрегації, кореляції, пріоритезації та аналізу подій і кіберінцидентів, а також генерація різноманітних звітів, повідомлень та візуального представлення даних для оперативного та обґрунтованого прийняття рішень [1]. Методику раціонального вибору SIEM для побудови SOC (Security Operation Center) наведено у [2].

У деяких джерелах [3] дані механізми розглядаються, як етапи загального процесу, який має назву – процес кореляції. Йому відводиться особливе місце в роботі SIEM-системи, оскільки його призначенням є виявлення кібератак, шкідливої активності, порушень політики безпеки та інші. Це забезпечується завдяки вирішенню широкого спектру задач, які він охоплює: визначення потенційних взаємозв'язків між різнорідною інформацією безпеки; групування низькорівневих подій до подій більш високого рівня; виявлення потенційних інцидентів на основі аналізу поведінки різних об'єктів інфраструктури та інші.

Технологічно, у складі SIEM-системи, метод кореляції включає послідовність дій над даними, яка направлена на виявлення певним способом ознак видалення, об'єднання та зв'язування інформації, що

обробляється, а також встановлення її причинності та пріоритетності [3]. Дані ознаки називають кореляційними ознаками. Для цього, на різних етапах процесу кореляції застосовуються велике різноманіття методів. Найбільш розповсюдженим є правило-орієнтований метод, проте в наслідок того, що він ґрунтується на класичних продукційних правилах, які не завжди в умовах неповноти та неточності інформації про кіберінциденти дають очікуваний результат, застосування його не завжди є ефективним.

Метою роботи є розробка моделі та правило-орієнтованого методу виявлення кіберінцидентів SIEM-системою на основі нечіткого логічного виводу.

Постановка задачі виявлення кіберінцидентів

Будь-який кіберінцидент характеризується множиною інформаційних ознак, на основі яких, у свою чергу, він може бути розпізнаним.

Нехай $O=\{o_i\} \mid i=\overline{1,n}$ – множина інформаційних ознак кіберінцидентів, які відбуваються в системі та представляються за допомогою множини $C=\left\{ C_j \mid C_j=\left(o_{j1}, o_{j2},\ldots, o_{jm}\right)\right\}$, $j=\overline{1,J}$, де $o_{ji}\in O$ – інформаційні ознаки, що асоціюються з кіберінцидентом C_j.

Тоді модель розпізнавання кіберінцидентів можна представити за допомогою кортежу [4]:

$$M=\langle K, O_i, R, C\rangle, \qquad (1)$$

де K – класифікатор ознак; $o_i\in O$ – множина ознак, що спостерігаються; $R=\{R_i\}$ – множина правил розпізнавання кіберінцидентів; C – кіберінцидент. (1)

Процес розпізнавання кіберінцидентів здійснюється на основі правил, зазвичай продукційних: $R_1:(K,O_i), R_2:(K,O_i),\ldots, R_l:(K,O_i)\to C$.

Проте, продукційні правила не у повній мірі відповідають умовам неповноти та неточності інформації про кіберінциденти. Для цього, застосовуються методи та моделі теорії нечітких множин на нечіткого логічного виводу [5].

Метод рішення задачі

З урахуванням наведеного, модель (1) може бути розвинутою та представленою у наступному видгляді:

$$MF=\langle KF, O_i, RF, C\rangle, \qquad (2)$$

де KF – нечіткий класифікатор, $RF=\{RF_i\}$ – множина нечітких правил розпізнавання кіберінцидентів:

$$RF_1:(K,O_v), RF_2:(K,O_v),\ldots, RF_l:(K,O_v)\to C.$$

З іншого боку, задача розпізнавання кіберінцидентів може розглядатися як задача їхньої ідентифікації [6], рішення якої полягає у знаходженні відображення:

$$O^* = \left(o_1^*, o_2^*, \ldots, o_n^*\right) \to c_j \in C = \left(c_1, c_2, \ldots, c_m\right), \qquad (3)$$

де O^* – множина ознак кіберінцидента; C – множина можливих кіберінцидентів. Область зміни ознак кіберінцидентів і вихідного значення результату ідентифікації вважаються відомими.

Основні труднощі, рішення задачі (3), полягають у наступному:

- для вірної ідентифікації кіберінциденту необхідно враховувати велику кількість різнорідних параметрів системи (кількісних та якісних), що, потребує високої кваліфікації офіцера з кібербезпеки, а також відповідного часу;
- відсутність аналітичних залежностей між кіберінцидентом та його ознаками.

Наведене підтверджує доцільність розробки моделі ідентифікації кіберінцидентів на основі нечітких правил та нечіткого логічного виводу.

При цьому, необхідно враховувати лінгвістичний характер типу кіберінциденту (вихідна змінна) та його ознак (вхідні змінні). У свою чергу, причинно-наслідкові зв'язки між кіберінцидентом та його ознаками повинні бути описані експертом зрозумілою мовою, а після цього, формалізовані у вигляді множини нечітких логічних правил.

При великій кількості ознак кіберінцидентів, доцільно побудувати дерево логічного виводу, яке визначає порядок вкладення висловлювань одне в одне.

Для кореляційного правила: якщо на одному комп'ютері з тією ж IP-адресою, 7 спроб автентифікації користувача завершилося невдало на протязі 10 хвилин, при цьому використовувалися різні ідентифікатори користувача, а також було здійснено успішний вхід користувача до системи з будь-якого комп'ютеру з тією ж IP-адресою, то ця подія має бути розглянутою офіцером з безпеки [6], структура дерева логічного виводу відповідає відношенням: $c = \xi_c(\alpha, \beta)$, $\alpha = \xi_\alpha(o_1, o_2, o_3, o_4, o_5)$, $\beta = \xi_\beta(o_3, o_4, o_5, o_6)$.

Тут через $o_1 \div o_6$ позначено ознаки кіберінциденту: o_1 - кількість невдалих спроб входу до системи; o_2 - кількість ідентифікаторів користувачів; o_3 - тривалість за часом спроб входу до системи; o_4 - кількість IP-адрес, що задіяні під час входу до системи; o_5 - кількість комп'ютерів, що задіяні під час входу до системи; o_6 - кількість вдалих спроб входу до системи. Через c_1 та c_2 позначено тип події, що відбувається в системі: α - успішний вхід до системи; β - неуспішний вхід до системи; c_1 – кіберінцидент; c_2- нормальний стан системи.

Для оцінки значень лінгвістичних змінних $o_1 \div o_6, \alpha, \beta$ застосовується єдина шкала якісних термів: Н - низький; нС - нижче за середній; С - середній; вС - вище за середній; В - високий.

Задача ідентифікації кіберінцидента відповідає математичній моделі ідентифікації об'єкту з дискретним виходом [5]. Для ідентифікації кіберінцидента c_1, співвідношення має наступний вигляд:

$$\mu^{c_1}(c) = \left[\mu^{B}(\alpha) \wedge \mu^{B}(\beta) \right] \vee \left[\mu^{вC}(\alpha) \wedge \mu^{вC}(\beta) \right],$$

де $\mu^{B}(\alpha) = \left[\mu^{\text{В}}(o_1) \wedge \mu^{\text{В}}(o_2) \wedge \mu^{\text{Н}}(o_3) \wedge \mu^{\text{Н}}(o_4) \wedge \mu^{\text{вС}}(o_5) \right]$, $\mu^{B}(\beta) = \left[\mu^{\text{нС}}(o_3) \wedge \right.$

$\left. \wedge \mu^{\text{Н}}(o_4) \wedge \mu^{\text{В}}(o_5) \wedge \mu^{\text{Н}}(o_6) \right]$, $\mu^{вC}(\alpha) = \left[\mu^{\text{вС}}(o_1) \wedge \mu^{\text{вС}}(o_2) \wedge \mu^{\text{нС}}(o_3) \wedge \mu^{\text{Н}}(o_4) \wedge \right.$

$\left. \wedge \mu^{\text{С}}(o_5) \right]$, $\mu^{вC}(\beta) = \left[\mu^{\text{нС}}(o_3) \wedge \mu^{\text{Н}}(o_4) \wedge \mu^{\text{вС}}(o_5) \wedge \mu^{\text{Н}}(o_6) \right]$;

а $\mu(c), \mu(\alpha), \mu(\beta), \mu(o_i)$ – відповідні функції належності.

Висновки

Для підвищення ефективності застосування правило-орієнтованого методу розпізнавання кіберінцидентів в умовах неповноти та неточності інформації, запропоновано модель, що ґрунтується на теорії нечітких множин та нечіткого логічного виводу. На основі моделі розроблено правило-орієнтований метод ідентифікації кіберінцидентів, шляхом відображення множини їх ознак на множину можливих класів кіберінцидентів та алгоритм його реалізації.

Отримані результати можуть бути використанні під час вирішення задачі виявлення кіберінцидентів SIEM-системою, яка входить до комплексу SOC.

Джерела

1 І. Субач, В. Кубрак, та А. Микитюк, "Архітектура та функціональна модель перспективної проактивної інтелектуальної системи SIEM-системи для кберзахисту об'єктів критичної інфраструктури", Information Technology and Security,Vol 7., Iss. 2., 2019, pp. 208–215, Режим доступу: https://doi.org/10.20535/2411-1031.2019.7.2.190570

2 I. Subach, V. Kubrak, and A. Mykytiuk, "Methodology of rational choice of security incident management system for building operational security center", CEUR Workshop Proceedings, 2019, 2577, p.p. 11-20, Режим доступу: http://ceur-ws.org/Vol-2577/paper2.pdf

3 А.Федорченко, Д. Левшун, А. Чечулин, и И. Котенко, "Анализ методов корреляции событий безопасности в SIEM-системах. Часть 1", Труды СПИИРАН, вып.4 (47), 2016, с. 5-27, DOI: 10.15622/sp.47.1.

4 Ю. Самохвалов, та С. Толюпа. “Кореляция событий в SIEM-системах на основе немонотонного вивода”, Захист інформації,.Том 19, № 1, 2017, с.5-9.
5 А.П. Ротштейн, *Интеллектуальные технологии идентификации: нечеткие множества, генетические алгоритмы, нейронные сети*, Винниця, Україна: УНИВЕРСУМ, 1999.
6 SIEM Rules or Models for Threat Detection? Exabeam, 2018.[Online]. Available: https://www.exabeam.com/siem/siem-threat-detection-rules-or-models/. Accessed on: November29, 2020.

РАЗРАБОТКА СИСТЕМЫ УПРАВЛЕНИЯ АВТОМАТИЗИРОВАННЫМ УСТРОЙСТВОМ ДЛЯ СЪЁМА ПЛОДОВ

Хорт Дмитрий Олегович [1], Кутырёв Алексей Игоревич[1], Пупин Даниил Сергеевич[1], Киктев Николай Александрович [2,3]

[1] Федеральный научный агроинженерный центр ВИМ, 1-й Институтский проезд, 5, 109428, Москва, Российская Федерация

[2] Национальный университет биоресурсов и природопользования Украины, ул. Героев обороны, 15, 03041 Киев, Украина

[3] Киевский национальный университет имени Тараса Шевченко, ул. Владимирская, 64/13, 01601 Киев, Украина, nkiktev@ukr.net

В программной среде САПР Autodesk Fusion 360 разработали 3D модель устройства для съема плодов. Согласно вычисленным нагрузкам провели прочностной расчёт. Разработали структурную схему системы управления автоматизированного съема плодов. Обосновали схемы подключения датчиков к микроконтроллеру и провели их калибровку.

Ключевые слова: съем плодов, возделывания плодовых культур, уборка урожая, система управления, автоматизированное устройство, манипулятор.

Введение

В системе механизированного процесса возделывания плодовых культур уборка урожая является важным завершающим этапом, который требует разработки новых, удобных, в том числе не повреждающих плоды автоматизированных технических устройств, установленных на роботизированные платформы, способных в автономном режиме проводить сбор урожая плодов. Поэтому разработки по созданию автоматизированных устройств для снятия плодов фруктовых насаждений с минимальными повреждениями (или без них) на высоте до 5 метров являются актуальной задачей. Существующие модели промышленных роботов не могут быть непосредственно применены для выполнения технологических процессов погрузки, выгрузки, сортировки и сбора урожая яблок. В частности, для сбора урожая необходимо разрабатывать специальные исполнительные устройства, захватные приспособления и новые алгоритмы их управления для сбора урожая плодовых насаждений в полевых условиях.

Постановка задачи

Для установления оптимальных конструктивных параметров устройств съема, обоснования параметров системы их управления и успешного внедрения данной технологии в производственный процесс необходимо проведение научных исследований. Самоходное роботизированное техническое средство с разрабатываемым автоматизированным захватом-манипулятором позволит качественно проводить технологическую операцию сбора урожая плодов в промышленных садовых насаждениях без участия человека. Цель исследования – разработка и обоснование параметров системы управления автоматизированного устройства для съема плодов яблони.

Материалы и методы

В результате проведенных исследований в программной среде САПР Autodesk Fusion 360 разработана 3D модель устройства для съема плодов (рис.1). В качестве исполнительного механизма использован линейный шаговый актуатор Nema 18, имеющий следующие технические характеристики: масса 80 г, длинна штока 110 мм, максимальная скорость хода штока 4 мм/с, точность позиционирования 0.001 мм, номинальное напряжение питания 4.2 В, максимальный потребляемый ток 0.5 А. Максимальные нагрузка на лапы захвата составляет 9,3 МПа, максимальная деформация 0,3 мм. Разработана структурная схема системы управления для автоматизированного съема плодов, которая включает в себя комплекс аппаратных и программных средств, микроконтроллер (МК), датчик срыва плода (ДСП), датчик давления плода (ДДП), драйвер шагового двигателя (ДШД), линейный актуатор (ЛА), компьютер управления манипулятором и системой распознавания плодов (К) (рис.2). Все элементы аппаратной части находятся внутри основания устройства для съема плодов, через аналоговые порты приходят данные с датчиков давления и срыва плода. Связь с компьютером осуществляется с помощью стандартного последовательного интерфейса UART. Программная часть устройства представляет собой программный код, который управляет положением шагового линейного актуатора, обрабатывает данные с датчиков давления и срыва плода, считывает и отправляет данные в компьютер. В качестве программируемого микроконтроллера использован Atmel ATmega328p в корпусе 28-PDIP. Для калибровки датчика срыва плода, он был закреплен на опытном образце автоматизированного устройства для съема плодов, который работал в совокупности с манипулятором. Работа манипулятора велась вручную с помощью пульта дистанционного управления. Робот-манипулятор эмитировал работу по сбору одного яблока, она начиналась от момента начала срыва, а заканчивалась моментом складывания яблока в емкость для транспортировки. Датчик срыва плода в процессе работы манипулятора, постоянно меняет свое положение, тем самым измерение на одной оси координат будет недостоверным, для точной регистрации момента срыва

плода проведены замеры ускорения устройства захвата по трём координатам x, y, z.

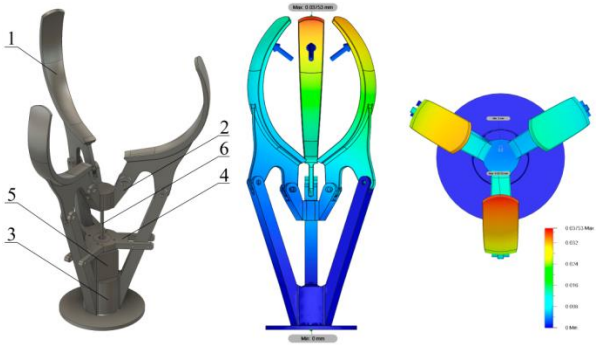


Fig. 11. 3D модель устройства для съема плодов, результаты анализа напряженно-деформированного состояния конструкции: 1 – лапы захвата; 2 – толкатель; 3 – основание; 4 – крепление шагового двигателя; 5 – линейный шаговый двигатель; 6 – шток линейного шагового двигателя

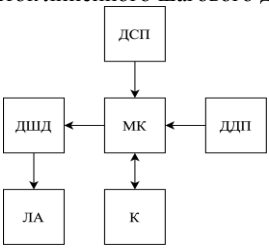


Fig. 2. Структура системы управления для автоматизированного съема плодов

Результаты и обсуждение

В результате проведенных исследований разработана интеллектуальная система управления оборудованием роботизированного устройства для съёма плодов, которая способна в кротчайшее время определить момент срыва плода с плодоножки, а также способна контролировать степень давления лап захвата на плод. Захват изготовлен методом 3D печати. В качестве материала использован полиамид 12 (ПА 12), более известный как нейлон, данный материал отличается высокой прочность износостойкостью, стойкостью к атмосферному воздействию. После сборки опытного образца автоматизированного устройства проведены лабораторные эксперименты.

Проведена проверка работоспособности разработанного захвата для сбора плодов, а именно яблок. Для этого захват был установлен на манипуляторе, после чего проведена серия экспериментов на искусственной модели яблоневого дерева с плодами различных размерных параметров (рис. 3.).

Fig. 3. Процесс проведения лабораторного эксперимента, захват плода разработанным устройством.

References

1. Khort D., Kutyrev A., Smirnov I., Osypenko V., Kiktev N. (2020) Computer Vision System for Recognizing the Coordinates Location and Ripeness of Strawberries. In: Data Stream Mining & Processing. DSMP 2020. Communications in Computer and Information Science, vol 1158. Springer, Cham.
2. P. Baranowski, W. Mazurek, J. Pastuszka-Wozniak. Supervised classification of bruised apples with respect to the timeafter bruising on the basis of hyperspectral imaging data // *Postharvest Biology and Technology.* 2013. N86. 249-258
3. A.K. Bhatt, D. Pant. Automatic apple grading model development based on back propagation neural network and machine vision, and its performance evaluation // *AI & Soc.* 2015. N30(1). 45-56.
4. Zhang, W. Huang, L. Gong, J. Li, C. Zhao, C. Liu, D. Huang Computer vision detection of defective apples using automatic lightness correction and weighted RVM classifier // *Journal of Food Engineering.* 2015. N146. 143-151.
5. O. Kleynen, V. Leemans, M.-F. Destain Development of a multi-spectral vision system for the detection of defects on apples // *Journal of Food Engineering*. 2005. N69. 41-49.

МАТЕМАТИЧНІ УМОВИ АДАПТОВАНОЇ ЗМІНИ ПОРЯДКУ НАДАННЯ ОБЧИСЛЮВАЛЬНИХ РЕСУРСІВ КОРИСТУВАЧАМ ХМАРНИХ ОБЧИСЛЕНЬ

Матов О.Я.
Інститут проблем реєстрації інформації НАН України

В доповіді розглянуті питання подальшого розвитку засад створення адаптивних інфраструктур хмарних обчислень, здатних динамічно адаптуватися до вимог користувачів та діючих особливостей і змін умов функціонування. Для такої адаптації запрапонована динамічна адаптивна змішана дисципліна надання обчислювальних ресурсів користувачам ХО [4]. Розроблені методи та аналітичні умови адаптації надання ресурсів користувачам хмарних обчислень. Ці умови надають можливість створювати технологію (механізми та алгоритми) використання адаптивної дисципліни (порядку) надання обчислювальних ресурсів користувачам. В свою чергу, це дозволяє забезпечувати часові вимоги різних користувачів на отримання своєчасних результатів обчислень чи найбільш ефективно використовувати наявні ресурси хмарних обчислень. . Це є актуальним для систем реального масштабу часу і, в першу чергу, для спеціальних інформаційних систем, що побудовані з використанням приватних хмар, і може бути критичним при обмежених обчислювальних ресурсах хмарних обчислень.

Аналітичні (формульні) умови адаптації розробляються на основі відповідних показників ефективності та математичних моделей хмарних обчислень. Стахостичний характер головних чинників та необхідність кількісної оцінки масових процесів на основі теорії імовірності обусловлює використання аналітичної моделі хмарних обчислень як багатопотокової та багатопріоритетної системи масового обслуговування з чергами зі змішаною дисципліною обслуговування. Модель враховує вірогідні відмови та різні особливості та має довільні закони розподілу для деяких вірогідних процесів.. Модель дозволяє вираховувати часову характеристику - час відгуку системи в умовах особливостей функціонування і відмов хмарних обчислень [2,3].

Показники ефективності надання ресурсів користувачам та математична модель

Розглянемо дві практичні задачі динамічної адаптації змішаної дисципліни надання ресурсів з відносно-абсолютними пріоритетами. Одними з основних показників ефективності ХО є показники, що базуються на оцінці часових характеристик цих систем і які необхідно підтримувати на заданому рівні Такі показники можуть задаватися договором між постачальником та користувачем ХО. Унаслідок випадкового характеру обчислювального процесу виникають додаткові затримки в обробці інформації, порушуються

припустимі обмеження на час її перебування в ХО, що негативно позначається на ефективності рішення цільових задач користувачів.

Для забезпечення необхідної ефективності ХО у таких ситуаціях необхідно підтримувати часові характеристики системи на заданому рівні. В умовах дефіциту обчислювальних ресурсів це можливо тільки за рахунок підвищення ефективності обчислювального процесу, зокрема, за рахунок адаптації дисципліни обслуговування.

Поряд з цим виникає друга задача - найбільш ефективного використання наявних обчислювальних ресурсів у кожен момент часу функціонування керуючих ХО. Цю задачу також можна вирішити шляхом адаптації дисципліни обслуговування.

Метою роботи є розробка методів та аналітичних умов адаптація надання обчислювальних ресурсів користувачам ХО за для забезпечення часових характеристик інформаційно-аналітичних систем та оптимізації використання ресурсів ХО.

В якості показника ефективності ХО беремо середню сумарну вартість (штрафу) часу відгуку ХО (часу затримки в ХО, часу очікування в чергах та часу надання ресурсів тобто перебування в ХО як у системі масового обслуговуванні (СМО)) на заявки (вимоги) користувачів. Для цього використаємо відомий функціонал [6]

$$C^{(S)} = \sum_{i=1}^{n} \alpha_i \lambda_i v_i^{(s)} ,$$

з чого маємо

$$C^{(\phi)} = \sum_{m=1}^{M} \sum_{n=1}^{N} \alpha(m,n)\lambda(m,n)v^{(\phi)}(m,n) , \qquad (1)$$

де

α_i - вартість (штраф) за одиницю часу відгуку ХО (затримки, перебування в ХО) заявок і-го потоку; λ_i - інтенсивність і-го потоку заявок; $v_i^{(S)}$ - середній час відгуку ХО заявок і-го потоку; n – кількість типів заявок; s – параметр, що характеризує спосіб організації обчислювального процесу; $v^{(\phi)}(m,n)(m=\overline{1,M}, \quad n=\overline{1,N_m})$ - середній час відгуку ХО заявок (m, n)-го потоку; α (m, n) - вартість одиниці часу відгуку ХО (затримки в ХО) заявок (m, n)-го потоку; λ (m, n) - інтенсивність (m, n)-потоку.

Цей показник ефективності базується на припущенні, що результати використання ресурсів користувачем знецінюються пропорційно часу їх затримки в ХО, тобто перебування в ХО як у СМО. Тоді цілями адаптації змішаної дисципліни обслуговування будуть або задоволення вимог вчасного перебування (m, n) заявок у системі, що задаються припустимими значеннями цього часу $v_{Д}(m,n)$, або мінімізація функціонала (1). Ці цілі досягаються шляхом відшукання відповідних оптимальних розбивок ϕ^0 на

відносни та абсолютні пріоритети, тобто задачі адаптації змішаної дисципліни обслуговування з відносно-абсолютним пріоритетом являють собою оптимізаційні задачі. Рішення задач відшукання оптимальної розбивки за допомогою відомих аналітичних методів оптимізації не представляється можливим. Єдиний шлях рішення цих задач – евристичний підхід, що не має формального обгрунтування, а спирається лише на специфіку задач (математичних моделей) і зв'язані з ними розуміння.

Досягнення цілей адаптації змішаної дисципліни обслуговування сполучено з потребою оцінки значення середнього часу відгуку ХО (перебування в ХО) заявок (m,n)-типу v (m,n) на ресурси. Тому виникає необхідність синтезу математичної моделі ХО зі змішаною дисципліною надання обчислювальних ресурсів (обслуговування).

Розробка математичних моделей хмарних обчислень чи інформаційних систем, що створені з використанням хмар, є важливим напрямком для виявлення та покращення їх характеристик [2,3]. Хмарні обчислення є об'єктами з високим рівнем невизначеності процесу функціонування. Тут зовнішню невизначеність потоку запитів на обчислювальні ресурси (ОР) доповнює внутрішня невизначеність ХО, що зв'язана з наявністю чи відсутністю необхідних ОР, випадкових відмов системи ХО, а також необхідність забезпечення певних часових характеристик для багатьох клієнтів. Саме це визначає необхідність введення адаптації у процес функціонування ХО. Крім того, введення адаптації у процес функціонування ХО зв'язано з необхідністю підтримки системи в оптимальному, а іноді і просто працездатному стані незалежно від численних факторів зовнішнього та внутрішнього характеру, що виводять ХО з необхідного цільового стану.

Використана аналітична модель, яка базується на роботах [2,3,6]. Модель дозволяє врахувати всі необхідні особливості для вирахування часових характеристик в умовах функціонування ХО з динамічною адаптивною змішаною (абсолютно – відносними пріоритетами) дисципліною надання обчислювальних ресурсів користувачам ХО з і врахуванням відмов та специфікою накопичення заявок на ресурси.

Методи та аналітичні умови адаптації надання ресурсів користувачам ХО

Адаптація змішаної дисципліни обслуговування з моделлю ХО полягає у відшуканні оптимальної розбивки потоків заявок по групах (рівням) абсолютного пріоритету (ϕ^0), тобто такої сукупності чисел $\{N_m\} m=\overline{1,M}$, при якій тимчасові характеристики моделі ХО забезпечували б відповідно до задачі (6) рівність:

$$\phi^0 = opt\{N_1, N_2, \ldots, N_M / \nu^{(\phi)}(m,n) \le \nu_{Д}(m,n), \phi \in \Phi, M = \min\} , \qquad (2)$$

та у відповідності з задачею рівність:

$$\phi^0 = opt\{N_1, N_2, \ldots, N_M / C^{(\phi)} = \min, \phi \in \Phi, M = \min\} . \qquad (3)$$

Оскільки кількість потоків заявок N є скінченною, то задача відшукання оптимальної розбивки ϕ^0 може бути вирішена шляхом повного перебору всіх можливих розбивок і вибору з них такої, котра задовольняє рівностям (17) і (18). Однак цей шлях для ХО реального масштабу часу неприйнятний, тому що кількість усіх можливих розбивок $\Phi = 2^{N-1}$ при великих N велике і реалізація методу повного перебору вимагає значних часових витрат. Тому виникає необхідність розробки таких методів адаптації, що дозволяють одержувати оптимальну розбивку в результаті розгляду обмеженої кількості варіантів групування.

Для відшукання розбивки ϕ^0, що забезпечує рівність (2), пропонується метод, суть якого полягає в почерговому задоволенні вимог до часу перебування заявок у системі, починаючи з першого потоку, шляхом послідовного формування спочатку першої, потім другої і т.д. груп абсолютного пріоритету. При цьому процес адаптації починається з розбивки, що відповідає дисципліні обслуговування з «чистим» відносним пріоритетом ($M=1$, $N_1=N$). У зв'язку з цим перша з розбивок, при якій виконується мета адаптації, характеризується мінімально можливою кількістю груп абсолютного пріоритету M, тобто є оптимальним.

Для відшукання розбивки ϕ^0, що задовольняє рівності (3), пропонується метод адаптації, суть якого складається в цілеспрямованому формуванні груп абсолютного пріоритету, починаючи з останньої, на основі аналізу знака прирощення величини середньої сумарної вартості перебування заявок у системі $\Delta C^{(\phi)}$. При формуванні чергової групи заявки потоків сформованих груп виключаються з розгляду, оскільки вони не впливають на середній час перебування в системі заявок потоків попередніх груп абсолютного пріоритету. Процес адаптації в цьому випадку починається з розбивки, що відповідає дисципліні обслуговування з «чистим» абсолютним пріоритетом ($M=N$, $N_m=1$), що також забезпечує мінімальне число груп M при виконанні мети адаптації.

Визначимо $\Delta C^{(\phi)}$. Припустимо, що попереднє *q-розбивка* має вид $N_1 = N_2 = \ldots = N_{l+1} = 1, \quad N_{l+2} = N_j = P - l - 1,$ де P – кількість потоків заявок, розглянутих на етапі формування чергової групи з номером j. При нумерації груп з останньої $P = N - \sum_{m=1}^{j-1} N_m$. Наступна $\phi-$розбивка відрізняється від *q-розбивки* тим, що потоки заявок двох останніх груп об'єднані в одну, тобто $N_1 = N_2 = \ldots = N_l = 1, \quad N_{l+1} = N_j = P - l$.

При $\Delta C^{(q,\phi)}$ при переході від *q-розбивки* до $\phi-$розбивки визначимо в такий спосіб:

$$\Delta C^{(q,\phi)} = C^{(\phi)} - C^{(q)} = \sum_{i=1}^{N} \alpha_i \lambda_i \Delta \nu_i^{(q,\phi)} \ , \qquad (4)$$

де $\Delta \nu_i^{(q,\phi)} = \nu_i^{(\phi)} - \nu_i^{(q)}$, $i = \overline{1,N}$.

З формули (4) випливає, що ϕ – розбивка вважається кращою у порівнянні з *q-розбивкою*, якщо $\Delta C^{(q,\phi)} \le 0$. У випадку $\Delta C^{(q,\phi)} > 0$ кращою є *q-розбивка*. Вираження $\Delta C^{(q,\phi)} = 0$ означає, що ϕ – розбивка за критерієм середньої сумарної вартості перебування заявок у системі не гірше *q-розбивки*, але забезпечує меншу кількість груп абсолютного пріоритету *M*.

Обчислимо $\Delta C^{(q,\phi)}$ на прикладі системи типу АС-І [3], для якої можна записати:

$$\nu_i^{(\phi)} = \begin{cases} \dfrac{b_i}{K_r - R_{i-1}} + \dfrac{K_r^3 \rho_0 \Delta_0 + \sum\limits_{r=1}^{i} \rho_r \Delta_r}{(K_r - R_{i-1})(K_r - R_i)}, & i = \overline{1,l}; \\ \dfrac{b_i}{K_r - R_l} + \dfrac{K_r^3 \rho_0 \Delta_0 + \sum\limits_{r=1}^{P} \rho_r \Delta_r}{(K_r - R_{i-1})(K_r - R_i)}, & i = \overline{l+1,P}. \end{cases} \qquad (5)$$

При переході від *q-розбивки* до ϕ – розбивки збільшення середнього часу перебування заявок у системі $\Delta \nu_i^{(q,\phi)}$

$$\Delta \nu_i^{(q,\phi)} = \begin{cases} 0 \qquad\qquad\qquad\qquad , & i = \overline{1,l}; \\ \dfrac{\sum\limits_{r=l+2}^{P} \rho_r \Delta_r}{(K_r - R_l)(K_r - R_{l+1})}, & i = l+1; \\ -\dfrac{b_i \rho_{l+1}}{(K_r - R_l)(K_r - R_{l+1})}, & i = \overline{l+2,P}. \end{cases} \qquad (6)$$

Тоді збільшення $\Delta C^{(q,\phi)}$ запишемо у виді

$$\Delta C^{(q,\phi)} = \alpha_{l+1} \lambda_{l+1} \Delta \nu_{l+1}^{(q,\phi)} + \sum_{i=l+2}^{P} \alpha_i \lambda_i \Delta \nu_i^{(q,\phi)} = \\ = \frac{\rho_{l+1}}{2(K_r - R_l)(K_r - R_{l+1})} \sum_{i=l+2}^{P} \lambda_i b_i^{(2)} \left(\frac{\alpha_{l+1}}{b_{l+1}} - \frac{2}{1+\phi_i^2} \frac{\alpha_i}{b_i} \right), \qquad (7)$$

де $\psi_i = \sqrt{D[t_i]} / b_i$ - коефіцієнт варіації часу обслуговування заявок *i-го* потоку ($D[t_i]$ - дисперсія часу обслуговування). При показовому

обслуговуванні заявок *i-го* потоку $\psi_i = 1$, а при детермінованому обслуговуванні - $\psi_i = 0$.

Аналіз виразу (22) показує, що доцільність переходу від *q-розбивки* до ϕ – розбивки визначається знаком вхідної в нього суми:

$$\Delta C_l = \sum_{i=l+2}^{P} \lambda_i b_i^{(2)} \left(\frac{\alpha_{l+1}}{b_{l+1}} - \frac{2}{1+\psi_i^2} \frac{\alpha_i}{b_i} \right). \qquad (8)$$

З цієї рівності випливає:

1) якщо $\Delta C_l \le 0$ для усіх $l = \overline{0, N-2}$, то оптимальною є дисципліна обслуговування з «чистим» відносним пріоритетом;
2) при $\Delta C_l > 0$ для всіх l оптимальною є дисципліна обслуговування з «чистим» абсолютним пріоритетом;
3) у випадку $\Delta C_l \le 0$ *или* $\Delta C_l > 0$ не для всіх l оптимальною є змішана дисципліна обслуговування.

Таким чином, для визначення доцільності переходу від *q-розбивки* до ϕ – розбивки досить по формулі (8) обчислити ΔC_l і проаналізувати результат. Зміна розбивки доцільна, якщо $\Delta C_l \le 0$.

Висновок

Розробка аналітичних умов адаптації надання ресурсів користувачам хмарних обчислень виконано на основі аналітичної моделі. Аналітичні умови дозволяють розробляти механізми та алгоритми адаптації ХО. Ці механізми та алгоритми враховують таки фізичні властивості ХО, як миттєву еластичність (динамічну міграцію, виділення і звільнення ресурсів для швидкого масштабування відповідно до потреб) та сервіси вимірювання (керування і оптимізація ресурсів за допомогою засобів вимірювання). Момент включення алгоритму адаптації визначається системою управління ХО при порушенні припустимих обмежень на час відгуку, зміні контрольованих параметрів системи (наприклад, її сумарного завантаження) чи показника ефективності системи понад граничні значення.

Посилання

1. Aleksandr Matov. Adaptation of cloud computing as optimization of the process of rendering services to users in the conditions of limited computing resources. // Selected Papers of the XIX International Scientific and Practical Conference "Information Technologies and Security" (ITS 2019). CEUR Workshop Proceedings. - Vol-2577. – pp. 210-221.
2. Матов О.Я. Аналітичні моделі багатопріорітетних хмарних дата-ценрів зі змішаною дисципліною надання послуг з урахуванням особливостей функціонування і можливих відмов. Реєстрація, зберігання і обробка даних. 2019. Т.21, №1 С.32 - 45.2

3. Aleksandr Matov. Mathematical models of cloud computing with absolute-relative priorities of providing of computer resources to users in conditions of functioning features and failures // CEUR Workshop Proceedings (ceur-ws.org). Vol-2318 urn:nbn:de:0074-2318-4. Selected Papers of the XVIII International Scientific and Practical Conference on Information Technologies and Security (ITS 2018) PP. 150-159 Matov A.Y.
4. Матов О.Я. Оптимізація надання послуг обчислювальними ресурсами адаптивної хмарної інфраструктури. Реєстрація, зберігання і обробка даних. 2018. Т.20, №3 С.83-90.
5. Матов А.Я., Шпилев В.Н., Комов А.Д. и др.. Организация вычислительных процессов в АСУ. Под ред. А.Я.Матова. Киев, 1989. – 200 с.

ДОСЛІДЖЕННЯ АЛГОРИТМІВ ВСТАНОВЛЕННЯ АСОЦІАТИВНИХ ПРАВИЛ У ІНТЕЛЕКТУАЛЬНИХ СИСТЕМАХ АНАЛІЗУ ДАНИХ

Безуглий Олексій Микитович
ІПРІ НАН України, Київ, abezugliy00@gmail.com

Мета дослідження полягає у порівнянні методів пошуку асоціативних правил у інтелектуальному аналізі даних з метою виявлення найефективнішого алгоритму. Метою аналізу є встановлення наступних залежностей: якщо в транзакції зустрівся деякий набір елементів X, то на підставі цього можна зробити висновок про те, що інший набір елементів Y також же повинен з'явитися в цій транзакції. Встановлення таких залежностей дає нам можливість знаходити дуже прості і інтуїтивно зрозумілі правила.

Ключові слова: Інтелектуальний Аналіз, Асоціативні Правила, Секвенційний Аналіз, БД, Закономірності, data mining, Apriori, Frequent Pattern-Growth (FPG), AprioriTID, AprioriHybrid, AprioriDHP, PRTITION, DIC (Dynamic Items Counting)

Вступ

Інтелектуальний аналіз даних – процес виявлення корисних відомостей з великих наборів даних.[1] В ході його проведення досліджуються різні варіанти подій, які можуть бути представлені у вигляді таблиці. Рядки являють собою будь-який варіант, а стовпці містять його параметри.

Секвенційний аналіз – це різновид інтелектуального аналізу даних (data mining). Об'єктом секвенційного аналізу є база послідовностей.

Секвенційний аналіз[1] виявляє часто повторювані послідовності подій. Тобто в якій послідовності відбуваються події або через якийсь проміжок часу після події "X", відбудеться подія "Y". Дані по одній самій події, але взяті з різних послідовностей.

Інтелектуальний аналіз складається з класифікації, кластерізації, ассоциації, послідовності, прогнозування, виявлення відхилень, оцінювання, аналізу зв’язків та візуалізація.

Класифікація – виявлення ознак що характеризують групи об’єктів. За цими ознаками нові об’єкти розподіляють по класам.

Кластеризація – є логічним продовженням класифікації. Результатом кластеризації є поділення об’єктів на групи.

Ассоциації – метою є пошук закономірностей між подіями на одному проміжку часу.

Пошук асоціативних правил полягає у визначенні частих наборів серед великих наборів даних. У контексті аналізу кошика покупця це пошук наборів товарів, які найбільш часто купують разом. Не менш важливим фактором є взаємозв'язок подій у часі. Ґрунтуючись на тому, які події найчастіше йдуть за іншими, з’являється можливість заздалегідь передбачати їх появу, що дозволить приймати більш своєчасні рішення.

Пошук асоціативних правил полягає у виявленні набору об'єктів в рамках однієї транзакції, тобто товари які найчастіше купуються разом за одну транзакцію.

Послідовність (послідовна ассоціація) – знаходить закономірності у часі між транзакціями.

Прогнозування – базується на історичних даних з метою оцінки пропущених чи майбутніх числових значень.

Виявлення відхилень(выкид) – метою є виявлення та аналіз даних, що відрізняються від загальної множини.

Оцінювання – метою є прогнозування безперервних значень ознак.

Аналіз зв'язків – пошук залежностей у наборі даних.

Візуалізація – використовуються графічні методи, що демонструють наявність залежностей у даних.

Актуальність

Актуальною задачею є порівняння методів пошуку асоціативних правил та знаходження оптимальних алгоритмів встановлення зв'язків серед наборами подій у великих базах даних. З метою економії ресурсів комп'ютеру та зменшення часових витрат на обробку великих даних (big data).

Асоціативні правила ефективно використовуються в сегментації покупців по поведінці при покупках, аналізі переваг клієнтів, плануванні розташування товарів в супермаркетах, крос-продажах, адресній розсилці. Однак, сфера застосування цих алгоритмів не обмежується лише однією торгівлею. Їх також успішно застосовують і в інших областях: медицині, для аналізу відвідувань веб-сторінок (Web Mining), для аналізу тексту (Text Mining), для аналізу даних по перепису населення, в аналізі та прогнозуванні збоїв телекомунікаційного устаткування то що.

Огляд існуючих рішень

Для проведення секвенційного аналізу існує два найпопулярніші алгоритма.

- Алгоритм Apriori – це один з перших алгоритмів, спрямованих на пошук асоціативних правил[2].
- Алгоритм Frequent Pattern-Growth (FPG) – один з найбільш ефективних алгоритмів пошуку асоціативних правил, який дозволяє не використовувати ресурсо-витратну генерацію кандидатів.

Основний матеріал

Суттю алгоритму **Apriori** є використання властивості підтримки: підтримка будь-якого набору об'єкта не перевищує мінімальну підтримку будь-якого з його підмножин: $Sup_F \leq Sup_E$ при тому, що $E \subset F$.

Алгоритм знаходить найчастіші набори даних в кілька етапів, кожний з яких складається з послідовності кроків:

- Формування кандидатів
- Підрахунок кандидатів

Недоліком даного алгоритму є процес генерації кандидатів, який вимагає великих обчислювальних і багаточасових витрат. До того ж, він вимагає багаторазового сканування бази даних і не враховує часову складову в наборах даних. Тому у даного алгоритму існує безліч модифікацій.

Формування кандидатів – це етап на якому алгоритм сканує БД та формує елементарних кандидатів. На цьому етапі підтримка кандидатів не формується.

Підрахунок кандидатів – етап на якому вираховується підтримка кандидатів, а також відсікаються кандидати з недостатнім рівнем підтримки, а найчастіші кандидати лишаються.

На першому етапі відбувається формування одно елементних кандидатів. Далі алгоритм підраховую підтримку цих кандидатів. Набори з рівнем підтримки менше заданого відсікаються.

Далі проходить етап формування двоелементних кандидатів, підрахування їх підтримки та видалення кандидатів з підтримкою менше мінімальної заданої. Алгоритм працює так поки не відскакує всю БД.

Якщо подивитися на алгоритм прямолінійно, на останньому етапі алгоритм формує n-елементний набор, підраховує їх підтримку та відсікає набори з рівнем підтримки з рівнем менш заданого. Однак алгоритм Apriori зменшує кількість кандидатів відсікаючи – апріорі – ті, які завідомо не можуть стати часто повторюваними на підґрунті інформації про відсічені кандидати на попередніх етапах роботи алгоритма.

Різновиди алгоритму Apriori необхідні для скорочення кількості сканувань бази даних, кількості наборів кандидатів або того та іншого. Було запропоновано декілька різновидів алгоритма Apriori: AprioriTID та AprioriHybrid.

Головна особливість алгоритму AprioriTID полягає у тому, що БД не використовується у підрахунку підтримки кандидатів набору після першого проходу. З цією метою використовується кодування кандидатів, яке було виконане на перших проходах. У наступних проходах розмір закодованих наборів може бути значніше менш ніж БД – таким чином економляться значні ресурси.

Різниця між Apriori і AprioriTID полягає у тому, що перші проходи Apriori по БД значно швидше ніж AprioriTID, а на етапах завершення AprioriTID швидше ніж Apriori. AprioriHybrid об’єднує найкращі фактори двох алгоритмів.

Алгоритм AprioriDHP, також має назву алгоритм хешування. В основі його роботи – імовірнісний підрахунок наборів-кандидатів. Він служить для скорочення числа підраховувань кандидатів на кожному етапі виконання алгоритма Apriori.

Алгоритм PRTITION розбиття (розділення) сканує БД шляхом розбиття її на розділи, що не перетинаються, які можуть вміститися в оперативній пам’яті. На першому етапі, за допомогою Apriori, виявляються «локальні» найчастіші набори даних. На другому етапі підраховується підтримка

кожного з наборів відносно всій БД. Таким чином на другому етапі виявляються всі потенціальні набори даних.

Алгоритм DIC (Dynamic Items Counting) розбиває БД на декілька блоків, кожен з яких відмічається так званими «початковими точками», а після циклічно сканує БД.

В основі алгоритму **Frequent Pattern-Growth (FPG)** лежить попередня обробка БД, в процесі якої вона трансформується в компактну деревоподібну структуру, яка називається Frequent-Pattern Tree - дерево популярних предметних наборів.[3] До основних переваг даного методу відносяться:

1. Стиснення БД в компактну структуру, яка забезпечує найбільш ефективне і повне вилучення частих предметних наборів;
2. При побудові FP-дерева використовується технологія «розділяй та владуй», яка дозволяє виконати декомпозицію однієї складної задачі на безліч більш простих;
3. Дозволяє уникнути витратної процедури генерації кандидатів, характерною для алгоритму Apriori.

В FP-дереві кожен елемент зображується у вигляді вузла з індексом, що вказує на кількість його повторень. Це – найбільш вигідний варіант подання бази даних транзакцій, так як в ній кожен елемент може неодноразово повторюватися.[4]

Зі збільшенням кількості даних в БД витрати часу на пошук найчастіших наборів для алгоритму Apriori ростуть на кілька порядків швидше, ніж для алгоритму FPG.

Висновок

Вузьким місцем в алгоритмі Apriori є процес генерації кандидатів в популярні предметні набори. Наприклад, якщо база даних (БД) транзакцій містить 100 предметів, то буде потрібно згенерувати 2^{100}~10^{30}кандидатів. Таким чином алгоритм Apriori менш ефективний ніж його аналог.

Крім цього, алгоритм Apriori вимагає багаторазового сканування БД транзакцій, а саме стільки разів, скільки предметів містить найдовший предметний набір. Тому було запропоновано ряд алгоритмів, які дозволяють уникнути генерації кандидатів і скоротити необхідну кількість сканувань набору даних.

Список літератури

1. https://izv.etu.ru/assets/files/izvestiya-3_2020-45-52.pdf стр. 45
2. http://bourabai.kz/tpoi/analysis5.htm
3. https://izv.etu.ru/assets/files/izvestiya-3_2020-45-52.pdf стр. 48-49
4. https://loginom.ru/blog/fpg#:~:text=FPG альтернативный алгоритм поиска ассоциативных правил,-19 октября 2020 text=Кроме алгоритма Apriori для поиска,часто встречающихся предметных наборов.

ANALYSIS OF CYBER EXERCISES APPROACHES

Vasyl Tsurkan[1[0000-0003-1352-042X]] and Valeriia Pokrovska[1[0000-0002-1318-5521]]
[1] Institute of special communication and information protection National technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute",
Kyiv-056, Ukraine
v.v.tsurkan@gmail.com, Hilariyap@gmail.com

Raising awareness of the organization's employees comes down to cyber exercise. They can target both an individual employee and specialists in general. This is implemented mainly in four stages of the organization of the cyber exercises. In the first stage, the purpose, scenarios, evaluating system of results for their execution, and the scenario-modeling environment is established. It is tested for compliance with the purpose of cyber exercises within the second stage. In the third stage, cybersecurity scenarios are being developed. The results of the execution are evaluated in the fourth stage. This is due to the relevance of the analysis of approaches to the organization of cyber exercises. In solving this problem, it was established that there is no uniform interpretation of this concept. First, such ambiguity of interpretations is associated with the direction of cyber exercises. Therefore, approaches to their organization are focused on obtaining theoretical knowledge, practical skills, and cybersecurity skills. Primarily, an incident detection and prevention approach is common. Also, the assessment of cybersecurity through penetration testing. The application of these approaches can be generalized and organized as a game.

Keywords: cybersecurity, scenario, incident, cyber exercises, cyber exercises approaches.

Introduction

Cyber exercises are organized through awareness-raising programs for cybersecurity organizations. It mainly comes down to training in an interactive form and is characterized by orientation both for the individual employee and for specialists as a whole [1].

There are four stages in the organization of cyber exercises. In the first stage, goals, scenarios, evaluating system of results for their execution, and the scenario-modeling environment is established. It is being tested for compliance with the goals of cyber exercises in the second stage. In the third stage, established cybersecurity scenarios are being worked out. While the results of their execution are evaluated at the fourth stage [2]. Depending on the purpose of cyber exercises and the training level of organization employees, it can be organized using different approaches.

Cyber exercises approaches

Now there is no unified interpretation of the “cyber exercises”, concept for example: “cyber training”, “cyber range”. Sometimes “cyberlearning” is used, which is used when organizing remote cyber exercises. It is focused primarily on obtaining theoretical knowledge. Unlike “cyberlearning”, a characteristic feature of “cyber training” and “cyber range” is the focus on obtaining practical cybersecurity skills [1], [3].

A typical approach to organizing cyber exercises is characterized by the acquisition of skills and abilities to respond to cybersecurity incidents. First, they focus on both their detection and prevention of manifestation in future activities [4]. In addition, it is possible to focus on assessing cybersecurity through penetration testing. This is the basis of an approach to identifying information vulnerabilities in cyberspace, including the use of social engineering [5]. At the same time, the use of a game approach to organizing cyber exercises is common. Within its framework, two teams are distinguished - attackers and “victims”.

Conclusion

Thus, the organization of cyber exercises is accompanied by the use of various concepts. Each of them determines their specificity, taking into account the orientation of both the individual employee and the specialists as a whole. Such features determine the choice of approaches to the organization of cyber exercises. In particular, they can focus on individual cybersecurity tasks, for example, incident response (“Cyber defense”). Also, to assess cybersecurity (“Compensation Testing”) or to reduce their solution to the game (“War game”).

References

1. Yamin, M., Katt, B., Gkioulos, V.: Cyber Ranges and Security Testbeds: Scenarios, Functions, Tools and Architecture. Computers & Security, vol. 88 (2020), https://doi.org/ 10.1016/j.cose.2019.101636, last accessed 2020/11/10.
2. Kick, J.: Cyber exercise playbook. The MITRE Corporation, Wiesbaden, Germany (2014).
3. Vykopal, J., Vizvary, M., Oslejsek, R., Celeda, P., Tovarnak, D.: Lessons learned from complex hands-on defence exercises in a cyber range, In: IEEE Frontiers in Education Conference (FIE). pp. 1-8. IEEE, Indianapolis, IN, USA, https://doi.org/10.1109/FIE.2017.8190713, last accessed 2020/11/10.
4. Matania, E., Yoffe, L., Goldstein, T.: Structuring the national cyber defence: in evolution towards a Central Cyber Authority. Journal of Cyber Policy, vol. 2, no. 1, 16-25 (2017), https://doi.org/10.1080/23738871.2017.1299193, last accessed 2020/11/10.
5. López de Jiménez R.: Pentesting on web applications using ethical – hacking. In: IEEE 36th Central American and Panama Convention (CONCAPAN XXXVI), pp. 1-6. IEEE, San Jose, Costa Rica, https://doi.org/10.1109/CONCAPAN.2016.7942364, last accessed 2020/11/10.

ЗМІСТ

Національна академія наук України
Інститут проблем реєстрації інформації НАН України

ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ І БЕЗПЕКА

Матеріали XX Міжнародної науково-практичної конференції

Випуск 20

Підп. до друку 15.12.2020. Формат 60x84 1/16. Папір офс. Гарнітура Times.
Спосіб друку – ризографія. Ум. друк. арк. 10,5. Обл.-вид. арк. 24,36. Наклад 100 пр.
Зам. № 15-200.

НТУУ «КПІ» ВПІ ВПК «Політехніка»
Свідоцтво ДК № 1665 від 28.01.2004 р.
03056, Київ, вул. Політехнічна, 14, корп. 15
Тел. (44) 406-81-78