НАЦІОНАЛЬНА АКАДЕМІЯ НАУК УКРАЇНИ

ІНСТИТУТ ПРОБЛЕМ РЕЄСТРАЦІЇ ІНФОРМАЦІЇ НАН УКРАЇНИ

ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ ТА БЕЗПЕКА

МАТЕРІАЛИ XXII МІЖНАРОДНОЇ НАУКОВО-ПРАКТИЧНОЇ КОНФЕРЕНЦІЇ

ВИПУСК 22

Київ – 2022

Рекомендовано до друку Вченою радою
Інституту проблем реєстрації інформації НАН України
(протокол № 14 від 20 грудня 2022 р.)

Інформаційні технології та безпека. Матеріали XXII Міжнародної науково-практичної конференції ІТБ-2022. – Київ: Інжиніринг. – 132 с. ISBN: 978-966-2344-85-1

До збірника увійшли матеріали доповідей, представлених на XXII Міжнародній науково-практичній конференції «Інформаційні технології та безпека» (ІТБ-2022, 16 листопада 2022 року, м. Київ, Україна).

У збірнику представлені матеріали, присвячені питанням функціональної стійкості інформаційних систем, безпеки та живучості критичних інфраструктур, комп'ютерного моделювання складних систем, технологій аналітики даних великих обсягів (Big Data), створення аналітичних систем на основі відкритих джерел інформації (OSINT), моделювання, аналізу та прогнозування процесів мережевої взаємодії, методів і засобів підтримки прийняття рішень.

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

ISBN 978-966-2344-85-1

АНАЛІЗ ТА ОЦІНЮВАННЯ ФУНКЦІОНАЛЬНОЇ СТІЙКОСТІ ІНФОРМАЦІЙНИХ СИСТЕМ, ЩО ПІДТРИМУЮТЬ ПРОЦЕСИ 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

Визначено поняття функціональної стійкості інформаційних систем, операційний цикл, що її характеризує. Обговорюються шляхи забезпечення функціональної стійкості інформаційних систем, а також особливості оцінювання цієї властивості.

Ключові слова: інформаційна система, функціональна стійкість, живучість.

Вступ

Функціонування більшості інформаційних систем (ІС), що задіяні у процесах управління, відбувається за умов постійної взаємодії із перманентно змінним зовнішнім середовищем. Вагому частину таких взаємодій складають різноманітні інформаційні конфлікти, які суттєво впливають на досягнення загальносистемної цілі. Інформаційні конфлікти можуть призводити до руйнування інформаційних ресурсів, порушення регламентів взаємодії та штатних інформаційних процесів, внаслідок чого порушується виконання системних і прикладних функцій, і відповідно виникають порушення в управлінні, які можуть представляти потенційну загрозу людському життю й навколишньому середовищу у разі критичності об'єкта управління. Імовірність виникнення надзвичайних ситуацій, ліквідація яких може потребувати значних матеріальних і людських ресурсів, змушує, оцінюючи ризики їх реалізації, розробляти засоби прогнозування, раннього виявлення та попередження, протидії руйнівним інформаційним впливам зовнішнього середовища.

Основна частина

ІС сьогодні є найважливішим інструментом забезпечення безпеки функціонування об'єктів управління, зокрема критичних інфраструктур. Для уникнення порушень в управлінських процесах

необхідно забезпечити наявність у ІС такої властивості, як функціональна стійкість, що дозволяє ІС зберігати і/або відновлювати виконання функцій в умовах різного роду збурюючих впливів, мінімізуючи ризики переходу в аварійний (небезпечний) стан [1]. Це інтегральна властивість, яка передбачає наявність в ІС певного рівня надійності, відмовостійкості, живучості та безпеки.

Функціональну стійкість можна характеризувати наступним операційним циклом Ω, що включає операції прогнозування w_1, попередження w_2, виявлення w_3, локалізації w_4, ізоляції w_5, парирування w_6 відмов, процедур реконфігурації w_7 і відновлення інформації w_8 . Послідовність і часові межі операцій залежать від специфіки дефектів і особливостей ІС. Для різних дефектів також можна визначити відповідні операційні цикли, так, наприклад, для уразливостей програмних компонент маємо [2]: w_{pr1}– прогнозування можливих характеристик атак на цю уразливість (можливість, спосіб, часові параметри); w_{pr2}– попередження втручання; w_{pr3}– детектування атаки через контроль вхідних даних; w_{pr4}– локалізація (ізоляція) компоненти, на яку здійснена атака; w_{pr7}– реконфігурація, що забезпечить компенсацію можливої відмови внаслідок атаки на уразливість; w_{pr8}– продовження виконання операцій сервісу і вибір стаціонарної конфігурації.

Функціональна стійкість ІС при проектуванні забезпечується традиційно введенням певної надмірності; впровадженням системи вбудованого контролю; формуванням контуру захисту від негативних впливів зовнішнього середовища; використанням компонентів із підвищеним рівнем захищеності і надійності. Та на жаль, додаткова надмірність веде до погіршення техніко-економічних характеристик ІС. Системи контролю відстежують заданий ряд параметрів, але не завжди забезпечують адекватну реакцію на нештатну ситуацію, до того ж не зменшується ймовірність виникнення таких ситуацій. Контур захисту мінімізує вплив зовнішніх факторів, але повністю його не виключає. Вибір елементної бази з підвищеним рівнем захищеності й надійності підвищує відмовостійкість ІС, та не забезпечує функціональної стійкості, коли відмова вже сталася.

Для оцінювання функціональної безпеки ІС використовуються різні процедури і техніки [2–4], які недостатньо систематизовані й узгоджені за вхідними і вихідними параметрами. Для мінімізації ризиків неточного оцінювання необхідно визначитись з порядком їх сумісного і паралельного використання.

Висновки

Функціональна стійкість є важливою характеристикою ІС, які задіяні в процесах управління, для забезпечення і оцінювання якої потрібно застосування спеціальних методів, що базуються на застосуванні діверсності, механізмів підвищення живучості, принципів багато параметричної адаптації до відмов тощо.

Перелік посилань

1. Oleksandr Dodonov, Olena Gorbachyk, Maryna Kuznietsova. Automated Organizational Management Systems of Critical Infrastructure: Security and Functional Stability. CEUR Workshop Proceedings 2021, 3241, pp.1-12. http://ceur-ws.org/Vol-3241/paper1.pdf
2. Харченко В.С., Яковлев С.В., Горбачик О.С. та ін. Забезпечення функціональної безпеки критичних інформаційно-керуючих систем. Харків: Константа, 2019. 272 с.
3. Королев А.Н. Функциональная устойчивость навигационно-информационных систем. // Изв. Вузов. Приборостроение. 2018. Т. 61, №7. С.559-565.
4. Додонов О.Г., Горбачик О.С., Кузнєцова М.Г. Підвищення безпеки критичних інфраструктур засобами автоматизованих систем організаційного управління. // Реєстрація, зберігання і обробка даних, 2022. Т. 24. № 1. С. 74-81.

OPTICAL RECORDING METHODS FOR LONG-TERM DATA STORAGE

Vyacheslav Petrov, Ievgen Beliak and Andriy Kryuchyn
Institute of information recording NAS of Ukraine Kyiv, Ukraine
petrov@ipri.kiev.ua, beliak1312@gmail.com, kryuchyn@gmail.com

Optical media have long been and remain a reliable option for archival storage. The only problem with their widespread use is the limited ability to meet the requirements for the storage of ever-growing volumes of data, which today cannot be fully implemented. If optical recording technology can provide at least tens of terabytes (TB) of capacity on a standard disk, then optical recording can be considered as a promising solution for archival storage [1]. The main disadvantage of modern optical recording systems is the low recording density, which leads to the high cost of storing data on optical media. The main problem with the limited capacity of optical media is the inability to reduce beyond the diffraction limit the diameter of the laser beam, which is used to record and read information. The task of overcoming the optical diffraction limit and increasing the resolution of optical recording systems, and thus increasing the capacity of optical media, is a key task for integrating optical storage technology with trends in information technology, such as storage and processing of large volumes of data and cloud storage.

The use of long-field optical systems for recording significantly limits the surface density of information recording, but allows to create three-dimensional (multilayer) media. Currently, the possibilities of creating three-dimensional optical (3D) storage technologies are not widely used, except for the creation of two and three-layer optical discs [2-4]. An important task that determines the widespread use of optical discs is the creation of media that provide both high recording density and long storage life. When creating optical media for long-term data storage, two central questions emerge. The first of these concerns the material of storage media likely to last long enough to convey a message to long enough period of time into the future. Nowadays special class of microfilms and optical disc-type media of with substrate made of synthesized materials are developed for long-term storage of information issue. The organizing of the modern distributed information system structure includes an analysis of the processes of working with data-in-use at the workstations and storing of data-in-rest at the servers storage and cloud services as it is shown at the Fig. 1.

Optical media is supposed to be highly perspective type of data-in-rest or “cold data” reliable storing. It should be noted that large-scale research is being carried out to create special optical media for long-term data storage based on the use of highly stable materials and recording methods, which support different approaches of digital reading [5-7].

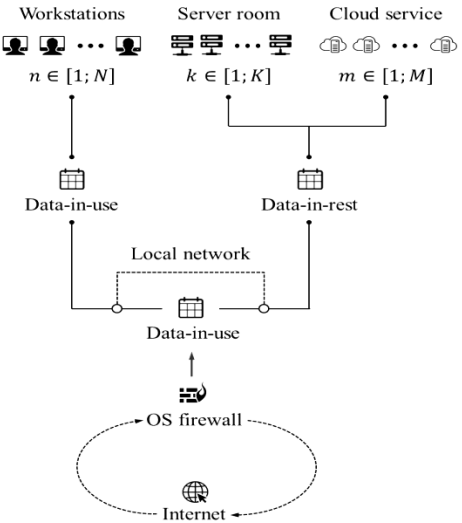

Fig. 1. Distributed information system structure

Faced with the lack of suitable options for storage device production, researchers around the world have re-embraced the use of chemically stable high-strength synthetic materials as adequately durable long-term storage media base. The second crucial problem when creating digital media concerns the choice of the form for the presentation and coding of data to be recorded which is relevant both to ensure the reliable long-term data storage system building and to ensure the unauthorized access prevention. The fact that the possibilities of optical media for storing large amounts of information are far from exhausted, evidenced by recent developments of optical media, which are based on the use of nanocomposite materials and near-field recording method [8-10]. Possibilities of creating optical disks of standard sizes with a capacity of up to 700 TB have been reported [11]. Special attention is paid to the creation of optical media for long-term storage of large amounts of data [5, 11]. The most impressive results in the creation of such media were achieved by Microsoft, which in the framework of the Project Silica

project proved the possibility of long-term storage of data in the volume of quartz glass. The recording is performed by a laser beam by deforming the silicon lattice to create layers of three-dimensional "records" in the thickness of the glass. To reproduce the information recorded in the glass, a machine learning algorithm is used, which decodes the inhomogeneities created by the rays of polarized light [12].

For long-term data storage (almost perpetual), optical media have significant advantages over other types of media. The size of this particular market segment is small compared to the general market of archival repositories [13]. To organize long-term data storage, special types of optical media must be created [5,7]

The crucial problem when creating long-term data storage media concerns the choice of the form for the presentation and coding of data to be recorded. It is even more challenging than choosing materials for storage media. To estimate the service life of storage devices, accelerated tests can be used. However, these methods are not applicable when justifying the choice of presentation and coding methods for data subject to long-term storage. It is proposed to record information on long-term storage media by means of placing diminished graphic or textual images onto the medium, which can be read by optical systems using appropriate magnification. The advantage of this way of presentation is that subsequent retrieval of information doesn't require special reading devices or software. The optical resolution required to read the data is defined by the diminution used for recording. Such data presentation is used on several types of sapphire and metal discs [5,7]. Presenting data as microimages complicates their processing and lowers the processing speed. To ensure high durability of long-term storage of information, it is proposed to use non-binary codes working with digital data on a symbol level, e.g. with bytes of information. Non-binary codes are used in channels with grouped errors as components of cascade codes to ensure error control on various types of optical media (Blu-ray, etc.) [7,14]. This coding method was chosen due to the fact that it had been widely and successfully used to produce billions of copies of compact discs of various types.

Improvements in optical discs development area allowed to increase volume of the storage. However, mentioned technologies are approaching fundamental limits. In particular surface-storage optical recording technology is characterized with diffraction limit while two-dimensional data storage systems resolution depends on the laser beam spot size that is focused onto an optical disc surface. In order to increase media density one should decrease the spot size. Therefore volumetric optical recording

has been developed to overcome the limit of density. Nowadays, there are two different methodologies for approaching three-dimensional optical data storage: one is hologram data storage and the other is three-dimensional bit-wise structures which could be presented as upgrade of optical disc systems. Three dimensional bit-wise optical data storage offers the potential for high recording capacity because bit information can be stored at different layers within a certain volume of the recording media. The information is coded in the form of micro-spots located inside of the substrate material. Thereby three-dimensional optical data storage using a nonlinear optical process are demonstrated to be most productive.

Research into the development of optical memory systems is also aimed at providing additional levels of security when storing large arrays of information. An example of such a medium is a new optical medium based on transparent glass ceramics (TGC) with photoluminescent nanocrystals LiGa5O8: Mn2 +, on which it is possible to record and reproduce data in the mode of photon capture / radiation. Highly ordered nanostructure of the material provides recording with high resolution and low error rate. The high transparency of the developed volumetric recording environment makes it possible to 3D-optical data storage and increase the level of information security by multiplexing signals, reproduction in intensity and wavelength [15].

CONCLUSIONS

1. Optical media have long been and remain a reliable option for archival storage. The only problem with their widespread use is the limited ability to meet the requirements for the storage of ever-growing volumes of data.
2. Volumetric optical recording has been developed to overcome the limit of density. The development of optical memory systems is also aimed at providing additional levels of security
3.As the data coding method, we propose to use non-binary codes working with digital data with grouped errors as components of cascade codes to ensure error control, which are used to produce compact discs.

Reference

1. How to store everything forever…using compact disks? https://blog.seagate.com/business/how-to-store-everything-forever-using-compact-disks/

2. Petrov VV, Kryuchyn AA, Shanoilo SM, Kravets VG, Kossko IO, Belyak EV, Lapchuk AS, Kostyukevich SO Super-dense optical recording of information. - National Academy of Sciences of Ukraine, Institute of Information Recording.– Kyiv: National Academy of Sciences of Ukraine, 2009.– 282 p.

3. Jiang Meiling, Zhang Mingsi, Li Xiangping, et al. Research progress of super-resolution optical data storage // Opto-Electronic Engineering, 2019, 46(3): 180649.doi:10.12086/oee.2019.180649.

4. Kenneth D. Singer , Irina Shiyanovskaya Co-extruded multilayer optical data storage media (Conference Presentation) // Proc. SPIE 11305, Ultra-High-Definition Imaging Systems III, 1130504 (9 March 2020); https://doi.org/10.1117/12.2553638

5. Petrov V.V., Z. Le, Kryuchyn A. A., Shanoylo S.M., M. Fu, Beliak Ie.V., Manko D.Yu., Lapchuk A.S., Morozov Ye.M. Long-term storage of digital information.– / National Academy of Sciences of Ukraine, Institute for Information Recording/. – Kyiv: Akademperiodyka, 2018. – 148 p. – ISBN 978-966-360-360-5

6. Li, W., Yang, Y., Yuan, D. (2015). Reliability assurance of big data in the cloud: Cost-effective replication-based storage. Waltham, MA: Morgan Kaufman is an imprint of Elsevier..

7. Petrov,V., Beliak,I., Kryuchyn,A. Shikhovets,A.(2020) Analysis of Methods for Creating Media for Long-Term Data Storage, IEEE 2nd International Conference on Advanced Trends in Information Theory (ATIT), pp. 238-241, doi: 10.1109/ATIT50783.2020.9349267.

8. S. Lamon, Y. Wu, Q. Zhang, X.Liu, M.Gu(2021) Nanoscale optical writing through upconversion resonance energy transfer // Science Advances :Vol. 7, no. 9, eabe2209 DOI: 10.1126/sciadv.abe2209.

9. Shiono T., Matsuzaki K., & Furumiya S. (2013).Near-field recording on phase-change nanoparticles and reflective reproduction from nanoantenna utilizing plasmonic resonance for high-density optical memory // Optics Express 2013, Oct 21;21(21):25532-43. –doi:10.1364/oe.21.025532.

10. Gu,M., Zhang,Q. & Lamon, S. (2016)Nanomaterials for optical data storage // Nat. Rev. Mater.1,16070 . https://doi.org/10.1038/natrevmats.2016.70

11. Scientists develop new 700 TB capacity optical discs https://the techzone.online/scientists-develop-new-700-tb-capacity-optical-discs/

12. Project Silica https://www.microsoft.com/en-us/research/project/project-silica/

13. 2021 Data Storage Outlook Report Insight: Storage Technologies and Advancements https://spectralogic.com/2021/06/03/2021-data-storage-outlook-report-insight-storage-technologies-and-advancements-blog/

14. 14.Schmalen, L., Alvarado, A., & Rios-Muller, R. (2016). Predicting the Performance of Nonbinary Forward Error Correction in Optical Transmission Experiments. Optical Fiber Communication Conference. doi:10.1364/ofc.2016.m2a.2.

15. Lin, S., Lin, H., Ma, C., Cheng,Y., Ye, S., Lin, F., Wang,Y. (2020) High-security-level multi-dimensional optical storage medium: nanostructured glass embedded with LiGa5O8: Mn2+ with photostimulated luminescence // Light: Science & Applications, 9(1). – doi:10.1038/s41377-020-0258-3.

МЕТОДИКА ВИЯВЛЕННЯ ОБ‘ЄКТІВ КІБЕРБЕЗПЕКИ НА БАЗІ ТЕХНОЛОГІЇ OSINT

Д.В. Ланде [1,2][0000-0003-3945-1178], О.О. Пучков [1][0000-0002-8585-1044], І.Ю. Субач [1][0000-0002-9344-713X]

[1] Національний технічний університет України "Київський політехнічний інститут імені Ігоря Сікорського", Київ, Україна
[2] Інститут проблем реєстрації інформації НАН України, Київ, Україна

dwlande@gmail.com, igor_subach@ukr.net

В інформаційних ресурсах мережі Інтернет міститься багато прихованих знань. Ці знання вносяться користувачами, які у сукупності утворюють своєрідне експертне середовище. У зв’язку з цим, основна задача технологій розвідки у відкритих джерелах (OSINT) полягає у виявленні та екстрагуванні прихованих експертних знань, їх узагальненні, а також подальшої аналітичної обробки. Для цього застосовуються лінгвістичні і статистичні методи, а також методи кластерного аналізу. В роботі запропонована методика екстрагування понять із текстів повідомлень мережевих джерел, що стосуються предметної області кібербезпеки, фільтрування цих понять за статистичними ознаками, рейтингування, формування мережі їх взаємозв’язків, кластеризації і візуалізації цієї мережі. Для створення програмної реалізації запропонованих підходів використовується мова програмування Perl у середовищі ОС Linux, а також засоби програмного забезпечення для моделювання, аналізу та візуалізації графів – Gephi.

Ключові слова: OSINT, об’єкти кібербезпеки, екстрагування понять, мережа термінів, веб-ресурси.

Постановка проблеми

Фахівцям, що працюють у визначеній предметній області, зазвичай відомі її основні поняття та об’єкти. Проте, з плином часу виникають нові поняття та нові об’єкти. У сфері кібербезпеки такими об’єктами можуть бути нові види кібератак, нові хакерські угрупування, нове деструктивне програмне забезпечення, аналітичні групи тощо. Можуть з’являтися нові змістовні зв’язки між такими об’єктами, що також потребує додаткового аналізу. В окремих групах об’єктів,

наприклад, злочинних хакерських угрупуваннях, можуть зміщуватися центри та об'єкти особливої уваги фахівців із кібербезпеки. Таким чином, виникає завдання постійного моніторингу інформації у межах визначеної предметної області. Така інформація може бути представлена в мережі Інтернет, до контенту якої, зокрема, документів, розміщених на веб-сайтах, може бути застосована технологія розвідки у відкритих джерелах (OSINT) [1].

Мета цієї роботи – створення і апробування методики визначення основних об'єктів кібербезпеки і зв'язків між ними на базі аналізу змістовної складової веб-простору, а також формування, кластеризація та аналітична обробка сформованих мереж об'єктів кібербезпеки. При вирішенні цих завдань в рамках запропонованої методики має бути проаналізовано тематичну частину кириличного сегменту веб-простору і соціальних мереж щодо публікацій у сфері кібербезпеки.

Основні кроки методики

1. На першому етапі методики формується інформаційний масив релевантних тематиці документів, для чого мають використовуватись наявні системи контент-моніторингу, наприклад система Cyber Aggregator [1].

Для отримання інформаційного масиву публікацій щодо кібербезпеки необхідно опрацювати тематичний запит до такої системи, наприклад, застосовувався запит:

Кібератака | кибератака |
Кібербезпека | кибербезопасность

У результаті отримується інформаційний масив релевантних документів великого обсягу (декілька тисяч документів на місяць).

2. На другому етапі на основі лінгвістичного і статистичного аналізу здійснюється екстрагування понять із предметної області, що містяться в документах отриманого на першому кроці інформаційного масиву. Основна ідея екстрагування об'єктів полягає у тому, що на цей час більшість нових понять в повідомленнях українською, російською, білоруською мовами позначаються латиницею або кириличними літерами, але в лапках. При цьому для екстрагування іменних сутностей також застосовується словник відомих іменних сутностей об'єктів кібербезпеки, які відшукуються в інформаційному масиві. Крім того,

виявляються не кириличні короткі словосполучення в інформаційному масиві.

3. На третьому етапі здійснюється сортування відібраних понять за частотою та фільтрація цих понять фахівцем-експертом.

4. На четвертому етапі здійснюється формування мережі відібраних понять [2]. Для цього визначаються неспрямовані зв'язки між поняттями. Два поняття вважаються зв'язаними, якщо вони входять в той самий сегмент документу із відібраного інформаційного масиву.

5. На п'ятому етапі здійснюється кластеризація відібраної мережі та їх кластеризація за алгоритмом модулярності, а також візуалізація із застосуванням системи Gephi [3].

Висновки

Запропоновано методику виявлення іменних сутностей об'єктів кібербезпеки із документів, представлених у мережі Інтернет. Методика враховує приховані знання, внесені експертним мережевим середовищем.

Кластерний аналіз і візуалізація отриманої мережі об'єктів кібербезпеки дозволяють наглядно спостерігати за станом і динамікою розвитку понятійної бази предметної області.

Перелік посилань

1. D. Lande, O. Puchkov, I. Subach, M. Boliukh, D. Nahornyi OSINT investigation to detect and prevent cyber attacks and cyber security incidents // Information Technology and Security. Том 9, N 2 (2021). - С. 209-218. DOI: doi.org/10.20535/2411-1031.2021.9.2.249921.
2. Lande, D., Dmytrenko, O. Creating Directed Weighted Network of Terms Based on Analysis of Text Corpora. 2020 IEEE 2nd International Conference on System Analysis and Intelligent Computing, SAIC 2020, 2020, 9239182
3. Cherven K. Mastering Gephi Network Visualization. – Packt Publishing, 2015. – 378 p. ISBN 78-1-78398-734-4.

МОДЕЛЬ АВТЕНТИФІКАЦІЇ КОРИСТУВАЧІВ ЗА ЇХ РУКОПИСНИМ ПІДПИСОМ

Горнійчук Іван Вікторович[1[0000-0001-6754-4764]],
Євецький Віктор Леонідович[1[0000-0002-5364-8076]],
Циганок Віталій Володимирович[1,2[0000-0002-0821-4877]],
Микитюк Артем В'ячеславович[1[0000-0002-8307-9978]]
[1] Інститут спеціального зв'язку та захисту інформації Національного технічного Університету України "Київський політехнічний інститут імені Ігоря Сікорського", Київ, Україна
[2] Інститут проблем реєстрації інформації Національної академії наук України, Київ, Україна
horniychuk.ivan@gmail.com, viktorevetsky@gmail.com,
tsyganok@ipri.kiev.ua, mukuta8888@gmail.com

Розглянуто функційну та структурну моделі автентифікації користувачів за їх рукописним підписом. Розглянуто методи автентифікації користувачів за їх рукописним підписом та методику їх застосування.
Ключові слова: біометричні характеристики, автентифікація користувача, рукописний підпис, вектор біометричних ознак.

Вступ

При управлінні доступом процедури ідентифікації та автентифікації повинні гарантувати відповідність між користувачем і його ідентифікатором, що запобігає несанкціонованому доступу (НСД) до інформації. Для підвищення ефективності автентифікації доцільно використовувати додаткові характеристики, властиві тільки конкретному користувачеві. Системи, що використовують біометричні характеристики користувача практично позбавлені цих недоліків, так як ідентифікатор нерозривно пов'язаний з самим користувачем і скористатися ним несанкціоновано практично неможливо [1][2].

Як біометричну характеристику доцільно використати рукописний підпис. Він є суспільно і законно визнаною біометричною характеристикою, що використовується для автентифікації людини [3].

Основний зміст дослідження

Запропоновано модель автентифікації користувачів за їх рукописним підписом з використанням мобільних пристроїв на основі

операційної системи Android як засобів введення.

В рамках запропонованої моделі запропоновано ряд методів, серед яких:

- метод встановлення захищеного з'єднання між мобільною та стаціонарною складовою [3];
- метод формування часових характеристик рукописного підпису користувача [3][4];
- метод формування вектору біометричних характеристик рукописного підпису користувача [4];
- метод формування біометричного еталону рукописного підпису користувача [4][5];
- метод прийняття рішення про істинність користувача на основі використання міри Хеммінга [4][5].

Запропоновано функційну модель у вигляді діаграми варіантів використання. Оскільки система складається із двох складових, функційну модель доцільно зобразити для кожної із складових: розглядаючи одну складову, інша буде актором для неї і навпаки.

Запропоновано структурну модель у вигляді діаграми компонентів. Реалізація моделі потребує системного інтерфейсу взаємодії з дисплеєм в мобільній складовій, що є інтерфейсом для вхідних даних. В якості вихідних даних є рішення про істинність користувача, що використовується для авторизації цього користувача в системі (надання чи ненадання доступу до неї).

Висновки

В ході роботи розроблено модель, методи та методику автентифікації користувачів за їх рукописним підписом. На розроблені в ході дослідження програмпі застосунки отримано свідоцтва про інтелектуальну власність.

Перелік посилань

1. Скородумов А. Плюсы и минусы биометрической идентификации. *Information Security / Информационная безопасность*. 2018. № 6. С. 31–33. URL: https://www.itsec.ru/articles/plyusy-i-minusy-biometricheskoy-identifikatsii (дата обращения: 20.01.2021).

2. Irwin L. GDPR: Things to consider when processing biometric data. *IT Governance Blog En*. URL: https://www.itgovernance.eu/blog/en/gdpr-things-to-consider-when-processing-biometric-data (date of access: 08.02.2021).

3. Horniichuk I., Yevetskyi V., Kubrak V. Applying mobile devices in biometric user authentication systems. *Collection "Information technology and security"*. 2019. Vol. 7, no. 1. P. 14–24. URL: https://doi.org/10.20535/2411-1031.2019.7.1.184213 (date of access: 02.10.2022).

4. Yevetskyi V., Horniichuk I. Selection of handwritten signature dynamic indicators for user authentication. *Collection "Information Technology and Security"*. 2020. Vol. 8, no. 1. P. 19–30. URL: https://doi.org/10.20535/2411-1031.2020.8.1.217994 (date of access: 02.10.2022).

5. Mackay D. J. C. Information theory, inference & learning algorithms. Cambridge, UK: Cambridge University Press, 2003. 640 p.

ADAPTIVE SOFTWARE SYSTEM FOR INTERNATIONAL ACTIVITY LEVEL ASSESSMENT

Oleksandr V. Koval [1], Valeriy O. Kuzminykh [1], Iryna I. Husyeva [1], Xu Beibei [2] and Zhu Shiwei[2]
[1] National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, 03056, Ukraine
[2] Information Institute, Qilu University of Technology (Shandong Academy of Sciences), Jinan, 250316, China

avkovalgm@gmail.com, vakuz0202@gmail.com, iguseva@yahoo.com, xubeibei1987@163.com, zhusw@sdas.org

Keywords: microservices, adaptation, event-oriented architecture, software system

Introduction

Among the criteria for evaluating the condition of universities and scientific institutions, the international activity level assessment of researchers in state scientific organizations and teachers of higher educational institutions is becoming more and more influential. For this purpose, various ratings are developed based on certain indicators of international activity. This is especially important when investing in scientific research by the state because the importance and relevance of such research are largely determined by the international evaluation and recognition level. Today, one of the most important indicators of the international activity level of a scientific institution is the number and quality of scientific publications placed in the most prestigious international bibliographic databases.

Collection and processing of such information require significant time and resources, given its size and daily replenishment, variety of storage formats and access to information repositories. Therefore, the task of collecting, processing, consolidating and analysing the bibliographic information and assessing the international activity level of researchers in state scientific organizations and teachers of higher educational institutions is a task of processing big data and requires special approaches in the construction of specialized software systems for its solution.

System Architecture

The peculiarity of the presented approach is the creation of an adaptive software system for collecting and processing consolidated information, which accumulates data on the international activities of individual researchers and institutions with the possibility of flexibly expanding the number of information sources depending on the request.

Since the system is focused on a changing number of sources, which differ in data structures and interaction interfaces, the main requirement is the creation of a flexible architecture with the possibility of constant modification for minimal costs and the inclusion of new data sources, exclusion of old and no longer needed, and modification of relevant ones.

The system's microservice architecture [1] is a set of independently deployed services. Each of them is responsible for a certain task and can be changed, supplemented and expanded. The system's key feature is the parameters' adaptation according to the user's request [2]. This is achieved by a generalized flexible data model, event-oriented microservice architecture and orchestration of software services.

Microservice architecture with the distribution of services according to the functional purpose is the basis of the implementation. This is due to the need to expand the functionality of the system without interfering with existing components. In addition, higher flexibility is needed when scaling the system to increase the throughput. The interaction of software services within the system and the organization of information flow is based on event-oriented architecture. This is due to the need to maintain a high throughput of the system for processing and supporting a large number of users and "heavy" (in terms of computation time) requests to the system [3].

Conclusions

The developed system provides an opportunity to collect information about the international activities of individual researchers or institutions from heterogeneous (in terms of storage form and composition) sources. The set of information sources of the system can be expanded without interfering with the operation of other components. Each data processor can provide only partial data, which is supplemented by other components. The system adapts to the user's request in such a way that data is taken only from those sources that meet the request's requirements and can satisfy them.

Also, the system is characterized by load adaptability and high throughput, which is achieved due to the use of modern approaches in the organization of the software system architecture.

The system provides users with a single point of entry for requests, information and analytical information on the international activities of both individual researchers and institutions.

Reference

1. Chris Richardson. Microservices Patterns: With examples in Java. СШІА, Minning, 2018 Book 1.

2. Adam Belnar. Building Event-Driven Microservices: Leveraging Organizational Data at Scale. СШІА, O'Reilly Media, 2020, Book 1.

3. Jeff Nicoloff, Stephen Kuensley. Docker in Action. USA, Minning, 2019.

ДОКУМЕНТО-ОРІЄНТОВАНИЙ ПІДХІД ДО ПОБУДОВИ СИСТЕМ УПРАВЛІННЯ ІНФОРМАЦІЙНОЮ БЕЗПЕКОЮ

В. Мохор[1], О. Бакалинський[1], Я. Дорогий[2], В. Цуркан[1, 2]

[1]Інститут проблем моделювання в енергетиці ім. Г. Є. Пухова Національної академії наук України, Київ, Україна
[2]Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», Київ, Україна
v.mokhor@gmail.com, baov@meta.ua, argusyk@gmail.com, v.v.tsurkan@gmail.com

Розглянуто передумови впровадження систем управління інформаційною безпекою в організаціях. Показано їх визначеність організаційними структурами, політиками, процедурами, настановами. Цим обґрунтовано та сформульовано документо-орієнтований підхід до побудови систем управління інформаційною безпекою.

Keywords: інформаційна безпека, система управління інформаційною безпекою, документо-орієнтований підхід.

Вступ

Системою управління інформаційною безпекою гарантується безпечність діяльності організацій і надання ними послуг завдяки належному оцінюванні ризиків. Відповідно до [1] вона визначається організаційними структурами, політиками, процедурами, настановами та пов'язаними з ними діями стосовно збереження властивостей інформаційних активів. Тож формулювання документо-орієнтованого підходу до побудови систем управління інформаційною безпекою є актуальним завданням [1, 2].

Формулювання документо-орієнтованого підходу

Документо-орієнтований підхід визначається сукупністю способів створення текстових специфікацій і проєктних друкованих та/або електронних документів [1, 3]. Це стосується реалізування вимог до побудови систем управління інформаційною безпекою. Вони

повинні бути доступні як задокументована інформація, наприклад, про сферу застосування, оцінювання, оброблення ризиків [4].

Орієнтованість на документи спонукує до забезпечення їх узгодженості. При цьому побудова систем управління інформаційною безпекою зводиться до оцінювання часу та зусиль для їх створення. Крім того, до відповідності текстовим специфікаціям і проєктним документам [3, 4].

Висновки

Отже, документо-орієнтований підхід до побудови систем управління інформаційною безпекою передбачає встановлення їх відповідності системі документів. Це призводить до складнощів оцінювання повноти, узгодженості та взаємозв'язків між вимогами та проєктом через їх викладення у різних текстових специфікаціях.

Перелік посилань

1. ISO/IEC 27000:2018. Information technology. Security techniques. Information security management systems. Overview and vocabulary. [Valid from 2018-02-07]. URL: https://www.iso.org/standard/73906.html.
2. ISO/IEC 27001:2013. Information technology. Security techniques. Information security management systems. Requirements. [Valid from 2013-09-25; revised 2018-12-04]. URL: https://www.iso.org/standard/54534.html.
3. Friedenthal S., Moore A., Steiner R. A Practical Guide to SysML. The Systems Modeling Language. Waltham : Elsevier, 2015. 599 p.
4. ISO/IEC 27003:2017. Information technology. Security techniques. Information security management systems. Guidance. [Valid from 2017-04-12]. URL: https://www.iso.org/standard/ 63417.html.

ВИКОРИСТАННЯ ТОПОЛОГІЧНОГО ПРОСТОРУ ДЛЯ ОЦІНЮВАННЯ РІВНЯ ЗАБЕЗПЕЧЕННЯ ФУНКЦІЙ КІБЕРБЕЗПЕКИ В КРИТИЧНІЙ ІНФРАСТРУКТУРІ

Зубок Віталій Юрійович, Давидюк Андрій Вікторович
Інститут проблем моделювання в енергетиці ім. Г.Є. Пухова НАН України, м. Київ, Україна
vitaly.zubok@gmail.com, andrey19941904@gmail.com

Фреймворк з покращення кібербезпеки критичної інфраструктури (NIST CSF) визначає п’ять функцій кібербезпеки, а також визначає всередині функцій певні категорії та підкатегорії заходів кіберзахисту, які забезпечують виконання кожної функції. Множина цих заходів та всі її підмножини можна розглядати як топологічний простір, що складається з п’яти секторів, окреслених п’ятьма функціями кібербезпеки, а ті, в свою чергу поділені, на підсектори (категорії заходів кіберзахисту). Реалізація кожного заходу кіберзахисту на об’єктах відрізняється рівнем впровадження, обумовленим багатьма суб’єктивними факторами. В роботі запропонований власний підхід до аналізу цього рівня та «вимірювання» його достатності для забезпечення функцій кібербезпеки. Він полягає в оцінюванні рівнів впровадження по кожній категорії заходів кіберзахисту та використанні чисельних значень для побудови топології заходів кіберзахисту певного об’єкта в кібербезпековому топологічному просторі. Запропоновано візуалізацію топології кіберзахищеності, яка дозволяє візуально порівнювати рівні забезпечення функцій кібербезпеки, визначаючи таким чином функціонально більш зрілі та менш зрілі об’єкти кіберзахисту.

Ключові слова: критична інфраструктура, топологічний простір, топологія заходів кіберзахисту.

Вступ

Кібератаки на технологічні мережі, індустріальні системи, зокрема, об’єктів критичної інфраструктури є однією з актуальних проблем і особливо – підчас військової агресії. Для ворога цифрова інфраструктура виробничих підприємств є пріоритетною ціллю кібератак. Тому ефективне проектування цифрової інфраструктури критичних об’єктів (або – об’єктів критичної інформаційної

інфраструктури (ОКІІ), а саме забезпеченням їхнього ефективного кіберзахисту є актуальною науково-прикладною проблемою.

Наразі відбувається адаптація українського законодавства та нормативно-технічної документації з кіберзахисту критичної інфраструктури та промислових систем до найкращих світових практик. Це довготривала і багатошарова робота, яка вимагає не лише перекладу та адаптації текстів відомих документів, а й врахування контекстів використання певних термінів та історично відмінних від України методів побудови комплексу заходів кіберзахисту. Дане дослідження присвячене одній з таких проблем, а саме – спробі адаптації двох понять «рівень впровадження» та «ступінь зрілості» у відношенні до процесів забезпечення кіберзахисту.

Кібербезпека критичної інфраструктури в NIST Cybersecurity Framework

У лютому 2013 року президент США Барак Обама підписав указ № 13636 "Посилення кібербезпеки критичних інфраструктур" [1], який серед іншого зобов'язав Американський інститут зі стандартизації NIST розробити базову модель захисту критичних інфраструктур «Framework for Improving Critical Infrastructure Cybersecurity», який зазвичай називають «Cybersecurity Framework» (надалі – CSF) [2]. Остання його редакція відбулась в 2018 році.

NIST зробив CSF практичним та конкретним. По суті, фреймворк придатний до застосування в життєвому циклі системи захисту будь-якого об'єкта - від корпоративної до промислової мережі. У ньому описано архітектуру промислової мережі з погляду її побудови та її безпеки у відповідності до так званої п'ятирівневої моделі Університету Пурдью (Purdue). Основою фреймворку є його ядро, де усі заходи захисту розбиті на 5 великих блоків (функцій) - ідентифікація, захист, виявлення, реагування та відновлення (рис.1).

У кожному з п'яти блоків потім виділяються кілька категорій заходів кіберзахисту, які, в свою чергу, декомпозуються до рівня підкатегорій. Для наочного розуміння приклад класифікації заходів по категоріях та підкатегоріях для однієї з п’яти функцій кібербезпеки наведена на рис. 2.

Function Unique Identifier	Function	Category Unique Identifier	Category
ID	Identify	ID.AM	Asset Management
		ID.BE	Business Environment
		ID.GV	Governance
		ID.RA	Risk Assessment
		ID.RM	Risk Management Strategy
		ID.SC	Supply Chain Risk Management
PR	Protect	PR.AC	Identity Management and Access Control
		PR.AT	Awareness and Training
		PR.DS	Data Security
		PR.IP	Information Protection Processes and Procedures
		PR.MA	Maintenance
		PR.PT	Protective Technology
DE	Detect	DE.AE	Anomalies and Events
		DE.CM	Security Continuous Monitoring
		DE.DP	Detection Processes
RS	Respond	RS.RP	Response Planning
		RS.CO	Communications
		RS.AN	Analysis
		RS.MI	Mitigation
		RS.IM	Improvements
RC	Recover	RC.RP	Recovery Planning
		RC.IM	Improvements
		RC.CO	Communications

Рисунок 1 – Ядро Cybersecurity Framework [2]

RECOVER (RC)	**Recovery Planning (RC.RP):** Recovery processes and procedures are executed and maintained to ensure restoration of systems or assets affected by cybersecurity incidents.	**RC.RP-1:** Recovery plan is executed during or after a cybersecurity incident
	Improvements (RC.IM): Recovery planning and processes are improved by incorporating lessons learned into future activities.	**RC.IM-1:** Recovery plans incorporate lessons learned
		RC.IM-2: Recovery strategies are updated

Рисунок 2 – Категорії та пвідкатегорії заходів кіберзахисту для забезпечення функції «Відновлення» [2]

Важливим елементом застосування фреймворку є складання організацією власного профілю. Профіль за фреймворком представляє результати на основі бізнес-потреб, які організація вибрала з категорій і підкатегорій заходів кіберзахисту. Профіль можна охарактеризувати як узгодження стандартів, інструкцій і практик з ядром CSF у конкретному сценарії впровадження. Профілі можна використовувати для визначення можливостей для покращення стану кібербезпеки шляхом порівняння «Поточного» профілю (стан «як є») із «Цільовим» профілем (стан «як має бути»).

Щоб розробити профіль, організація може переглянути всі категорії та підкатегорії та, виходячи з рушійних факторів бізнесу/місії та оцінки ризиків, визначити, які є найважливішими; він може додавати категорії та підкатегорії, якщо це необхідно для усунення ризиків організації. Потім поточний профіль можна використовувати для визначення пріоритетів і вимірювання прогресу на шляху до цільового профілю, одночасно враховуючи інші бізнес-потреби, включаючи економічну ефективність та інновації. Профілі можна використовувати для проведення самооцінки та спілкування всередині організації або між організаціями.

CSF також вводить поняття рівнів впровадження заходів кіберзахисту. Рівні впровадження забезпечують контекст того, як організація розглядає ризики кібербезпеки та процеси, які існують для управління цим ризиком. Рівні описують ступінь, до якого практики управління ризиками кібербезпеки організації демонструють характеристики, визначені в Рамковій основі (наприклад, усвідомлення ризиків і загроз, повторюваність і адаптивність). Рівні характеризують практику організації в діапазоні від часткового (рівень 1) до адаптивного (рівень 4). Ці рівні відображають перехід від неформальних, реактивних реакцій до підходів, які є гнучкими та інформованими про ризики. Під час процесу вибору рівня організація повинна враховувати свою поточну практику управління ризиками, середовище загроз, правові та нормативні вимоги, цілі бізнесу/місії та організаційні обмеження.

Представлення множини заходів кіберзахисту CSF у вигляді топологічного простору

Поняття топологічного простору використовується у загальній топології. Визначення топологічного простору спирається лише на теорію множин, і є найбільш загальним поняттям математичного простору, що дозволяє визначити наступні концепції, такі як безперервність, зв'язність та конвергентність [3].

Нехай існує множина елементів X. Система T відкритих підмножин її елементів є топологією на X, що відповідає вимогам, які звуться аксіомами топології:

– об’єднання довільного сімейства підмножин L з елементів X належить T: $\forall L', L'' \in T : \left(L' \cap L'' \in T\right)$

– перетин довільного скінченного сімейства L з підмножин елементів X належить T: $\forall L', L'' \in T : \left(L' \cup L'' \in T\right)$

– сама множина X та порожня множина належать *T*:

$$\varnothing \in T,\ X \in T$$

А впорядкована пара (*X, T*) має назву *топологічного простору*.

Окремим випадком топологічного простору є простір з дискретною топологією. У дискретному топологічному просторі множина точок не є безперервною, всі точки простору в певному сенсі ізольовані одна від одної. Топологією дискретного топологічного простору (дискретною топологією) є сімейство всіх його підмножин, що відповідають аксіомам топології. Особливістю дискретної топології є те, що її базою послуговують всі підмножини множини *X*, що складаються з одного елемента.

Нехай $s \in S$ - це підкатегорія заходів зі скінченої множини підкатегорій *S*, а $m \in M$ – рівень впровадження. Назвемо кортеж *c*(*s*,*m*), що поєднує підкатегорію заходів кіберзахисту та рівень впровадження цієї заходів підкатегорії, *вектором кіберзахисту* (підкатегорія заходів вказує напрямок, а рівень впровадження вказує довжину вектора).

Множина всі можливих векторів кіберзахисту разом з усіма можливими комбінаціями векторів утворює, відповідно до визначення, топологічний простір. Кожен вектор при цьому, а також будь-яка їхня підмножина є елементами топології.

Резюмуючи, нехай існує множина *C* всіх векторів *c*. Нехай існує система елементів множини цих векторів *T*, до якої належать порожня множина, сама множина всіх векторів, та будь-які комбінації (об'єднання та перетини) з цієї множини векторів:

$$\exists T\ :\ \varnothing, C \in T;\ \forall C', C'' \in T : \left(C' \cap C'' \in T;\ C' \cup C'' \in T\right)$$

У такому випадку система *T* відповідає визначенню топології на множині векторів *C*, а пара (*C,T*) відповідає визначенню топологічного простору, де множина векторів кіберзахисту є «носієм» топології.

Тепер за допомогою представлених визначень формалізуємо поняття профілю кіберзахисту. Профіль кіберзахисту *P* є підмножиною векторів кіберзахисту, такою, що кожній підкатегорії відповідає лише один вектор. За визначенням, всі підмножини векторів належать до топології, отже профіль кіберзахисту є елементом топології, утвореної на множині векторів кіберзахисту.

Надалі використовуватимемо такі назви для зроблених визначень:

- топологічний простір кібербезпеки;
- топологія кіберзахищеності;

– вектор кіберзахисту.

Визначення рівня впровадження заходів кіберзахисту

Наказом Адміністрації Держспецзв'язку [4] затверджено методичні рекомендації щодо підвищення рівня кіберзахисту критичної інформаційної інфраструктури, в яких перекладено та адаптовано положення CSF. У тому числі, визначені такі рівні впровадження заходів кіберзахисту:

1.Частковий
2.Ризик-орієнтовний
3.Повторюваний
4.Адаптивний.

Щоб зрозуміти цю класифікацію, необхідно дослідити інші підходи та зробити порівняння, наприклад, рівнів впровадження заходів кіберзахисту рівням зрілості IT процесів, визначених Методологією COBIT 5 [5], та рівнів захищеності згідно ISA 62443-1-1 [6], представленої в табл.1.

Таблиця 1. Порівняння рівнів впровадження та зрілості

Рекомендації/NIST CSF	**ISA 62443-1-1 (ISA 62443-3-3 Annex A)**	**COBIT 5**
1.Частковий	0. Не має засобів захисту 1. Захищений від випадкового/ненавмисного порушення безпеки	0. Incomplete process (Не існуючий) 1. Performed process (Початковий)
2.Ризик-орієнтовний	2. Захищений від слабкого навмисного втручання (невеликі ресурси та знання низька мотивація, прості засоби)	2. Managed process (Повторюваний але інтуїтивний)
3.Повторюваний	3. Захищений від наполегливого навмисного втручання (задіяни суттєві ресурси, підключено професіоналів і спеціалізовані засоби, але атакуюча сторона має обмеження в можливостях та/або ресурсах та/або часі)	3. Established process (Визначений)
4.Адаптивний.	4. Захищений від цілеспрямованої	4. Predictable process

	наполегливої атаки (коли атакуюча сторона має ресурси на регулярні ворожі дії, постійно вдосконалює механізми атаки, вичікує, шукає нові вразливості).	(Керований та вимірюваний) 5. Optimising process (Оптимізо-ваний)

Досягнення рівня впровадження заходів кіберзахисту визначається за повнотою виконання цих заходів з урахуванням поточної практики щодо реалізації заходів кіберзахисту та управління ризиками кібербезпеки на ОКІ, характеристики загроз кібербезпеки, законодавчі та нормативні вимоги, комерційні та стратегічні цілі ОКІ, вимоги до кібербезпеки в ланцюзі поставок програмного/апаратного забезпечення, організаційні та інші обмеження.

Візуалізація профілю кіберзахисту

В незалежності від того, яка методика використовується для визначення рівня впровадження заходів кіберзахисту певної підкатегорії, профіль кіберзахисту може бути представлений у вигляді сукупності векторів по всіх підкатегоріях, наприклад:

$$P = \{(ID.AM-1;\ 2),(ID.AM-2;\ 4)(ID.AM-3;\ 3),....,(RC.CO-3;\ 1)\}$$

У такому вигляді профіль матиме довжину понад 100 векторів. Можна представити спрощений, менш деталізований профіль, якщо усереднити рівень впровадження заходів кіберзахисту для кожної категорії заходів кіберзахисту:

$$СЦ_{кат} = \frac{\sum V_{підкат}}{N_{кат}},$$

де

$СЦ_{кат}$ – мінімальний середній рівень впровадження заходів;

$V_{підкат}$ – рівень впровадження підкатегорії заходів, яка належить до категорії;

$N_{кат}$ – кількість підкатегорій заходів, що належать до категорії.

Спрощений профіль складатиметься всього з 23 усереднених векторів векторів і матиме вигляд:

$$P = \{(ID.AM; 2),(ID.BE; 2)(ID.GV; 3),....,(RC.CO;2)\}$$

Такий спрощений профіль можна візуалізувати за допомогою пелюсткової діаграми (рис.3).

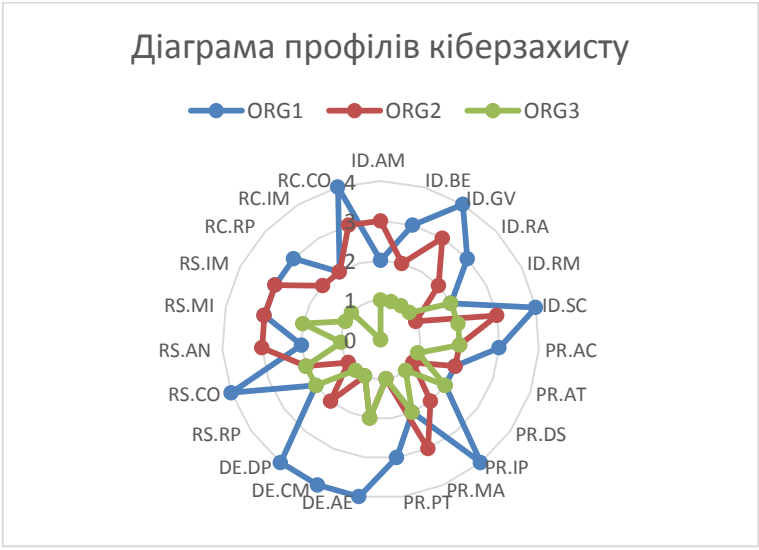

Рисунок 3 – Візуалізація спрощених профілів кіберзахисту

На рисунку 3 представлено спрощені профілі кіберзахисту трьох уявних ОКІІ в системі координат, схожій на радіальну. По кругу розташовані категорії заходів кіберзахисту. Відстань точки від центру (радіус) відповідає рівню впровадження. Очевидно, що площа фігури, оточеної певним кольором, тим більша, чим вищий середній рівень впровадження заходів кіберзахисту ОКІІ. Форма і розмір фігури можуть слугувати для оцінювання ступеню зрілості функцій кібербезпеки ОКІІ.

Висновки

Модель кібербезпеки, яка запропонована в Cybersecurity Framework та адаптована в рекомендаціях Адміністрації держспецзв'язку, дозволяє представити профіль кіберзахисту у вигляді топології. Запропонований підхід до отримання чисельної оцінки рівня

впровадження заходів кіберзахисту дає додаткові можливості для «вимірювання» достатності рівня для забезпечення функцій кібербезпеки завдяки побудові топології заходів кіберзахисту певного об'єкта в кібербезпековому топологічному просторі.

Запропоновано спосіб графічного представлення профілів кіберзахисту, який дозволяє візуально порівнювати рівні забезпечення функцій кібербезпеки, визначаючи таким чином функціонально більш зрілі та менш зрілі об'єкти кіберзахисту.

Перелік посилань

1. Foreign Policy Cyber Security Executive Order 13636. The White House. URL: https://obamawhitehouse.archives.gov/issues/foreign-policy/cybersecurity/eo-13636 (дата звернення: 11.11.2022).

2. Cybersecurity Framework. NIST. URL: https://www.nist.gov/cyberframework (дата звернення: 11.11.2022).

3. Виро О.Я., Иванов О.А., Нецветаев Н.Ю., Харламов В.М. Элементарная топология. – М.: МЦНМО. – 2010. – 352 с.

4. Наказ Адміністрації Держспецзв'язку від 06 жовтня 2021 року № 601 Про затвердження Методичних рекомендацій щодо підвищення рівня кіберзахисту критичної інформаційної інфраструктури. https://cip.gov.ua/.
URL: https://cip.gov.ua/ua/news/nakaz-administraciyi-derzhspeczv-yazku-vid-06-zhovtnya-2021-roku-601-pro-zatverdzhennya-metodichnikh-rekomendacii-shodo-pidvishennya-rivnya-kiberzakhistu-kritichnoyi-informaciinoyi-infrastrukturi (дата звернення: 11.11.2022).

5. COBIT 5 A business Framework for the Governance and Management of Enterprise IT

6. IEC/TS 62443-1-1 Industrial communication networks – Network and system security – Part 1-1: Terminology, concepts and models. Welcome to the IEC Webstore. URL: https://webstore.iec.ch/preview/info_iec62443-1-1%7Bed1.0%7Den.pdf (дата звернення: 11.11.2022).

МОНІТОРИНГ ОБ'ЄКТУ КРИТИЧНОЇ ІНФРАСТРУКТУРИ З ДОПОМОГОЮ БПЛА

Кучеров Д.П., Шмельова Т.Ф.
Національний авіаційний університет, Київ, Україна
d_kucherov@ukr.net, shmelova@ukr.net

Розглядається завдання моніторингу високовольтних ліній електромережі безпілотними літальними апаратами (БПЛА). Обговорюються алгоритми для виконання цього завдання.

Ключові слова: БПЛА, моніторинг, управління, високовольтна електромережа

Вступ

До об'єктів критичної інфраструктури належать передусім об'єкти банківської діяльності, транспорту та електроенергетики. Оскільки останні мають велике значення, їх стану слід приділяти серйозну увагу. Можливим рішенням може бути моніторинг стану високовольтних ліній електромережі за допомогою БПЛА.

Для контролю цілісності високовольтних ліній електромережі БПЛА має бути оснащений засобами навігації, радіозв'язку та відеокамерами. Політ БПЛА здійснюється на висоті, достатньої для контролю відеокамерами. Для підвищення ефективності контролю використовується мала група БПЛА у складі 3-5 літальних апаратів, маршрут якої пролягає вздовж лінії досліджуваної лінії електромережі. Група візуально шукає обриви ліній електромережі та пошкодження трансформаторних підстанцій та фіксує їх координати.

При русі за маршрутом група БПЛА повинна бути здатною виконувати нескладні маневри такі, як зліт, посадка, прямолінійний політ вздовж лінії передачі електроенергії, баражування на місці пошкодження і повернення до місця старту.

Мета статті полягає у розробленні алгоритмів управління БПЛА для забезпечення необхідної якості моніторингу та контролю відеокамерами.

Основний зміст

Груповий політ може розглядатися з лідером, без лідера та з віртуальним лідером. Основні проблеми організації групового польоту з лідером полягають у необхідності забезпечення

доступності лідера до членів команди, підтримки зв'язку та вирішення проблем польоту, пов'язаних, наприклад, з доланням перешкод [1; 2]. Лідер є технічно складним елементом групи, від якого залежить ефективність виконання цільового завдання групи БПЛА «моніторинг». Але лідер є слабкою ланкою в управлінні групою, вихід його з ладу унеможливлює виконання польотного завдання. Для подолання цієї проблеми має бути вирішена задача вибору нового лідера та передача йому управління. Дія без лідера передбачає просторове та тимчасове рознесення членів групи для організації польоту, виключення зіткнень та організації безпечної діяльності всіх членів групи. Перевагами такої організації є незалежне виконання кожним членом групи поставленого завдання, вихід з ладу одного апарату не впливає на результат кінцевої задачі, завдання виконане хоча б одним апаратом, вважається виконаним. Потребують вирішення проблеми, пов'язані з необхідністю ретельного планування маршруту та забезпеченням повного контролю за діями членів групи, особливо у разі відсутності взаємодії в критичні моменти вирішення цільового завдання «моніторинг» та неузгодженою діяльністю членів групи.

Група з віртуальним лідером, на відміну від розглянутих, може вирішувати завдання при виході з ладу одного з членів групи, здатна взаємодіяти один з одним для ефективного вирішення задачі при втраті одного з членів групи, не втрачаючи при цьому віртуального лідера. Таким чином група з віртуальним лідером компенсує недоліки, які мають місце з лідером і без лідера. Як недолік такої організації слід зазначити необхідність підтримки деякої структури щодо лідера, відповідно до якого розраховуються координати кожного члена групи. Віртуальний лідер також може бути в групі без лідера за необхідності, наприклад, для завдання «подолання перешкод». Віртуальний лідер може бути метою польоту або центром групи. При виконанні тривалих за часом завдань організація з віртуальним лідером є кращою.

В доповіді розглядається мала група БПЛА, яка вирішує завдання моніторингу об'єктів високовольтної лінії електромережі з виконанням підзавдань: рух за маршрутом; підтримка структури групи в русі або реконфігурації при подоланні перешкод; визначення пошкодження на лінії та фіксація його координат.

Висновки

Ефективним засобом рішення завдання моніторингу високовольтних ліній електромережі може стати застосування групи БПЛА.

Запропоновані алгоритми та результати моделювання підтверджують теоретичні напрацювання.

Перелік посилань

1. Kucherov D., Fu M., Kozub A. Synthesis of the laws of motion control of a UAV group with natural obstacles. Ch.7 in book “Automated Systems in the Aviation and Aerospace Industries”. Ed. T. Shmelova, Yu. Sikirda, N. Rizun, D. Kucherov, K. Dergachov– USA: IGI-Global Publ, 2019. P. 193 – 219.

2. Shmelova T., Lazorenko V., Burlaka O. Unmanned Aerial Vehicles for Smart Cities: Estimations of Urban Locality for Optimization Flights. Ch. 15 in book “Methods and Applications of Geospatial Technology in Sustainable Urbanism”. Ed. J. Tenedório, R.Estanqueiro, C.Delgado, – USA: IGI-Global Publ, 2021.– P. 444-477.

ОНТОЛОГІЧНИЙ ПІДХІД ДО КЕРУВАННЯ ДРОНАМИ НА ОСНОВІ МУЛЬТИАГЕНТНОЇ СИСТЕМИ ТА РОЄВОЇ ВЗАЄМОДІЇ

Гладун Анатолій[0000-0002-4133-8169], Хала Катерина[0000-0002-9477-970X]
Міжнародний науково-навчальний центр інформаційних технологій та систем НАНУ та МОНУ, Київ, Україна

glanat@yahoo.com, cecerongreat@ukr.net

Одним із великих викликів сьогодення є застосування роїв безпілотних літальних апаратів (БПЛА) на основі мультиагентних систем (МАС), щоб забезпечити переваги децентралізованого керування, які полягають у сумісному виконанні дронами поставленого завдання, живучості системи, доступності і масштабованості у системах захисту критичних інфраструктур.

Для обміну даними між дронами необхідна інтероперабельна інформація, яка забезпечує однозначний контент, а також формалізовані знання про рольові функції усіх дронів та умови співпраці між ними, умови переназначення ролей при збоях або втратах, ідентифікація інших БПЛА. Ці завдання можуть бути покладені на онтологічну модель, яка формалізує певні знання про політику дій системи у різних складних ситуаціях і допомагає МАС краще виконати поставлені завдання.
Ключові слова: онтологія, представлення знань, ройовий інтелект, дрони, БПЛА, ієрархічна структура керування, адаптивна онтологія.

Вступ

Сьогодні БПЛА, або дрони, привертають велику увагу різних сферах застосування від військових до цивільних, оскільки вони можуть виконувати завдання великої складності. Сфери застосування БПЛА з МАС і роєвим інтелектом сьогодні охоплюють від інспекції висотних споруд і мостів (США, Німеччина) [1,2], пошуково-рятувальних робіт (Бразилія, США) до військового для виконання воєнних операцій та захисту критичних інфраструктур.

Застосування методів штучного інтелекту та машинного навчання, дозволяє розробити високонадійні системи БПЛА з різноманітними функціями, які мають традиційні роботи, від комп'ютерного зору, розпізнавання звуків до логічного виведення нових знань на основі тих, які є в наявності [3].

Для забезпечення функціонування МАС використовується група онтологій, що визначають контрольований словник термінів, дають змогу визначати функцію, контекст і ситуації, необхідні для обміну семантичною інформацією для предметної області (ПрО) в МАС або між оператором та агентом.

Онтологічний підхід до забезпечення взаємодії рою БПЛА

Онтології забезпечують формальну специфікацію концептуалізації в чітко визначений і однозначний спосіб, таким чином, вони сприяють обміну та повторному використанню знань. Онтології можна використовувати для міркування про об'єкти та їхні атрибути у визначеному домені [5]. Системи засновані на роях складаються із множини агентів, що взаємодіють між собою та зовнішнім середовищем застосовують для створення плану руху БПЛА, розподілу робіт між агентами і працюють у взаємодію з онтологією ПрО.

Фундаментальними для функціонування БПЛА на основі чотири типи класів онтологічної моделі:

1) класи *інформаційних об'єктів*, такі як інструкції польоту та розвідки, політики підтримання живучості та надійності системи та завдання від операторів або запрограмовані в БПЛА;
2) класи *агентів*, такі як оператори БПЛА та автономні БПЛА, які надсилають, отримують і виконують завдання;
3) класи *процесів*, такі як процеси польоту, зв'язку та спостереження, передбачені цими директивами та виконані цими агентами; і
4) класи *ролей*, такі як командир, оператор, ведучий і приманка, які призначаються учасникам цих процесів і диктують приписи, за які відповідає кожен учасник.

Онтологічно визначені сценарії є зрозумілими та можуть використовуватися як операторами, так і агентами, що створює спільне розуміння як контенту сценарію, так і сутностей, про які йдеться в контенті.

Висновки

Однією з визначальних характеристик спільних дій є те, що учасники співпраці мають намір досягти однієї мети та досягти її

разом як група. Онтологія розширює дані в базі знань за допомогою узгодженої структури відношень і чітко визначених термінів для формування логічних висновки з даних, позначених термінами з цієї онтології.

Перелік посилань

1. G. Chmaj and H. Selvaraj, Distributed processing applications for UAV/drones: a survey," in Progress in Systems Engine. Springer, 2015.
2. Brambilla, M., et al., ‘Swarm robotics: a review from the swarm engineering perspective’, Swarm Intelligence, vol. 7 (2013), pp. 2–3.
3. Гладун А.Я. Рогушина Ю.В. Семантичні технології: принципи та практики .- К.: Вид-во «Універсаріум». 319 с.
4. Гриценко В.И., Гладун А.Я., Рогушина Ю.В Семантическое распознавание информационных объектов на основе онтологического представления знаний о предметной области в задачах интелектуального управления. Кібернетика і обчислювальна техніка. 2014, №4 (178). С. 5-22.
5. Гладун А.Я., Рогушина Ю.В. Патент України №113890 «Спосіб персонального пошуку інформації», 2015.

ФОРМУВАННЯ МЕРЕЖІ ВЧЕНИХ У СФЕРІ КІБЕРБЕЗПЕКИ

Д.В. Ланде [1,2][0000-0003-3945-1178](dwlande@gmail.com),
А.О Снарський [1,2][0000-0002-4468-4542](asnarskii@gmail.com),
О.О. Дмитренко [1,2][0000-0001-8501-5313](dmytrenko.o@gmail.com),
Лі Чень [3,4], Лі Сяньї [3,4], Го Цзяньпін [3,4] (jianpingdou@126.com)

[1] Національний технічний університет України "Київський політехнічний інститут імені Ігоря Сікорського", Київ, Україна
[2] Інститут проблем реєстрації інформації НАН України, Київ, Україна
[3] Цілу Технологічний Університет Академії наук провінції Шаньдун, Цзинань, КНР
[4] Інститут інформаційних досліджень Академії наук провінції Шаньдун, Цзинань, КНР

У цій роботі розглядаються мережі вчених, в яких враховуються не тільки відносини співавторства, але й тематична близькість наукових інтересів. Особливість наведеного підходу полягає в урахуванні дескрипторів тематик, які приписують собі окремі автори. Змістовна кореляція тематик визначає вагу зв'язку між вченими у мережах, що розглядаються. При реалізації цього підходу застосовується спеціальний алгоритм сканування ресурсів наукометричного сервісу для одержання репрезентативного набору авторів або співавторів як основи (вузлів) мережі. Для створення програмної реалізації запропонованих підходів і методів використовується мова програмування Perl, а також засоби програмного забезпечення для аналізу та візуалізації графів Gephi.

Ключові слова: Мережа Вчених, Наукометричний Сервіс, Зондування Інформаційної Мережі, Дескриптори Тематик, Кластерний Аналіз, Кібербезпека.

Постановка проблеми

На цей час актуальною є задача вибору експертних груп, прогнозування [1] спільної роботи вчених в різних галузях, зокрема, в сфері кібербезпеки. Якщо враховувати відношення співавторства та/або спільних інтересів різних вчених, то можна сформувати мережі, які можна використовувати для рішення цієї задачі.

Метою даної роботи є представлення нового підходи до побудови мережі вчених шляхом зондування наукометричних сервісів.

Пропонується методика формування і подальшого дослідження мережі вчених шляхом цілеспрямованого зондування наукометричних мереж. Під зондуванням мереж розумітимемо вибірку невеликого обсягу найважливішого змісту з великих мереж, які з технологічних причин не підлягають повному скануванню [2]. Існують різні методики формування мереж вчених, зокрема, мереж співавторства.

У багатьох сучасних дослідженнях мереж застосовуються механізми їх моніторингу, після чого робляться висновки щодо топології таких мереж. У роботі [3] показано, що цей підхід є хибним. Отримані в результаті моніторингу образи первинних мереж, частково відображаючи їх властивості, найчастіше суттєво відрізняються. Властивості цих образів суттєво залежать від алгоритмів, за якими здійснюється моніторинг.

Алгоритм

Зондування опорної інформаційної мережі здійснюється за таким алгоритмом:

Крок 1. Вибирається базовий дескриптор, який визначається як базовий для зондування (спочатку, у найпростішому випадку вибирається один вузол – cyber security).

Крок 2. Для обраного дескриптора/дескрипторів засобами наукометричного сервісу є всі вчені – автори, які приписали собі ці дескриптори. Автор розміщуються у відсортованому порядку – на початку показуються автори з найбільшими цитуваннями. Для побудови мережі шляхом зондування розглядаються автори зі значенням цитування не менше 10 000.

Крок 3. Складається перелік дескрипторів від знайдених авторів, що відповідають первинній темі cyber security. Зокрема, на перших сторінках авторів по першому дескриптору знаходяться такі дескриптори, як access control models architectures, secure cloud and IoT computing, Wireless Security, Network Security, Intrusion Detection, Deception Detection, Cloud Forensics access control тощо.

Крок 4. Для кожного з авторів розглядаються їх співавтори також із значенням цитування, не менше 10 000. З цих співавторів як вузли мережі розглядаються ті вчені, дескриптори яких близькі до первинної тематики cyber security. Таким чином, були знайдені такі дескриптори, як Network Security, Computer Security, Data breach analysis, Cyber crime investigation, IT Security, Security and Privacy тощо.

Крок 5. Для всіх обраних дескрипторів вибираються автори, яким ці дескриптори приписані. Якщо список авторів зі значенням цитування, більшим за 10 000, для всіх вибраних дескрипторів вичерпано, то процес завершується. В іншому випадку здійснюється перехід до кроку 2.

Вагове значенням зв'язків, яке ставиться у відповідність між вузлами-авторами в мережі, дорівнює загальній кількості спільних дескрипторів. Якщо існує відношення співавторства, то до ваги зв'язків між вузлами додається деяка константа (експертна оцінка – число 5).

На Рис. 1 наведено фрагмент мережі вчених у галузі кібербезпеки. Мережа має високу зв'язність та явно виражені кластери.

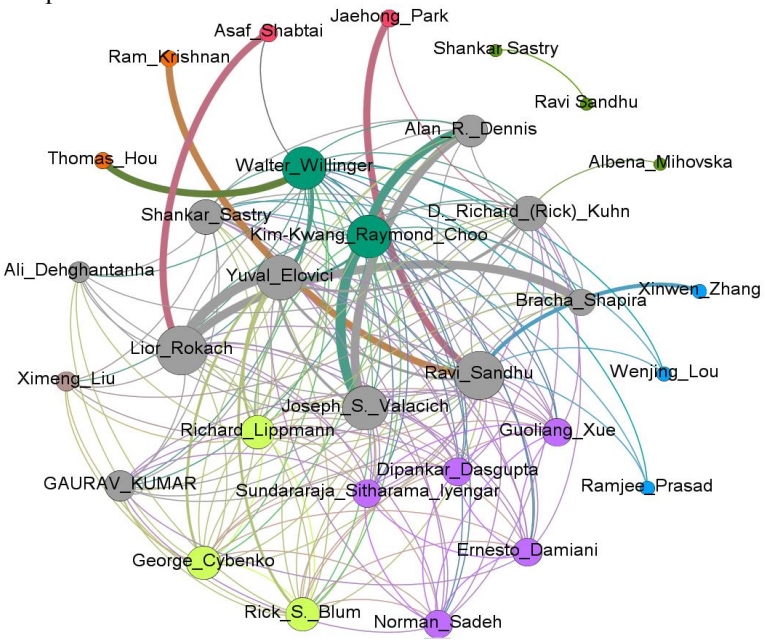

Рисунок 1 – Фрагмент мережі вчених

Висновки

Запропоновано та реалізовано підхід до формування мереж вчених у рамках предметної галузі кібербезпеки, обмежувальними

елементами якого є деякі маркери знань (дескриптори), заздалегідь задані вченими як учасниками проекту Google Scholar.

Слід зазначити принципову відмінність запропонованої моделі автоматичного формування мереж від існуючих, які базуються на безпосередній участі експертів. У межах цього дослідження для побудови мережі в якості числових значень відношень між ученими використовується змістовна кореляція базових дескрипторів. Програма зондування мережі використовує знання, закладені авторами, таким чином експертне середовище у цьому випадку суттєво розширюється.

Модель застосовувалась для сфери кібербезпеки в рамках сервісу Google Scholar, але запропонований підхід можна використовувати і для інших наукових галузей, або для інших наукометричних сервісів.

Літературні джерела

1. Lande D., Fu M., Guo W., Balagura I., Gorbov I. & Yang H. Link prediction of scientific collaboration networks based on information retrieval. World Wide Web : Internet and Web Information Systems. – Iss. 23, pp. 2239-2257 (2020). DOI: doi.org/10.1007/s11280-019-00768-9.

2. Ландэ Д.В., Балагура И.В., Андрущенко В.Б. Построение сетей соавторства по данным сервиса Google Scholar Citations // Открытые семантические технологии проектирования интеллектуальных систем (OSTIS-2016): материалы VI междунар. науч.-техн. конф. (Минск 18-20 февраля 2016 года). – Минск: БГУИР, 2016. – С. 233-237.

3. Lande D., Dmytrenko O. Research of Topological Properties of Network Reflections Obtained Using Different Algorithms for Scanning Initial Networks. In: Shkarlet S. et al. (eds) Mathematical Modeling and Simulation of Systems. MODS 2021. Lecture Notes in Networks and Systems, Vol. 344. Springer, Cham. (2022) https://doi.org/10.1007/978-3-030-89902-8_26.

ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ ВИЯВЛЕННЯ КІБЕРІНЦИДЕНТІВ SIEM-СИСТЕМАМИ В ІНФОРМАЦІЙНО-КОМУНІКАЦІЙНИХ СИСТЕМАХ

Субач Ігор Юрійович[1,2] [0000-0002-9344-713X],
Могилевич Дмитро Ісакович[1] [0000-0002-4323-0709],
Микитюк Артем В'ячеславович[1] [0000-0002-8307-9978],
Кубрак Володимир Олександрович[1] [0000-0001-8877-5289],
Фесьоха Віталій Вікторович[2] [0000-0001-6612-1970].

[1] Інститут спеціального зв'язку та захисту інформації Національного технічного Університету України "Київський політехнічний інститут імені Ігоря Сікорського", Київ, Україна
[2] Військовий інститут телекомунікацій та інформатизації імені Героїв Крут, Київ, Україна

igor_subach@ukr.net, mogilev11@ukr.net, mukuta8888@gmail.com, volodymir.kubrak@ukr.net, vitaliifesokha@gmail.com

Розглянуто питання виявлення кіберінцидентів, які виникають в ході функціонування інформаційно-комунікаційних систем (ІКС), на основі застосування в контурі кіберзахисту SIEM-системи, як основи для її побудови. Зроблено висновок про необхідність застосування в ній новітніх інформаційних технологій, реалізованих на основі моделей та методів штучного інтелекту, інтелектуального аналізу даних, обробки великих обсягів даних та машинного навчання. Запропоновано модель формалізації причинно-наслідкових зв'язків між змінними "ознаки кіберінцидентів-типи кіберінцидентів", на основі опису цих зв'язків природною мовою із застосуванням теорії нечітких множин та лінгвістичних змінних.

Ключові слова: кібербезпека, кіберінцидент, SIEM-система, нечіткі множини, нечіткі продукційні правила.

Вступ

Основою побудови ефективної системи кіберзахисту має бути застосування проактивної SIEM-системи (Security Information and Event Management) – систем управління інформацією та подіями безпеки [1].

Ефективність функціонування системи, що розроблена на базі запропонованої моделі, суттєво залежить від застосування в ній новітніх

інформаційних технологій, реалізованих на основі моделей та методів штучного інтелекту (ШІ), інтелектуального аналізу даних (Data Mining – DM), обробки великих обсягів даних (Big Data), машинного навчання (Machine Learning - ML) та інших. Дані технології дозволяють мінімізувати участь людини під час вирішення задачі реагування на кіберінциденти, які відбуваються в ході функціонування ІКС, тим самим підвищуючи оперативність та обґрунтованість рішень, які вона приймає [2].

Модель ідентифікації кіберінцидентів SIEM-системою

Загальновідомо, що ефективність застосування будь-якої інтелектуальної системи, в основному, залежить від потужності її бази знань. Проте, проведений аналіз показує, що найбільш розповсюдженим методами виявлення кіберінцидентів SIEM-системами є правило-орієнтовані методи, які ґрунтуються на класичних продукційних правилах. Відповідно, для представлення знань в базі знань SIEM-системи застосовується продукційна модель надання знань.

Проте, дана група методів, в умовах неповноти та неточності інформації про кіберінциденти, які виникають в ході функціонування ІКС, не завжди дають очікуваний результат, а, відповідно, застосування їх є неефективним.

Для усунення даного недоліку, в роботі [3] запропоновано застосування моделей та методів, що ґрунтуються на теорії нечітких множин на нечіткого логічного виводу.

Відповідно до цього, модель виявлення (розпізнавання) кіберінцидентів SIEM-системою може бути представленою у виді [3]:

$$MF = \langle KF, O_i, RF, C \rangle, \quad (1)$$

де KF – нечіткий класифікатор;

$RF = \{RF_i\}$ – множина нечітких правил розпізнавання кіберінцидентів:

$$RF_1 : (K, O_v), RF_2 : (K, O_v), \ldots, RF_l : (K, O_v) \to C .$$

Ґрунтуючись на [4] задача розпізнавання кіберінцидентів може розглядатися як задача їхньої ідентифікації, а рішення її полягає у знаходженні відображення:

$$O^* = (o_1^*, o_2^*, \ldots, o_n^*) \to c_j \in C = (c_1, c_2, \ldots, c_m), \quad (2)$$

де O^* – множина ознак кіберінцидента; C – множина можливих кіберінцидентів.

Для вирішення на практиці задачі (2) найбільш широкого розповсюдження набули методи параметричної ідентифікації, наприклад, метод найменших квадратів, метод максимальної правдоподібності, метод середніх нев'язок, метод стахостичної апроксимації та ін. [5].

Основними недоліками цих методів, які ускладнюють їх використання є наступні:

- моделі об'єктів типу "входи-вихід", як правило не мають чіткої інтерпретації;
- відсутня можливість роботи з вхідними та вихідними змінними якісного типу;
- відсутня можливість використання досвіду експерта про структуру об'єкту, що формалізується у вигляді логічних висловлювань типу "ЯКЩО-ТО".

Усунення наведених недоліків полягає у застосуванні іншого підходу, який ґрунтується на моделях та методах інженерії знань, зокрема, моделях надання знань у базі знань про ознаки кіберінцидентів (ОКІ), що виникають в ході функціонування ІКС та типи кіберінцидентів (ТКІ), для їхньої подальшої ідентифікації на основі збору та обробки експертної інформації за допомогою теорії нечітких множин.

Основною ідею цього підходу є формалізація причинно-наслідкових зв'язків між змінними "ОКІ-ТКІ", шляхом опису цих зв'язків природною мовою із застосуванням теорії нечітких множин та лінгвістичних змінних. Це дозволяє здійснити математичну формалізацію природно-мовних висловлювань щодо їхнього застосування для вирішення задачі ідентифікації кіберінцидентів у постановці задачі (2).

Висновки

Запропонована модель ідентифікації кіберінцидентів SIEM-системою, на відміну від існуючих, застосовує нечіткі правила для розпізнавання кіберінцидентів, що виникають в ході функціонування інформаційно-комунікаційної системи. Це, у свою чергу, дозволяє усунути неповноту та неточність інформації про кіберінциденти у процесі їх ідентифікації.

Перелік посилань

1. І. Субач, В. Кубрак, А. Микитюк. Архітектура та функціональна модель перспективної проактивної інтелектуальної системи SIEM-системи для кберзахисту об'єктів критичної

інфраструктури. *Information Technology and Security 2019.* Vol 7. Iss. 2. P. 208–215. DOI: 10.20535/2411-1031.2019.7.2.190570.

2. І. Субач, В. Фесьоха, Н. Фесьоха. Аналіз існуючих рішень запобігання вторгненням в інформаційно-телекомунікаційні мережі, відкритих на основі загальнодоступних ліцензій. *Information Technology and Security* 2017. Vol 5. Iss. 1. P. 29–41. DOI: 10.20535/2411-1031.2017.5.1.120554.

3. Ihor Subach, Artem Mykytiuk, Volodymyr Kubrak, Stanislav Korotayev. Rule-oriented method of cyber incidents detection by SIEM based on fuzzy logical inference. *Information Technologies and Security 2021*. Aachen, Germany: CEUR Workshop Proceedings. Vol. 2859. P. 210-219. DOI: 10.5281/zenodo.7123656.

4. А.П. Ротштейн, Интеллектуальные технологии идентификации: нечеткие множества, генетические алгоритмы, нейронные сети. – Винниця: УНИВЕРСУМ, 1999.

5. Зайченко Ю.П. Исследование операций: нечеткая оптимизация. – Киев: Вища школа, 1991.

РОЗВІДУВАЛЬНИЙ АНАЛІЗ ДАНИХ З ВИКОРИСТАННЯМ ІНСТРУМЕНТАЛЬНИХ ЗАСОБІВ НАДБУДОВИ ASPE

А.І. Кузьмичов
Інститут проблем реєстрації інформації НАН України, Київ, Україна
akuzmychov@gmail.com

На конкретних прикладах з використанням надбудови Excel ASPE показано, як за допомогою інструментів розвідувального аналізу наборів даних великих обсягів здійснюються процедури: зменшення розмірів, кластеризації, перетворень типів даних, підсилений вибірковий аналіз, відновлення та майнінгу текстів.
Ключові слова: Розвідувальний аналіз даних, датамайнінг, машинне навчання, Analytic Solver Data Mining, ASPE.

Датамайнінг та Розвідувальний аналіз даних (Data Mining & Exploratory Data Analysis) – міждисциплінарна методологія на тлі «великих даних», новітні інформаційні технології і процедури, зорієнтовані на виявлення вад в наборах табличних даних великих обсягів про об'єкти, згідно поставлених цілей. Цей аналіз стосується потенційно чи реально пошкоджених даних, де неодмінна передобробка і перетворення спеціальними інструментальними засобами з метою подальшого вилучення корисної і цінної інформації про об'єкти, представленими цими даними, в лаконічних формах, як-от візуалізацією. Він охоплює різноманітні підходи експериментальної техніки і засоби досліджень, використовується у різних сферах практики, зазвичай, для передбачувальних цілей засобами машинного навчання.

Великі дані, зобов’язані успіхам IT у вигляді автоматизованої реєстрації, передачі, зберігання, накопичення та перетворення даних різного формату і походження, висунули специфічну проблематику щодо їх продуктивного використання – mining of data: за буквальним значенням mine (копальня) мова йде про видобування корисної інформації, якою зазвичай цікавляться: розвідники, дослідники, геологи, слідчі тощо, тобто, «копати, якнайглибше, аби докопатися» за визначеним інтересом.

Основні процеси, задачі, моделі, методи, алгоритми

1. **Підсилений вибірковий аналіз** (Oversampling): у наборі даних із тисяч записів і десятків змінних у дуже малій кількості представлені записи, що сильно відрізняються від інших, мета: так організувати вибірки (тренувальну, валідаційну) випадковим відбором, аби повністю зберегти ці записи, які небезпідставно названо «успішними».
2. **Відбір впливових змінних/властивостей** (Feature Selection, Principal Components): для розуміння наявних даних серед десятків змінних треба визначити невелику їх кількість, що суттєво впливають на статистичні оцінки класифікації та передбачення, із можливістю візуалізації стиснутого набору даних.
3. **Виправлення/поновлення відсутніх значень** (Missing Data Handling): зазвичай набір даних, що «майнять», «сирий» чи пошкоджений, випадково чи штучно, і виявлені пропуски, помилки, повтори – джерело шуканої інформації, для цього існує відповідний інструментар.
4. **Перетворення типів даних** (Transform Continuous & Categorical Data) здійснюється задля розв'язання задач класифікації, передбачення, аналізу часових рядів та пошуку правил асоціації, де існують відповідні вимоги до типів даних.
5. **Кластеризація** (K-Means & Hierarchical Clustering) суттєво сприяє розумінню набору даних, можливо, невідомого походження чи пошкодженого в результаті шкідливих дій, де результатом є розділення багатовимірних записів на потрібне/рекомендоване число кластерів та їх візуалізація.
6. **Класифікація** (Discriminant Analysis, Logistic Regression, k-Nearest Neighbors, Classification Tree, Naive Bayes, Neural Network, інструменти Ensemble: bagging, boosting, random trees, Find Best). Рекомендована колекція кращих інструментів статистичного аналізу та машинного навчання для побудови моделей та визначення за ними класів записів нових даних.
7. **Передбачення** (Linear Regression, k-Nearest Neighbors, Regression Tree, Neural Network, інструменти Ensemble: bagging, boosting, random trees, Find Best, Association Rules). Рекомендована колекція кращих інструментів статистичного аналізу та машинного навчання для побудови прогнозних моделей та визначення за ними значень цільових змінних у записах нових даних.

Суттєво, що наведений інструментар представлений у доступному й знайомому інформаційному середовищі Excel, не потребує будь-якого програмування, для роботи достатньо шкільного освітнього рівня в математиці/інформатиці. В той же час, професійно-орієнтована надбудова ASP та її освітня версія ASPE підтримують використання будь-яких джерел даних та операційних платформ: Web/Win/Mac.

Перелік посилань

1. Додонов О. Г., Кузьмичов А. І. Датамайнінг в Excel. Розвідувальний аналіз даних з використанням надбудови Analytic Solver Data Mining: Київ: Ліра-К, 2022. – 250 с.
2. Ragsdale C. Spreadsheet Modeling and Decision Analysis. A Practical Introduction to Business Analytics, 9-ed. Cengage, 2022. – 908 p.
3. Myatt G., Johnson W. Making Sense of Data. A Practical Guide to Exploratory Data Analysis and Data Mining, 2-ed. – Wiley, 2014. 248 p.
4. Тьюки Дж. Анализ результатов наблюдений. Разведочный анализ. Пер. с англ. М.: Мир, 1981. – 694 с.

ОЦІНКА ВРАЗЛИВОСТІ ПРОТОКОЛУ RADIUS НА ОСНОВІ РОЗШИРЕНОЇ АВТОМАТНОЇ МОДЕЛІ

Гальчинський Л.Ю., НТУУ «КПІ», Середа А.С.
НТУУ «КПІ», ім.Ігоря Сікорського, Київ,

hleonid@gmail.com, ars.sereda2016@gmail.com

Ключові слова: протокол RADIUS, вразливості, скінченний автомат

Вступ

Бурхливий розвиток Інтернету породив широке розповсюдження корпоративних мереж, в яких співробітники працюють у віддаленому режимі. Разом з з'являється і нові небезпеки. Зазвичай віддалений доступ загалом надається за IP-адресою працівника. Однак існує вірогідність, що неавторизований користувач може ввійти в систему тією ж самою IP-адресою. Через це, має бути запроваджена додаткова автентифікація. Зрозуміло, що пакети, якими обмінюються в такій мережі, мають бути зашифровані так, щоб інші особи не змогли отримати доступ до конфіденційної інформації. Саме для реалізації цих завдань і забезпечення додаткового рівня безпеки використовується протоколи під загальною назвою AAA, представником яких є протокол RADIUS[1]. Проте практика показала певні недоліки цього протоколу, які породжують вразливості мережі в такому захищеному режимі.

Метою даної роботи було створення розширеної автоматної моделі протоколу для пошуку потенційно слабких станів, які можуть спричинити вразливості,

Пакет протоколу RADIUS та його потенційні вразливості

Протокол RADIUS бере за основу пакети формату UDP. Кожен пакет RADIUS складається умовно з трьох секцій:

- секція ідентифікаторів, за домогою якої в протоколі визначаються режими роботи та реалізуються запити і відповіді;
- секція автентифікації для автентифікації відповіді від сервера RADIUS та для застосування алгоритму приховування пароля;

- та секція атрибутів, в якій містяться дані про конкретний тип аутентифікації, авторизації, інформацію про облік і конфігурацію для пакетів запиту та відповіді.

Пристрій, який функціонує як клієнт RADIUS, отримує інформацію про користувача, включаючи ім'я користувача та пароль, і надсилає цю інформацію на сервер RADIUS. Потім сервер RADIUS автентифікує користувачів відповідно до інформації, після чого виконує авторизацію та облік для користувачів, якщо перевірки пройдені успішно, або відхиляє отриманий запит. Повідомленнями аутентифікації між сервером RADIUS і клієнтами RADIUS обмінюються за допомогою спільного ключа. Пакет RADIUS має 16-октетне поле Authenticator, яке містить дані цифрового підпису всього пакета. Дані підпису обчислюються за допомогою алгоритму MD5 і спільного ключа. Приймач пакетів RADIUS має перевірити, чи правильний підпис, і відкидає пакет, якщо підпис неправильний. Цей механізм покращує безпеку обміну повідомленнями між клієнтами RADIUS і сервером RADIUS. Крім того, паролі користувачів, що містяться в пакетах RADIUS, шифруються за допомогою спільних ключів перед передаванням пакетів, щоб запобігти крадіжці паролів користувачів під час передачі в незахищеній мережі. Сервер отримує пакет RADIUS Access-Request і перевіряє, чи володіє він загальним з клієнтом секретом. Якщо сервер не володіє спільним секретом з клієнтом, запит відкидається без відповіді.

Розширена автоматна модель протоколу RADIUS

Для моделювання складної поведінки запропонованого протоколу розглянуто модель скінченного автомата FSM, в її розширеній версії EFSM. Підставою для вибору моделі скінченного автомата є те, що процес авторизації в RADIUS має скінченний життєвий цикл, в ході якого елементи мережі обмінюються повідомлення переходячи зі стану в стан, починаючи від запиту користувача на авторизацію в системі до вдалої авторизації, або відмови сервера. EFSM[2] допомагає краще зрозуміти семантику моделі системи, коли кількість станів і переходів збільшується.

Розширений автомат кінцевих станів (EFSM) M – це кортеж (S, T, E, V), де: S – набір станів, T – набір переходів, E – набір подій, V – набір змінних. Переходи мають вихідний стан source(t) ∈ S, цільовий стан target(t) ∈ S і мітку lbl(t). Мітки переходів мають вигляд e1[c]/a, де e1∈E, c – умова, а a - послідовність дій.

Висновки

Така модель дозволяє створити скінченний автомат з умовними переходами. Програмна реалізація описаної моделі дозволила виявити стани, як потенційні джерела вразливостей протоколу RADIUS.

Джерела

1. RFC 2865. Remote Authentication Dial In User Service (RADIUS). [Електронний ресурс] / C.Rigney, A. Rubens, W. Simpson, S. Willens. – 2000. – Режим доступу до ресурсу: https://tools.ietf.org/html/rfc2865.

2. Control dependence for extended finite state machines. Fundamental Approaches to Software Engineering / K.Androutsopoulos, D. Clark, M. Harman, L. Tratt., 2009. – (Springer).

ОЦІНЮВАННЯ РИЗИКІВ СКЛАДНИХ СИСТЕМ З ВИКОРИСТАННЯ МЕТОДІВ Q-АНАЛІЗУ

Полуциганова В.І., Смирнов С.А.
Національних технічний університет України
«Київський політехнічний інститут ім. Ігоря Сікорського», Київ, Україна
vipol-ipt@lll.kpi.ua, sergsmr@gmail.com

В роботі розглянуто метод оцінювання ризиків для структурно та функціонально складних, в яких вразливості, а як наслідок і збитки можуть реалізовуватись сумісно і залежати один від одного. Для таких випадків модифіковано класичну формулу оцінювання ризиків. За допомогою Q-аналізу описано структурні залежності між вразливостями, що дало більш чітку оцінку збиткам.
Ключові слова: ризик, вразливості, Q-аналіз, симплекційний комплекс

Вступ

В роботі представлено метод оцінювання ризиків, який базується на тому, що вразливості системі можуть бути залежними та одночасно реалізовуватися. Через це виникають складні залежності між ними. Щоб їх описати та проаналізувати доречно використовувати [3]. Система вразливостей тоді описується як симплекційний комплекс, взаємозв'язки обраховуються та оцінюються за допомогою Q-аналізу. Це дає більш чітку картину як будувати систему безпеки в реальній системі та краще розуміти проблеми, які з цим пов'язані. Для таких систем в даній роботі представлений метод для обрахунку ризиків з врахуванням взаємозалежними вразливостями, а як наслідок і збитками.

Метод обрахунку ризиків для складних систем

Для того щоб провести аналог симплексійного комплексу для вразливостей в кібербезпеці потрібно шукати не тривіальні зв'язки між ними. Наприклад, наша система має складну структуру. Кожна з її підсистем має свій набір вразливостей $A_1, A_2, \dots, A_n$[1]. При цьому кожна з них може опосередковано впливати на вразливості з інших підсистем. Тоді виникають залежності схожі на залежності в сиплекційному комплексі. Якщо вразливість ніяк не впливає і не залежить від інших виникає симплекс точка. Якщо дві вразливості

взаємопов'язані то виникає ребро між симплексами. Якщо 3 вразливості залежні виникає грань, якщо 4 – тетраедр і так далі [4]. Якщо ж вразливості з однієї з підсистем впливають на вразливості з іншої або при реалізації однієї з вразливостей виникають вразливості в іншій, то може виникати складні залежності в комплексі. Виходить, що для системи деякі вразливості можуть запускати каскадні залежності, що призводить до порушення цілісності. Для таких випадків слід прораховувати ризики, але в класичну формулу для ризиків потрібно внести зміни, адже вона базується на припущені, що всі втрати та ймовірності їх настання незалежні[2]. Але не лише ризик можна прораховувати у цьому випадку. Окрім цього, можливо провести Q-аналіз на виявлення складних зв'язків між вразливостями. Тоді можна поліпшити попередню формулу так, щоб не враховувати деякий симплекс в комплексі (іншими словами вразливість в системі вразливостей) декілька разів. В доповіді розглянуто приклад складної функціональної залежності між збитками пов'язаними з несправностями авто. Показано як впливає структурна залежність слабких місць, та як їх виявити за допомогою Q-аналізу. Прораховані збитки в залежності від сценарію реалізації несправностей.

Висновки

Оцінювання ризиків важливий етап в побудові безпеки будь-якої системи. При цьому вразливості та збитки можуть бути структурно взаємозалежними, що ускладнює прорахунки. За допомогою Q-аналізу, в роботі, було побудовано та оцінено структуру вразливостей. Наведено модифікований метод обрахунку ризиків з врахуванням функіональних конфігурацій.

Посилання

1. Качинський А. Б. Безпека складних систем / – Київ: Юстон, 2017. – 498 с.

2. Вишняков Я. Д., Радаев Н. Н. Общая теория рисков. –2 изд. – М. : Академия, 2008.—368 с.

3. Atkin R. H. "Mathematical structure in human affairs", Heinemann Educational Books, (1973); 143. doi: 10.1137/1018064.

4. Beaumont J.R., Gatrell A.C. "An introduction to Q-analysis". Catmog 34, 1982.

MALWARE DYNAMIC ANALYSIS SYSTEM BASED ON VIRTUAL MACHINE INTROSPECTION AND MACHINE LEARNING METHODS

Alan Nafiiev[1] , *Hlib Kholodulkin*[2] and *Andrii Rodionov*[1]

[1] National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Institute of Physics and Technology, Build. 1, 37, Prosp. Peremohy, Kyiv, 03056, Ukraine
[2] Taras Shevchenko National University of Kyiv, Faculty of Computer Science and Cybernetics, Akademika Hlushkova Ave, 4, Kyiv, 03680, Ukraine

Nowadays, malware authors are creating ever more advanced and sophisticated malware that is almost impossible to detect using static analysis. Even when using dynamic analysis, a malicious file can recognize being executed by the virtual environment and change its code. The aim of this work is to implement the process of collecting and processing behavioral characteristics for a set of different exe-files based on a dynamic analysis system, where the file is not able to detect being observed and can show its nature. Also, based on the collected data, machine learning models have been built and experiments have been carried out with training and test sets.

Keywords: malware dynamic analysis, virtual machine introspection, feature selection, hypervisor, malware detection

Introduction

Malware analysis can be performed using both static and dynamic methods. As a rule, dynamic analysis is more efficient way to determine the functionality of the analyzed program. However, the success of detecting a malicious file depends on the type of a software agent responsible for collecting data and their completeness during dynamic analysis. Therefore, malware authors equip their programs with the ability to detect the agent and bypass such systems. As a result, a malicious file may recognize that it is running on an isolated system and delay its launch, or change its behavior to mislead the analysis system. Therefore, it is critical for dynamic malware analysis systems to provide a real file execution environment with hidden monitoring capability to hide the presence of executable data collection pattern.

In this work, we used the Drakvuf Sandbox malware analysis system [1], which allows you to track malware at the user and OS kernel levels without the need to install an agent in the guest OS. The system is built on the Xen virtualization platform, uses the LibVMI API and the DRAKVUF engine.

Dataset Construction

In the experiment, the training set includes 240 files, 120 benign and 120 malicious. The test set includes 100 files, 50 benign and 50 malicious. Using Drakvuf sandbox, each .exe file was launched on a windows 7 virtual machine. For five minutes the dynamic analysis system reads internal kernel functions used by kernel drivers, tracing system calls, file operations like file creation and deletion made by malware. Thus, we get a .txt file with the detailed behavioral profile of the executable file in dynamics. Each line in the .txt file corresponds to one action that takes place in the operating system during the execution of the executable file. This action is represented as a dictionary data type in python. There is a set of parameters and their values. The following is an example of such an action:

{"Plugin": "filetracer", "TimeStamp": "1657826673.189062", "PID": 560, "PPID": 436, "TID": 572, "UserName": "SessionID", "UserId": 0, "ProcessName": "\\Device\\HarddiskVolume2\\Windows\\System32\\svchost.exe", "Method": "NtOpenFile", "EventUID": "0x8c8", "FileName": "\\??\\C:\\Windows\\System32\\DriverStore\\en-US\\", "ObjectAttributes": "OBJ_CASE_INSENSITIVE"}

ProcessName - The name that the system uses to identify the process to the user
Method - contains functions from the Windows library ntdll.lib
FileName - the name of file that the system is using in a certain moment
In the end, 340 .txt files have been collected with the result of dynamic analysis for each .exe file.

Feature Selection

Now that we have the results of dynamic analysis, there is a fundamental task of collecting and processing data. Based on these data, the final matrix will be formed, which will be sent to the SVM machine learning algorithm. For analysis, 3 parameters were selected (Method, FileName,

ProcessName), which contain significant information about the action taken in the operating system. To form the features on which SVM is trained, the N-gram method is used, where transition probabilities are used to build a Markov chain instead of 2-grams as the basic data representation. The following subsections will describe the process of generating the final matrix based on each of the 3 parameters.

Parameter "Method"

The first step in feature generation is parsing data from all .txt files. We received 340 arrays of the following form:

[NtOpenFile, NtCreateFile, …, NtWriteFile, NtQueryAttributesFile] (1)

It turned out that in all .txt files the Method parameter takes only 7 unique values: NtReadFile, NtWriteFile, NtCreateFile, NtOpenFile, NtSetInformationFile, NtQueryAttributesFile, NtOpenDirectoryObject.

Let's show the principle of formation of the final matrix on the example of the sequence of values (1). Based on such an array, a two-dimensional quadratic adjacency matrix of 7*7 dimensions is constructed and for each pair of values in the matrix we counted how many times the first value immediately followed the second.

Our Markov chain can be represented as a graph in which the vertices are the states of the process (all possible values of the Method parameter), and the edges are transitions between states, and p_{ij} is written on the edge from i to j - the probability of transition from one value to another. As a result, we get a quadratic transition matrix $= \left|\left|p_{ij}\right|\right|$, dimension 7*7, on which the following conditions are imposed:

$$p_{ij} \geq 0,$$

$$\forall i \sum_{j}^{N} p_{ij} = 1$$

Further, this matrix is converted into a vector form and completed into the final matrix.

Parameter "FileName"

After parsing all .txt files, it turned out that there are 25440 unique values of the FileName parameter. It would be possible to build the final matrix in the same way as in the case of the Method parameter. However, in this case, the adjacency matrix will be very large (25440 x 25440) and the

dimension of the final matrix will be (340 x 647 million), which is too large for the power of modern computers. In addition, most of the cells in such a matrix will be zeros. Therefore, the problem arises to choose a smaller number of the most significant features.

To form a set of such features, a method is used where values that occur in total the most times in all .txt files are taken. So, 3 final matrices with different dimensions were formed based on 3 sets of values of the FileName parameter. The first set contains the 88 most common values. The second is 326 and the third is 712.

Parameter “ProcessName”

The array of unique ProcessName parameter values contains 312 elements. However, most of these elements are values that contain the name of the .exe file being launched. That is, such a value is found only in one .txt file out of 340, which means that it is not of a particular value when training a machine learning model. Therefore, from the set of 312 unique values, those that contain their own file names were removed. The resulting array had 65 elements. Based on this set, adjacency matrices with dimensions (65 x 65) and one final matrix were built for each .exe file.

All three parameters

After the formation of all the final matrices, the idea arose to make a model that would be based on information not only from one parameter, but from all three at once. Each .exe file corresponds to a vector of numeric values that describes the nature of the file. The final matrix consists of such vectors. To combine three vectors from three different matrices, we can sum these vectors together. Then one, the most “complete” in terms of information, final matrix will be obtained. To form such a matrix, the following final matrices were taken: a matrix based on the Method parameter with 7 features, a matrix based on the ProcessName parameter with 47 features, and a matrix based on the FileName parameter with 88 features.

As you can see, the dimensions of the three matrices are different, and in order to sum the vectors, we need to achieve the same dimension. For this, the principal component method was used, with the help of which the ProcessName and FileName matrices were reduced to the dimension of the Method matrix.

Training and Metrics

Machine learning models are trained using the SVM algorithm with a square exponential kernel. The algorithm hyperparameters are selected using cross-validation. The test final matrices were formed based on the features that were obtained from the training set. As an assessment of the accuracy of the models under study, F-score is used, which is a joint assessment of such metrics as precision and recall. Also, for the analysis of the models, the plots of the roc_auc and pr_auc curves were used. Table 1 shows the results of the accuracy of all models. All models demonstrated fairly high accuracy rates for all metrics.

Table 1

	F-score		Precision		Recall		Roc_AUC	Pr_AUC
	0	1	0	1	0	1		
Method_7	0.9230	0.9166	0.8888	0.9565	0.9600	0.8800	0.9747	0.9608
ProcessName_47	0.9411	0.9387	0.9230	0.9583	0.9600	0.9200	0.9659	0.9432
FileName_88	0.9320	0.9278	0.9056	0.9574	0.9600	0.9000	0.9676	0.9420
FileName_376	**0.9607**	**0.9591**	0.9423	0.9791	0.9800	0.9400	0.9773	0.9628
FileName_726	0.9183	0.9215	0.9375	0.9038	0.9000	0.9400	0.9828	0.9764
F88+P47+M_7	0.9491	0.9447	0.9438	0.9506	0.9545	0.9390	**0.9860**	**0.9880**

Conclusions

In this study, we have created 6 machine learning models: 3 models based on the FileName parameter, a model based on the Method parameter, a model based on the ProcessName parameter, and a model that is based on all three parameters at once. Out of these three parameters, the FileName parameter turned out to be the most optimal one in terms of its ability to detect malicious files. Namely, the model containing the average number of features - 376. It showed the highest F-score metrics and almost the best values of the roc and precision-recall curves. Based on this, it can be argued that not always a model with a large number of features will show the best result. It is necessary to find the optimal set of features, since the stage of their formation is probably the most important step in the process of detecting malicious files. It is also worth noting the high performance of the model based on three parameters. The idea of summing the vectors of three adjacent matrices justified our expectations, since the performance of the roc and precision-recall curves turned out to be the highest among all models. It is also worth noting that despite the fact that the model based on the Method parameter has only 7 features, the

accuracy results were at a sufficient level among other models with a lot more features.

References

1. Tamas K Lengyel, Steve Maresca, Bryan D Payne, George D Webster, Sebastian Vogl, and Aggelos Kiayias. Scalability, fidelity and stealth in the drakvuf dynamic malware analysis system. In The 30th Annual Computer Security Applications Conference, pp.s 386–395, 2014.
2. Xen Project. Available at: https://xenproject.org/
3. LibVMI. Available at: http://libvmi.com/

ВИЯВЛЕННЯ І ПОБУДОВА МОРФОЛОГІЧНИХ ТАБЛИЦЬ НА ОСНОВІ РЕЗУЛЬТАТІВ АНАЛІЗУ СЛАБКОСТРУКТУРОВАНИХ ДАНИХ

Панкратова Наталія Дмитрівна, Савченко Ілля Олександрович
Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», Київ, Україна
savil.ua@gmail.com

В доповіді розглядається класифікація типів морфологічних таблиць, які доцільно використовувати в процесі сценарного аналізу. Пропонуються способи виявлення і побудови таблиць різних типів з семантичних мереж.

Ключові слова: сценарний аналіз, метод морфологічного аналізу, морфологічні таблиці, семантичні мережі, бази знань.

Модифікований метод морфологічного аналізу (МММА) є потужним інструментом якісного аналізу в задачах, де об'єкти характеризуються неточністю, невизначеністю, неповнотою, нечіткістю інформації і через це мають велику кількість альтернативних варіантів реалізації [1].

Від коректного, адекватного, продуктивного опису об'єктів аналізу часто залежить успішність процесу передбачення в цілому. Тому побудова морфологічних таблиць в МММА є надзвичайно важливим етапом, що забезпечує відповідність результату реальній картині світу.

Побудова морфологічних таблиць є творчим процесом, кінцевий результат якого може бути забезпечений тільки людиною-аналітиком. Через великі обсяги даних у складних багатопараметричних задачах цей процес може бути занадто трудомістким для проведення його вручну.

Розглянемо типи об'єктів, які можуть бути утворюючими для морфологічної таблиці.

Опис об'єкта. Морфологічна таблиця описує заданий матеріальний або нематеріальний об'єкт або систему.

Опис стану. Морфологічна таблиця описує стан деякого об'єкта або системи. Невизначеність полягає в тому, що саме відбувається або відбуватиметься в об'єкті морфологічного дослідження.

Опис дії. Морфологічна таблиця описує певну подію або взаємодію між об'єктами. Невизначеність може полягати у стані цієї

системи (контекст події), у об'єктах, які беруть участь у події, і в характеристиках перебігу самої події.

Як бачимо, кожний наступний вид морфологічної таблиці може потребувати наявності попередніх.

Різним типам морфологічних таблиць можна співвіднести фрагменти мережі сутностей. Для опису об'єкта з точки зору архітектури бази знань параметрами і альтернативами морфологічної таблиці можуть стати:

1. Параметр – класифікація об'єкта за деяким розрізом, альтернативи – підкласи об'єкта в цьому розрізі;
2. Параметр – характеристика об'єкта, альтернативи – можливі значення або діапазони значень цієї характеристики;
3. Пункти 1, 2, 3 для елементів (підоб'єктів) об'єкта.

Процедура конструювання морфологічної таблиці може виконуватись рекурсивно, з глибиною деталізації в залежності від потреб задачі. Процес побудови аналітиком морфологічної таблиці ґрунтується на інтерактивному режимі взаємодії із базою знань.

При побудові морфологічної таблиці для опису стану об'єкта або системи основними групами параметрів в морфологічній можуть бути:

1. Значення показників або характеристик, які має об'єкт дослідження.
2. Значення показників або характеристик, які мають пов'язані об'єкти або системи (опис їх стану). Серед них можуть бути підсистеми об'єкта дослідження, надсистеми (опис зовнішнього впливу, контекст) і пов'язані об'єкти того ж рівня.
3. Опис пов'язаних об'єктів (включаючи рішення, що могли бути прийнятими щодо об'єкта дослідження і вплинути на його стан).

В описі стану системи також можуть бути об'єкти, які включаються в морфологічну таблицю за наведеною вище процедурою опису об'єктів.

Параметрами для морфологічної таблиці, що описує певну подію, є:

1. Характеристики і підкласи події, що розглядається (сама подія з точки зору бази знань тут розглядається як об'єкт).
2. Опис причетних об'єктів та/або їх станів.
3. Опис стану системи, в рамках якої відбувається подія (контекст).
4. Причини виникнення події (якщо їх доцільно розглядати).

Таким чином, запропоновані прийоми видобування морфологічних матриць дозволяють використовувати великі об’єми неструктурованої інформації для напівавтоматичної побудови моделей МММА.

Перелік посилань

1. Панкратова Н.Д., Савченко І.О. Морфологічний аналіз. Проблеми, теорія, застосування. Навчальний посібник. – Наукова думка. – 2015. – 245 с.

ІНСТРУМЕНТАРІЙ ПЛАТФОРМИ ТРАНСФЕРУ ЗНАНЬ ДЛЯ СТРАТЕГІЧНОГО ПЛАНУВАННЯ

Віталій Циганок [1,2,3] [0000-0002-0821-4877],
Антон Астахов [1],
Володимир Мінас [1],
Максим Коновалюк [2] [0000-0003-4601-3790]

[1] Інститут проблем реєстрації інформації Національної академії наук України, Київ, Україна
[2] Національний технічний університет України "Київський політехнічний інститут імені Ігоря Сікорського", Київ, Україна
[3] Київський національний університет імені Тараса Шевченка, Київ, Україна
tsyganok@ipri.kiev.ua, astakhov.anton97@gmail.com, vminas@i.ua, konovalyuk@gmail.com

Дослідження направлене на вирішення глобальної проблеми – трансферу знань від осіб, що володіють знаннями (досвідом, інтуїцією) у певній сфері до осіб, що потребують цих знань задля вирішення практичних задач. Застосування цього теоретичного підгрунтя призначено на практиці для стратегічного планування у різних сферах, особливо, у слабко структурованих предметних областях. Запропоновано технологію та відповідний програмний інструментарій платформи трансферу знань для побудови стратегічних планів.

Ключові слова: трансфер знань, стратегічний аналіз, цільове динамічне оцінювання альтернатив.

Мета дослідження

Метою дослідження є розробка програмного інструментарію платформи трансферу знань для стратегічного планування у слабко структурованих предметних областях.

Методологія

В основу технології покладено групову побудову ціле-орієнтованої моделі системи (предметної області) [1, 2], яка створюється шляхом декомпозиції головної стратегічної цілі та враховує часові та ресурсні властивості компонентів системи та взаємозв'язків між ними. Програмний інструментарій дозволяє віддалено інженерам

знань та запрошеним ними експертам надавати знання про предметну область для створення адекватної моделі системи.

Сутність підходу

Стратегічне планування [3] можна розбити на такі етапи:

1. Визначення основної стратегічної цілі;
2. Декомпозиція цілі на підцілі, що впливають на головну ціль в даній предметній області;
3. Визначення атомарних цілей в межах компетентності особи, що приймає рішення (ОПР);
4. Визначення ресурсів, необхідних для впровадження кожного заходу (проекту);
5. Визначення тривалості та затримок при реалізації кожного проекту;
6. Визначення оптимального розподілу ресурсів між даними проектами, що дозволить максимізувати ефективність досягнення основної стратегічної цілі при заданому фінансуванні;
7. Забезпечення можливості корекції стратегічного плану.

Метод цільового динамічного оцінювання альтернатив (МЦДОА) [4, 5] дозволяє здійснювати оцінювання, базуючись на побудованій експертним шляхом моделі предметної області. Метод дає змогу використовувати найбільш загальні моделі слабко-структурованих предметних областей, що достатньою мірою повно та адекватно відображають особливості тієї, чи іншої предметної області.

Ступінь досягнення цілі в запропонованій моделі визначається наступним чином:

$$d_i(t) = \begin{cases} 0, & \text{if} \quad D_i(t) < T_i \\ T_i, & \text{if} \quad D_i(t) = T_i \\ f\big(D_i(t)\big), & \text{if} \quad T_i < D_i(t) < 1 - \sum_j \left| w_{ij}^{(k-)} \right|, \\ 1, & \text{if} \quad 1 - \sum_j \left| w_{ij}^{(k-)} \right| \le D_i(t) \le 1 \end{cases}$$

де $D_i(t) = \sup\limits_k \sum_j w_{ij}^{(k)} d_j(t)$; T_i – поріг досягнення i-ї цілі; $f\big(D_i(t)\big)$ – функція ступеня досягнення i-ї цілі в момент часу t; $w_{ij}^{(k-)}$ – ЧКВ j-ї

цілі в *k*-й групі сумісних цілей, який має негативний вплив на *i*-у ціль.

Результати

Розроблено методи колективного отримання та обробки знань. На їх основі реалізовано програмний інструментарій платформи трансферу знань для стратегічного планування у слабко структурованих предметних областях. Розроблено теоретичні основи та методи достовірного отримання колективних знань в певній предметній області, їхнього узгодження, узагальнення для подальшого застосування у різних сферах.

Висновки

Запропоновані теоретичні засади та методи для достовірного отримання та застосування колективних знань у різних галузях, наявність яких дозволила прийти до створення та практичного застосування інструментарію для стратегічного планування у різних сферах.

Перелік посилань

1. Saaty, Thomas L. Mathematical Principles of Decision Making (Principia Mathematica Decernendi). *Comprehensive coverage of the AHP, its successor the ANP, and further developments of their underlying concepts*, Pittsburgh, 2009.
2. Тоценко В.Г., Методи та системи підтримки прийняття рішень. Алгоритмічний аспект. *Видавництво «Наукова думка»*, Київ, 2002.
3. Mintzberg H., Quinn J. B., & Ghoshal S. The strategy process (Rev. European). Prentice Hall, 1998.
4. Totsenko, V.G. On One Approach to the Decision Making Support while Planning Research and Development. Part II. The Method of Goal Dynamic Evaluation of Alternatives. *Journal of Automation and Information Sciences*. *33*(4), 2001, 82-90.
5. Циганок В.В. Удосконалення методу цільового динамічного оцінювання альтернатив та особливості його застосування. *Реєстрація, зберігання і обробка даних*. т.15. №1. 2013. С. 90-99.

ОГЛЯД МЕТОДІВ ПАРНИХ ПОРІВНЯНЬ, ЩО ВРАХОВУЮТЬ РАНЖИРУВАННЯ

Олег Андрійчук[1], *Сергій Каденко*[1]
[1]Інститут проблем реєстрації інформації Національної академії наук України
andriichuk@ipri.kiev.ua, seriga2009@gmail.com

У доповіді представлений стислий огляд методів підтримки прийняття рішень на основі парних порівнянь альтернатив, що враховують ординальне співвідношення між ними. Основну увагу зосереджено на авторському методі експертних парних порівнянь, який базується на певній послідовності оцінювання альтернатив. Ця послідовність визначається попереднім ранжируванням альтернатив, що порівнюються. Даний метод дозволяє підвищити достовірність результатів експертизи, отримати більш узгоджені експертні дані, а також, за необхідності, знизити трудомісткість експертизи за рахунок зменшення необхідної кількості парних порівнянь.

Ключові слова: слабо структурована предметна область, експертна оцінка, матриця парних порівнянь, ранжирування.

Вступ: проблематика

У слабо структурованих предметних областях часто виникає потреба у прийнятті рішень, що полягають у виборі найкращої альтернативи, або побудові рейтингу альтернатив із заданої множини за певним цільовим критерієм. Водночас, слабо структурована предметна область характеризується високим рівнем невизначеності, що не дозволяє формально описати предметну область та альтернативні варіанти рішення, а також отримати кількісну інформацію про них. Відповідно, часто, єдиним джерелом такої інформації в слабко структурованих предметних областях є експертні оцінки (зокрема, задані у вигляді парних порівнянь (ПП)) альтернатив, критеріїв, факторів та варіантів рішень. Зауважимо, що у реальних слабо структурованих предметних областях, коли виникає необхідність у експертному оцінюванні, часто застосовуються примітивні процедури, що базуються на принципі ad hoc. Втім, застосування більш формальних та універсальних процедур експертного оцінювання дозволило б підвищити достовірність результатів експертиз, що проводяться.

ПП – це співвідношення невідомих відносних ваг альтернатив, що порівнюються. На основі матриці парних порівнянь (МПП) альтернатив обчислюються їхні відносні ваги. Для цього застосовується якийсь із багатьох наявних методів (власного вектору [1] або інший).

Для того, щоб порівняти n альтернатив між собою необхідно виконати $n(n-1)/2$ ПП. Як бачимо, із зростанням розмірності, кількість ПП зростає зі швидкістю $O(n^2)$. Тому актуальною є проблема розробки методів, які б дозволяли зменшити кількість необхідних ПП і при цьому уникнути втрат та спотворень експертної інформації.

Стан проблеми: наявні методи, що враховують інформацію про ранжирування

Низка методів ПП передбачає використання апріорної інформації про ранжирування альтернатив. Більшість із них базується на результатах експериментів з когнітивної психології [2], які продемонстрували, що людина більш точно оцінює об'єкти, якщо вони подаються для оцінки у порядку спадання ступеня вираженості критерію оцінки (від найкращого до найгіршого, від найбільшого до найменшого тощо).

Історично першим методом, який використовує даний принцип, можна вважати багатокритеріальний метод TOPSIS [3]. У ньому, на основі багатокритеріальних експертних оцінок множини альтернатив, генеруються «ідеальна найгірша» та «ідеальна найкраща» альтернативи. Потім складається рейтинг альтернатив за відстанню до цих ідеальних об'єктів.

Протягом останнього десятиліття з'явилося кілька одно-критеріальних методів ПП, що використовують інформацію про ординальне співвідношення між альтернативами. Так, метод "best\worst" [4] передбачає (замість повного перебирання пар альтернатив) порівняння усіх альтернатив із заданої множини лише з найкращою та найгіршою із них. Метод TOP2 (best\second best) [5] передбачає порівняння усіх альтернатив із заданої множини лише з першою та другою альтернативами у ранжируванні.

Метод порівняння найвіддаленіших альтернатив

Ще один метод, що використовує інформацію про ранжирування альтернатив, запропоновано авторами даної доповіді [6]. Метод передбачає порівняння альтернатив у порядку спадання відстані між

ними у ранжируванні. Відповідно, ПП здійснюються у кілька черг. Нехай альтернативи пронумеровані відповідно до ранжирування: $a_1 > a_2 > \cdots > a_n$, де a_i це альтернатива з номером та рангом i, $i = \overline{1,n}$, а n – загальна кількість альтернатив. Тоді черговість виконання експертом ПП (яка забезпечує найвищу адекватність результатів ПП уявленням експерта) є наступною:

черга 1: (a_1, a_n) (ранги відрізняються на $(n-1)$);

черга 2: (a_1, a_{n-1}) або (a_2, a_n) (ранги відрізняються на $(n-2)$);

черга 3: (a_1, a_{n-2}) або (a_2, a_{n-1}) or (a_3, a_n) (ранги відрізняються на $(n-3)$);

…

черга $(n-1)$: (a_1, a_2) або (a_2, a_3) або … або (a_{n-1}, a_n) (ранги відрізняються на 1).

Результати експериментального дослідження методу

Експериментальне дослідження даного методу, детально описане у [6], продемонструвало, що відносні ваги альтернатив, обчислені методом власного вектору на основі МПП, що відповідає вказаній послідовності ПП, є більш адекватними уявленням експертів про предметні області, у яких вони компетентні (аніж ваги, отримані на основі інших послідовностей ПП).

Більш того, останні дослідження авторів показали, що й самі МПП, отримані на основі вказаної послідовності ПП, є, загалом більш узгодженими, аніж МПП, отримані на основі інших послідовностей.

Таблиця 1. Результати експериментального дослідження методу

Тип послідовності парних порівнянь	Число експертів, які присвоїли послідовності відповідний ранг			Середнє значення відношення узгодженості оцінок CR, %	Частка (у %) та кількість прецедентів, у якій результати ПП – найбільш узгоджені
	1	2	3		
A (спочатку – найвіддаленіші)	45	22	16	10,73	36,14 (30)
B (довільні)	15	38	30	11,61	32,53 (27)
C (спочатку – найближчі)	23	23	37	11,88	31,33 (26)

У Таблиці 1 показані результати експериментальних досліджень методу. Загалом на даний момент було проведено 83 змістовні прецеденти експерименту. Як можна побачити з таблиці, більшість

експертів віддає перевагу вказаній послідовності ПП (вважає відносні ваги альтернатив, отримані на основі цієї послідовності, найбільш адекватними власним глибинним уявленням). При цьому, значення середнього показника (не)узгодженості (CR) [1] МПП, побудованих на основі вказаної послідовності ПП, є меншим, аніж для інших послідовностей. Також, загалом, за значенням CR, у більшості випадків (30 з 83) запропонована послідовність ПП є лідером з-поміж трьох послідовностей ПП, що досліджувалися.

Ідея експерименту з порівняння методів

Очевидним напрямком подальших досліджень має стати порівняння запропонованого методу ПП («порівняння найвіддаленіших») із іншими подібними методами (TOP2 та best\worst). Етапи експерименту, що базується на принципах імітаційного моделювання, наступні: 1) Генерування випадкового вектору ваг; 2) Побудова ідеально узгодженої МПП; 3) Зашумлення ІУМПП; 4) Обчислення ваг на основі зашумленої МПП (різними методами – «ПП найвіддаленіших», TOP2, best\worst); 5) Обчислення та порівняння похибок обчислення ваг. Втім, виникає декілька питань щодо умов такого порівняльного експерименту.

По-перше, TOP2 та best\worst є неповними методами ПП. Обидва методи передбачають здійснення лише $2n-3$ порівнянь. У той же час, метод «порівняння найвіддаленіших» дозволяє як повністю вибудувати МПП, так і мінімізувати кількість порівнянь до мінімальної ($(n-1)$ ПП) [7]. Відтак, постає питання, скільки саме ПП із відповідних черг (див. вище) слід задіяти у експерименті? Друге питання: як саме доповнювати неповну МПП до повної (щоб можна було застосувати метод власного вектору для обчислення відносних ваг альтернатив)? Третє питання: які ще методи, окрім власного вектору, можна задіяти у експерименті для обчислення ваг; чи вони працюють для неповних МПП?

Висновки

У доповіді наведено огляд методів експертного оцінювання, що враховують попередні ранжирування альтернатив. Показано, що урахування ординального співвідношення між альтернативами дозволяє скоротити кількість ПП без суттєвих втрат та спотворень експертної інформації. Запропоновано авторський метод ПП, що базується на порівняннях найвіддаленіших альтернатив у ранжируванні. Отримано експериментальні результати, які свідчать

про підвищення узгодженості експертних ПП та можливість, за необхідності, максимально скоротити кількість ПП. Подальші дослідження за тематикою передбачають проведення експерименту з порівняння авторського методу та інших аналогічних методів ПП, зокрема best\worst та TOP2.

Reference

1. Saaty, T. (1980). The analytic hierarchy process. New York: McGraw-Hill.
2. Stevens, S. S. & Galanter, E. (1957) Ratio Scales and Category Scales for a Dozen Perceptual Continua, Journal of Experimental Psychology, Vol. 54, No 6: 377-411.
3. Hwang, C.L. & Yoon, K. (1981) Multiple attribute decision making: methods and applications: a state-of-the-art survey. Berlin, New York: Springer-Verlag.
4. Rezaei, J. (2015) Best-worst multi-criteria decision-making method. Omega, 53, 49-57, https://doi.org/10.1016/j.omega.2014.11.009.
5. Szádoczki, Z., Bozóki, S., Juhász, P. et al. (2022) Incomplete pairwise comparison matrices based on graphs with average degree approximately 3. Ann Oper Res. https://doi.org/10.1007/s10479-022-04819-9
6. Andriichuk, O., Tsyganok, V., Kadenko, S., & Porplenko, Y. (2020) Experimental Research of Impact of Order of Pairwise Alternative Comparisons upon Credibility of Expert Session Results, in: Proceedings of the 2020 IEEE 2nd International Conference on System Analysis & Intelligent Computing (SAIC), pp. 1-5, http://dx.doi.org/10.1109/SAIC51296.2020.9239126
7. Kadenko, S., Tsyganok, V., Szádoczki, S. et al. (2021). Improvement of Pair-wise Comparison Methods Based on Graph Theory Concepts. CEUR Workshop Proceedings; Vol. 3241, 46-55.

СИСТЕМА ЗАБЕЗПЕЧЕННЯ ПРОВЕДЕННЯ ВСТУПНОЇ КАМПАНІЇ З ВИКОРИСТАННЯМ ХМАРНИХ ТЕХНОЛОГІЙ

Гнатієнко Григорій Миколайович[1], Іларіонов Олег Євгенович[1], Мирутенко Лариса Вікторівна[1] Власенко Оксана Олександрівна[1], Гамоцька Сніжана Леонідівна[1]

[1] Київський національний університет імені Тараса Шевченка, вул. Володимирська, 64/13, 01601 Київ, Україна

Вступ

В умовах повномасштабного вторгнення в Україну рашистських військ з'являються нові виклики у всіх сферах життєдіяльності нашої країни. Зокрема, війна, розв'язана російською федерацією, внесла значні корективи у підготовку та проведення Вступної кампанії (ВК) до вищих навчальних закладів України (ВНЗ) щодо проведення фахових вступих іспитів до магістратури.

Проблема

Необхідно забезпечити справедливий, своєчасний та професійний відбір кращих абітурієнтів до українських ВНЗ. При цьому подбати про забезпечення зручності як абітурієнтів, так і викладачів, безпеку усіх учасників вступної кампанії, оптимізацію усіх чинників, які стосуються цієї події.

Постановка задачі

Усі аспекти ВК мають бути належним чином формалізовані, проблема забезпечення якості ВК має бути структурована, результати ВК мають бути кількісно оцінені.

Критерії

Основними критеріями якісного проведення ВК на факультеті інформаційних технологій Київського національного університету імені Тараса Шевченка (ФІТ) було вибрано такі:

f_1 – забезпечення максимальної відкритості та інформованості суспільства про перебіг ВК;

f_2 – забезпечення конфіденційності інформації про персональні дані абітурієнтів, склад екзаменаційних білетів тощо;

f_3 – забезпечення анонімності абітурієнтів, у тому числі і для викладачів, які беруть участь у ВК;

f_4 – забезпечення адекватності оцінювання знань абітурієнтів;

f_5 – забезпечення варіативності екзаменаційних білетів з метою поставити різні хвилі абітурієнтів у рівні умови;

f_6 – забезпечення надійності функціонування системи.

Варіанти рішень

На вибір варіанту рішення впливають кілька парамерів: p_1 – способи забезпечення процесу проведення ВК, p_2 – інструментарій комунікації усіх учасників процесу, p_3 – інструментарій оцінювання абітурієнтів та інші параметри.

Таким чином, маємо гіперпаралелепіпед (ГП) можливих варіантів рішень:

$P \ni p = (p_1, ..., p_n)$, n – кількість параметрів.

Модель вступної кампанії

Таким чином, проблема визначення кращоїдля ВНЗ конфігурації параметрів рішень може бути формалізована у класі задач багатокритеріальної оптимізації (ЗБКО)

$$f_i(p) \to opt,\ i = 1, ..., k, \qquad (1)$$

$$p \in P.$$

Варіанти забезпечення

Було визначено варіанти проведення фахових вступних іспитів до магістратури:

- в режимі оффлайн;
- онлайн з використанням сервера ФІТ;
- онлайн з використанням власного сервера кожною кафедрою;
- онлайн з використанням виділеного сервера в датацентрах України;
- онлайн з використанням виділеного сервера в датацентрах а межами України
- онлайн з використанням технологій PaaS.

Інструментарій комунікації

В умовах, що склалися, серед існуючих платформ комунікації найбільш популярними у нашому ВНЗ стали: Zoom, Google Meet та Microsoft Teams.

Інструментарій оцінювання

Серед доступних варіантів для проведення тестування було попередньо виділено: дистанційне системе навчання Moodle, використання Google форм або тестування на базі платформи MicrosoftTeams.

Висновки

Авторами було структуровано проблему супроводження ВК на рівні факультету. Для цього за допомогою експертних технологій було розглянуто основні елементи системи, згенеровано ГП можливих значень параметрів, сформульовано та формалізовано і розв'язано ЗБКО виду (1). Результати ВК на ФІТ доводять ефективність створеної авторами системи забезпечення проведення ВК

Список літератури

1. Волошин О.Ф., Гнатієнко Г.М., Кудін В.І. Послідовний аналіз варіантів. Технології та застосування. – К.: Стилос, 2013, 304 с.

2. Гнатієнко Г.М., Снитюк В.Є. Експертні технології прийняття рішень: Монографія. – К.: ТОВ «Маклаут», 2008. – 444 с.

ГРУПОВА ЗАДАЧА ПЕРЕСЛІДУВАННЯ ДЛЯ ДРОБОВИХ ДИФЕРЕНЦІАЛЬНИХ СИСТЕМ З ЧИСТИМ ЗАПІЗНЮВАННЯМ

Барановська Л.В.[1[0000-0003-0024-8180]], Мухін В.Є.[1[0000-0002-1206-9131]]
[1] Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», Київ, Україна,

lesia@baranovsky.org

Розглянуто групову задачу переслідування для лінійних дробових диференціальних систем з чистим запізнюванням. Розроблено схему методу розв'язувальних функцій для даної задачі з використанням новітнього представлення формули Коші. Сформульовано достатні умови завершення гри та методику практичного знаходження розв'язувальних функцій.

Ключові слова: конфліктно-керований процес, прийняття рішень в умовах конфлікту, диференціальні ігри, диференціальні ігри з дробовими похідними, теорія ігор.

Вступ

На сьогоднішній день широкого застосування, зокрема, у теорії прийняття рішень в умовах конфліктів, набули математичні моделі з дробовими диференціальними рівняннями з запізнюванням. У теорії конфліктно-керованих процесів актуальними, особливо у воєнній практиці, є задачі групового переслідування. Ставиться задача знаходження стратегій переслідувачів (необов'язково оптимальних), які гарантують розв'язність задачі групового переслідування за довільних допустимих керуваннях втікача. При розв'язанні задач групового переслідування існує два основних підходи. Перший з них пов'язаний з позиційним переслідуванням і розвиває правило екстремального прицілювання. Ця методика дає обґрунтування закону переслідування за погонною кривою. Другий підхід базується на використанні обернених функціоналів Мінковського [1]. Для реалізації цієї методики необхідна інформація про передісторію керування втікачі, включаючи його миттєве значення. Коло задач групового переслідування, які можна розв'язати у цьому випадку, значно ширше. З даного підходу випливає правило паралельного переслідування, добре відомого інженерам-проектувальникам.

Постановка задачі

Розглянемо конфліктно-керований процес

$$\left({}^{C}D_{0+}^{\alpha} z_i\right)(t) = -A_i z_i(t-h_i) + f_i\big(u_i(t), v(t)\big), \qquad (1)$$
$$t \geq 0, h_i > 0,\ i = 1, \dots, \nu,$$
$$z_i(t) \equiv \varphi_i(t),\ \acute{z}_i(t) \equiv \acute{\varphi}_i(t), -h_i \leq t \leq 0,$$

де ${}^{C}D_{0+}^{\alpha}$ визначено як дробову похідну Капуто порядку $\alpha \in (1,2)$ з нульовою нижньою межею [2]; $z_i = (z_{i1}, \dots, z_{in})^T$: $[-h_i, \infty) \to R^{n_i}$ є розв'язком, що задовольняє (1) при всіх $t \geq 0$; A_i є сталими дійсними ненульовими матрицями порядку n_i; блок керування задається функціями f_i: $U_i \times V \to R^{n_i}$, неперервними за сукупністю змінних; $u_i(t) \in U_i$, $v(t) \in V$ сукупності непорожніх компактів; φ_i: $[-h_i, 0] \to R^{n_i}$ є довільними диференційовними функціями, які визначають початкові умови.

Сформулюємо задачу переслідування, задану системою (1). Термінальна множина складається з циліндричних множин M_i^*, де $M_i^* = M_i^0 + M_i$, $i = 1, \dots, \nu$, M_i^0 – лінійні підпростори з евклідових просторів R^{n_i}, M_i –непорожні компакти з відповідних ортогональних доповнень L_i до M_i^0 в R^{n_i}.

Мета переслідувачів (u_i) – вивести траєкторію процесу на термінальну множину M_i^* за найменший час. Мета втікача (v) – уникнути зустрічі або максимально відтермінувати момент потрапляння на термінальну множину M_i^*.

Схема методу розв'язувальних функцій

Нехай π_i –оператори ортогонального проєктування з R^{n_i} в L_i. Введемо багатозначні відображення $W_i(t,v) = \pi_i \boldsymbol{S}_{h_i,\alpha}(A_i(t)^{\alpha}) f_i(U_i, v)$, $W_i(t) = \cap_{v \in V} W_i(t,v)$ [3]. Розглянемо умову Понтрягіна: $W_i(t) \neq \emptyset$ для всіх $t \geq 0, i = 1, \dots, \nu$.

Для конфліктно-керованого процесу (1) за умови виконання умови Понтрягіна для початкового стану $\varphi_i(.)$ згідно зі схемою методу розв'язувальних функцій знайдено достатні умови для приведення траєкторії процесу з початкового стану на термінальну множину в наперед заданий момент часу з використанням відповідної квазістратегії.

Література

1. Chikrii A.A. Conflict controlled processes. Boston; London; Dordrecht: Springer Science and Business Media, 2013. 424 p.

2. Ahmed M. Elshenhab, Xing Tao Wang. Representation of solution for linear fractional systems with pure delay and multiple delays. *Mathematical Methods in the Applied Sciences*. 2021, Vol. 44. P. 12835–12850.

3. Baranovska, L.V. Pursuit Problem for Fractional Differential Systems with Pure Delay. *Cybern Syst Anal* 58, 409–416 (2022).

https://doi.org/10.1007/s10559-022-00473-y.

АНАЛІЗ І РОЗРОБКА МАТЕМАТИЧНИХ МОДЕЛЕЙ ОЦІНЮВАННЯ ІНВЕСТИЦІЙНИХ РИЗИКІВ НА ФІНАНСОВИХ РИНКАХ

Кузнєцова Наталія Володимирівна[1], Батейко Едуард Анатолійович[1]

[1] Інститут прикладного системного аналізу Національного технічного університету України «Київський політехнічний інститут ім. І.Сікорського», Київ, Україна
natalia-kpi@ukr.net, bateikoeduard@gmail.com

Проведено аналіз існуючих підходів і математичних моделей для прогнозування інвестиційних ризиків і запропоновано власні математичні моделі на основі методологій Value-at-Risk, Conditional VaR. Для практичного моделювання було обрано фондовий ринок та різні за напрямом діяльності компанії S&P 500. Досліджено ціни активів компаній у розрізі промислового сектора за останні 5 років, побудовано і оброблено часові ряди цін акцій у вигляді прибутку за один день для кожної акції, розроблено моделі VaR, cVaR, Monte Carlo VaR.

Ключові слова: інвестиційні ризики, VaR, CVaR, фінансовий ринок, часові ряди.

Вступ

Фінансові інструменти для примноження власного капіталу є одним з найбільш поширених об’єктів для дослідження і розробки математичних моделей. Це пов’язано з підсвідомим бажанням людей примножити власні заощадження, поклавши їх на депозити, інвестувавши в дорогоцінні метали, вклавши в нерухомість, криптовалюту або в акції найбільш відомих компаній. Вибір і можливості інвестування суттєво обмежуються наявними матеріальними засобами, законодавством, доступом і можливістю для інвестування на фінансових (фондових) ринках, а головне − готовністю і толерантністю інвестора до ризику. Головною метою інвестування є збереження коштів від інфляції та примноження їх. Тому необхідним є створення якісного інвестиційного портфеля, враховуючи як ризики фондового ринку, так і людський фактор.

Оцінювання інвестиційного ризику

За методологією Value at Risk (VaR) можна з певним довірчим рівнем обчислити верхню межу втрат в результаті змін факторів ризику у довірчому інтервалі:

$$P(Loss_t(k) < VaR_t(k)) = (100 - \alpha)\%,$$

де $Loss_t(k)$ – фактичні втрати на момент часу t за період k днів, $VaR_t(k)$ – прогнозовані втрати на момент часу t за період k днів, α – довірчий рівень. Conditional Value at Risk (CVaR) визначає кількість ризику або «товщину хвоста» для інвестиційного портфеля і розраховується через середньозважене значення «екстремальних» втрат у хвості, що виходять за межі граничного значення VaR.

$CVaR = E(X \mid X > VaR)$, тобто $CVaR = \frac{1}{1-c} \int_{-1}^{VaR} xp(x)dx$,

де $p(x)$ – це щільність розподілу втрат, c – точка відсікання на розподілі, встановлена аналітиком як поріг VaR, VaR – погоджена верхня межа VaR.

Тоді очікувані втрати або прибуток інвестора будуть визначатись як середнє значення VaR в межах певного довірчого інтервалу (квантилю).

Для моделювання було обрано різні галузі інвестування з портфелю S&P 500, виявлено найбільш поширені напрями: промислові підприємства, IT компанії, фінансові компанії і компанії, які відносяться до галузі охорони здоров'я. Було розраховано середній денний прибуток за допомогою моделей VAR і CVAR з довірчим інтервалом 0.95 та виконано моделювання інвестиційного портфеля на 100 днів за допомогою моделі Monte Carlo VAR з початковими інвестиціями розміром 10000 доларів.

Було проведено 4 різних типи експериментів для різного розміру часових рядів на весь розмір навчальних даних з 1 квітня 2016 року по жовтень 2021 року. В експериментах оцінювалися різні часові інтервали інвестування (1 день, 3, 7 та 30 днів). Та відповідно з різним параметром часу для оцінювання і планування доходу інвестора на вікно у 1, 7, 30 та 90 днів. Далі було виділено найбільш цікаві для інвестування компанії і сформовані портфелі так званих "блакитних фішок". Було побудовано моделі Var, cVar, Expected Porfolio Return на 180, 365 днів і весь період інвестування. В результаті моделювання і прогнозування виявлено, що найбільш

ефективний спосіб інвестування − це інвестувати в “блакитні фішки” на всьому часовому інтервалі.

Висновки

Виконане дослідження показало можливість побудови і комбінування різних моделей на основі VaR, cVar, Monte Carlo Var з методами інтелектуального аналізу даних для розробки стратегії інвестування і оцінки можливих прибутків та втрат в залежності не лише від волатильності фінансових рядів, а й від толерантності до ризику самого інвестора. Даний підхід може бути застосований при розробці мобільних агентів для роботи на фінансових ринках, які за початково заданими умовами і з урахуванням відношення інвестора до ризику будуть обирати стратегію і поведінку на ринку та фінансові інструменті (акції), які відповідають очікуваним для інвестора волатильності і доходом за портфелем.

СЕМАНТИЧНИЙ ПІДХІД ДО БАГАТОКРИТЕРІАЛЬНОГО СПІВСТАВЛЕННЯ СКЛАДНИХ ІНФОРМАЦІЙНИХ ОБ`ЄКТІВ

Рогушина Юлія Віталіївна[1], Гладун Анатолій Ясонович[2]
[1]Інститут програмних систем НАНУ, Київ, Україна,
[2]Міжнародний науково-навчальний центр інформаційних технологій та систем НАНУ та МОНУ, Київ, Україна
ladamandraka2010@gmail.com, glanat@yahoo.com

Запропоновано моделі та методи співставлення складних інформаційних об'єктів на основі як складової прийняття рішень в інтелектуальних інформаційних системах на основі використанні онтологій відповідної предметної області та метрик визначення семантичної близькості. Розроблено алгоритм семантичного співставлення об'єктів з подібною структурою, що визначена спільною онтологією, який дозволяє генерувати множину критеріїв співставлення та визначати їх ієрархію для поточної ситуації.
Ключові слова: складний інформаційних об'єкт, прийняття рішень, онтологія, семантична подібність.

Вступ

Одним з ключових питань у розробці складних систем є підвищення ефективності прийняття рішень у проблемних ситуаціях, таких як керування ризиками, менеджмент комплексних проектів та розподіл людських ресурсів. Системи підтримки прийняття рішень (Decision Support System) є специфічним класом інтелектуальних інформаційних систем (ІІС), які допомагають фахівцям вибрати/сформувати потрібну альтернативу серед набору припустимих варіантів при прийнятті відповідальних рішень, і часто поєднують математичні методи та моделі пошуку рішення з евристичними, логіко-лінгвістичними моделями та методами, що базуються на знаннях фахівців – експертів, моделях людських міркувань та накопиченому досвіді. Проблеми у прийнятті рішень в таких системах стосуються вибору найбільш інформативних ознак, за якими порівнюються потенційні рішення. Ситуацію ускладнює необхідність аналізу значної кількості параметрів об'єктів, з яких складаються потенційні рішення, та потреба в уніфікації їх структури. У відкритому середовищі дефіцит часу на обробку інформації та її швидкі зміни обмежує можливість використання

традиційних методів прийняття рішень, що забезпечують знаходження оптимального рішення. Одним з напрямків розв'язання цієї проблеми є застосування елементів штучного інтелекту та моделювання знань предметної області (ПрО) задачі.

Проблеми моделювання знань на сучасному етапі для ІІС найбільш повно реалізуються на основі онтологічного аналізу [1] як окремого випадку семантичного аналізу [2]. Онтології забезпечують інтероперабельність та однозначну інтерпретацію знань ПрО, і тому вони можуть розглядатися як зовнішні джерела інформації для різних ІІС. Тому доцільним є використання підсистем підтримки прийняття рішень в ІІС на основі онтологічного інженерінгу.

Прийняття рішень у відкритому середовищі пов'язане з пошуком актуальної інформації щодо змін у цьому середовищі, що стосуються як самих об'єктів, щодо яких приймаються рішенні, так і тих критеріїв, які впливають на цей вибір. У багатьох інтелектуальних інформаційних системах рішення – це структурований набір елементів, кожен з яких має власну структуру та пов'язаних з іншими елементами різноманітними семантично навантаженими відношеннями. В такому випадку процес прийняття рішень потребує співставлення таких наборів, які надалі ми розглядаємо як складні інформаційні об'єкти (СІО).

Кожен СІО подається як впорядкований набір з більш ніж одного *інформаційного об'єкта* (ІО), які пов'язані між собою відношеннями та відповідають вимогам щодо структури та значень властивостей ІО. В кожному конкретному випадку така структура визначається на основі знань щодо предметної області (ПрО) та специфіки задачі. Через те, що зараз використання онтологій є практично стандартом для формалізації знань щодо ПрО у Web-орієнтованих ІІС, надалі ми застосовуємо саме елементи онтологічного аналізу для опису моделі та методів співставлення СІО.

З точки зору онтологічного аналізу ІО – це будь-які класи або екземпляри онтології. ІО, що відповідають класам онтології, характеризуються своєю структурою – набором властивостей та їх характеристик, а також припустимими відношеннями з іншими класами онтології. ІО, що пов'язані із екземплярами класів онтології, також мають значення (всіх або деяких) цих властивостей, серед яких можуть бути відношення з іншими екземплярами онтології. Таким чином, СІО може розглядатися як підмножина онтології ПрО, що виокремлюється відповідно до задачі користувача.

Приклади СІО, що генерується на основі організаційної онтології певного підприємства, – це команда та обладнання, які

використовуються для виконання певного проекту або для проведення наукової конференції. В цьому випадку прийняття рішення потребує вибору такої підмножини особистого складу організації, що найбільш ефективно виконає проект. Приклади СІО, що базуються на онтології навчального закладу, – вибір спеціальності або побудова навчального плану для отримання визначеного набору компетенцій, тоді як прийняття рішення пов'язане з отриманням необхідних компетенцій за найменший час або з найбільшою повнотою, яку можуть забезпечити викладачі даної установи [1].

Співставлення складних інформаційних об'єктів

Співставлення СІО є складовою прийняття рішень в ІІС. При цьому проблеми генерації рішення як послідовності певних дій в динамічному інформаційному оточенні знаходяться поза сферою розгляду цієї роботи. Основна увага приділяється багатокритеріальному порівнянню тих СІО, що відповідають вимогам задачі.

Особливістю запропонованого підходу є те, що співставляється відносно невелика кількість СІО, тобто не всі теоретично можливі, а лише ті, які можуть бути обрані в поточній ситуації. Таким чином, проблема полягає не у пошуку оптимального (за певними критеріями) рішення, а у виборі прийнятного рішення з набору наявних. Наприклад, для виконання певного проекту потрібно обрати групу співробітників з певного підрозділу, а не взагалі з усіх людей. Тому припустима ситуація, коли всі можливі рішення є незадовільними і змінюють ситуацію тільки на гіршу. Наприклад, виконання проекту недостатньо компетентними співробітниками призведе до втрати часу та ресурсів, але потрібний результат не буде отримано. При цьому значущість критеріїв може змінюватися з часом через зміни в динамічному інформаційному оточенні. Найбільш поширеним прикладом зміни пріоритетів є вартість виконання робіт та швидкість отримання результатів.

Співставлення СІО забезпечує обґрунтування пошуку інформації на змістовному рівні. Приклади семантичного пошуку: знайти групу людей з певною кваліфікацією, що працюють в одній організації, яка відповідає умовам конкурсу; визначити країни, в яких проводилися наукові дослідження певної тематики, результати яких були опубліковані в обраній множині журналів за певний період часу тощо. Але якщо результатом пошуку є, як правило, набір ІО одного або кількох класів, тоді як обмеження використовуються тільки для

відбору прийнятних варіантів, то у багатьох інших задачах сам результат є множиною різних сукупностей ІО різних типів, кожна з яких відповідає певним умовам. Прикладом використання СІО є ієрархічне агрегатне оцінювання (ІАО), що забезпечує співставлення команд з кількома рівнями ієрархії шляхом аналізу індивідуальних та колективних результатів тестів.

Етапи співставлення СІО

У загальному випадку задача співставлення СІО, які мають різну структуру та базуються на різних онтологіях, потребує вирівнювання цих онтологій та пошуку подібності між їх структурними елементами – ІО та їх поєднаннями. В даній роботі ми розглядаємо окремий випадок такої задачі, коли всі СІО, що потрібно співставити, базуються на єдиній онтології та мають подібну структуру, який потребує виконання таких етапів [3]:

- створення еталонної моделі СІО, яка відображає вимоги користувача;
- генерація набору наявних СІО, які структурно відповідають еталонній моделі, на основі інформації про поточний стан середовища на певний момент часу та семантичну подібність між елементами СІО [4];
- вибір серед наявних СІО тих, що не суперечать вимогам користувача;
- пошук критеріїв оцінки СІО з використанням знань з онтології ПрО;
- визначення рівня значущості кожного окремого критерію на поточний момент співставлення на основі експертних оцінок та евристик ПрО з використанням методу аналізу ієрархій;
- визначення кількісної оцінки кожного СІО на основі сукупності критеріїв.

Висновки та перспективи подальшої роботи

Співставлення СІО, що мають подібну структуру, є необхідною складовою в аналізі та порівнянні СІО побудовані на основі різних онтологій та мають різну структуру. В цьому випадку спочатку потрібно: 1. виконати вирівнювання базових онтологій ПрО й знайти відповідності між їх концептами та відношеннями; 2. знайти в СІО, що співставляються, підмножини з подібною структурою; 3. співставити такі підмножини за розглянутим вище алгоритмом.

Запропонований підхід ієрархічного агрегатного оцінювання колективів, орієнтований на інтегроване багатокритеріальне прийняття рішень в умовах, коли сам набір критеріїв залежить від особливостей задачі і може бути створений на основі знань ПрО, але в різні моменти часу відносна значущість цих критеріїв може динамічно змінюватися. Цей підхід у перспективі планується розширити за допомогою засобів і методів керування знаннями, інтелектуального аналізу даних і машинного навчання, які використовуються для отримання компетентних знань. Запропоновані теоретичні моделі та методи можуть застосовуватися для підтримки таких актуальних для воєнного стану задач, як керування ризиками, швидке налаштування промисловості на випуск важливої продукції, відновлювальне будівництво, динамічна адаптація команд, організацій, колективів з багаторівневою ієрархічною структурою (проєктні групи, групи дослідників, формування військових підрозділів, експертні комісії, медичні групи швидкого реагування) до виконання важливих оперативних завдань при відсутності достатніх компетенцій, навичок та досвіду [5]. Таке комплексне співставлення дозволяє виявити слабкі та сильні сторони різних команд та їх учасників для конкретних завдань, внести корективи до їх складу або провести додаткове навчання.

Список посилань

1. Siegemund, K. Contributions To Ontology-Driven Requirements Engineering : dissertation to obtain the academic degree Doctoral engineer (Dr.-Ing.) / K. Siegemund – Dresden: Technischen Universität Dresden, 2014. – 236 p.
2. Fitsilis, P. Ontologies for Software Project Management: A Review / P. Fitsilis, V. Gerogiannis, L. Anthopoulos // Journal of Software Engineering and Applications. – 2014. – №7. – P. 1096- 1110.
3. Resnik P. Using information content to evaluate semantic similarity in a taxonomy. Proc. of the14th international joint conference on Artifcial intelligence, V. 1. 1995, P.448–453.
4. Рогушина Ю.В., Гладун А.Я. Використання онтологічних знань для багатокритеріального співставлення складних інформаційних об`єктів // Проблеми програмування, 2022, №2-3.
5. Гладун А.Я., Рогушина Ю.В., Лєсаж М. Онтологічний підхід до агрегованого оцінювання роботи колективів з багатьма рівнями ієрархії // Information Technology and Security (Інформаційні технології та безпека) НТУ-КПІ, 2022, 10(1). С.83–97.

ВРАХУВАННЯ ОСОБЛИВОСТЕЙ ЕКСПЕРТНИХ ЗНАНЬ В СИСТЕМАХ ОРГАНІЗАЦІЙНОГО УПРАВЛІННЯ ПРИ ФОРМУВАННІ ІНФОРМАЦІЙНОГО РЕСУРСУ

В.В. Юзефович[1], О.В. Андрійчук[1], Є.О. Цибульська[1],
Ніколай Стоянов[2]
[1]Інститут проблем реєстрації інформації Національної академії наук України, Київ, Україна
[2]Болгарський оборонний інститут імені професора Цвєтана Лазарова, Софія, Болгарія

uzefv71@gmail.com, andriichuk@ipri.kiev.ua, evts68@gmail.com, n.stoianov@di.mod.bg

В роботі розглянуто особливості отримання та обробки експертних знань, характерних для систем організаційного управління (СОУ), які в подальшому складають інформаційну основу баз даних та баз знань інформаційного ресурсу СОУ. Виділено п'ять різновидів аналітичної діяльності в рамках СОУ. Зазначено найбільш характерні для них когнітивні викривлення та запропоновано можливі шляхи їх зменшення або зниження їх впливу на якість інформаційного ресурсу.

Ключові слова: аналітик, експертні знання, когнітивні викривлення, інформаційний ресурс, система організаційного управління.

Вступ

Аналіз різних визначень системи організаційного управління показує, що її ключовою ознакою є наявність впливу в межах об'єкта та суб'єкта управління людського фактору [1]. Людина (аналітик, фахівець, експерт) в таких складних системах є важливим джерелом даних та знань, які, зокрема, використовуються при формуванні інформаційного ресурсу СОУ. Процесу отримання та обробки такої інформаціїпритаманні певні особливості, у тому числі - наявність характерних для неї когнітивних викривлень [2], зменшення яких або зниження їх впливу на якість вмісту інформаційного ресурсу є актуальним завданням, оскільки якість інформації (знань) прямо впливає на якість управлінських

рішень.Для СОУ характерною є ієрархічна обробка інформації, де фахівці нижніх рівнів ієрархії, які є важливим першоджерелом інформації про стан системи та середовища її існування, є аналітиками, що не в повній мірі відповідають вимогам до експерта. Крім того, враховуючи доцільність у багатьох випадках скорочення циклу управління та з інших причин – експертні оцінювання в даній предметній області переважно здійснюються лише на верхніх рівнях системи управління. Зазначені особливості СОУ потребують додаткового аналізу та врахування можливого додаткового прояву когнітивних викривлень інформації.

Отже, метою даної роботи є аналіз особливостей отримання та обробки експертних знань в системах організаційного управління для визначення можливих шляхів зменшення впливу когнітивних викривлень на якість вмісту інформаційного ресурсу СОУ.

Особливості експертних знань

Виділимо наступні загальні особливості експертних знань, які, на наш погляд, слід враховувати при формуванні інформаційного ресурсу СОУ.

1) Відповідно до дослідження [2], при вирішенні задач експертного оцінювання наявні когнітивні викривлення даних та знань можуть суттєво вплинути на його результат.

2) Психофізіологічні обмеження людини [2] не дозволяють їй одночасно оперувати більш ніж 9-ма об'єктами.

3) Експертне оцінювання потребує часових затрат, є дорогим процесом і тому краще його застосовувати лише за наявності нагальної необхідності. Бажано, за можливості, використовувати раніше побудовані бази знань (БЗ), їх фрагменти та шаблони проектування [3].

4) Експерт може пропускати сеанси оцінювання, не відповідати на певні питання через брак часу, сильну зайнятість, втому, небажання, тощо. Тому слід забезпечувати можливості обробки неповної експертної інформації [4].

5) Потреби в узагальнених та систематизованих знаннях поетапно деталізуються (процес декомпозиції) на більш низьких рівнях ієрархії формування знань і, навпаки, деталізована інформація на шляху «знизу вверх» агрегується для забезпечення інформаційних потреб її споживачів [5].

Формування інформаційного ресурсу СОУ з урахуванням особливостей експертних знань

Діяльність фахівців СОУ, які на різних її рівнях залучені до процесів формування, обробки (агрегування) та аналізу інформації (даних), будемо називати аналітичною діяльністю. Для досягнення мети даної роботи необхідно виділити та проаналізувати види аналітичної діяльності, характерні для різних ієрархічних рівнів СОУ.

У [1] представлено узагальнену схему формування інформаційного ресурсу деякої СОУ, яка, після зміщення акценту в сторону визначення в ній місця аналітичної діяльності персоналу, буде мати вигляд, показаний на рис. 1. Аналіз рис. 1 показує, що в рамках СОУ може бути виділено принаймні п'ять різновидів аналітичної діяльності та, відповідно, груп фахівців.

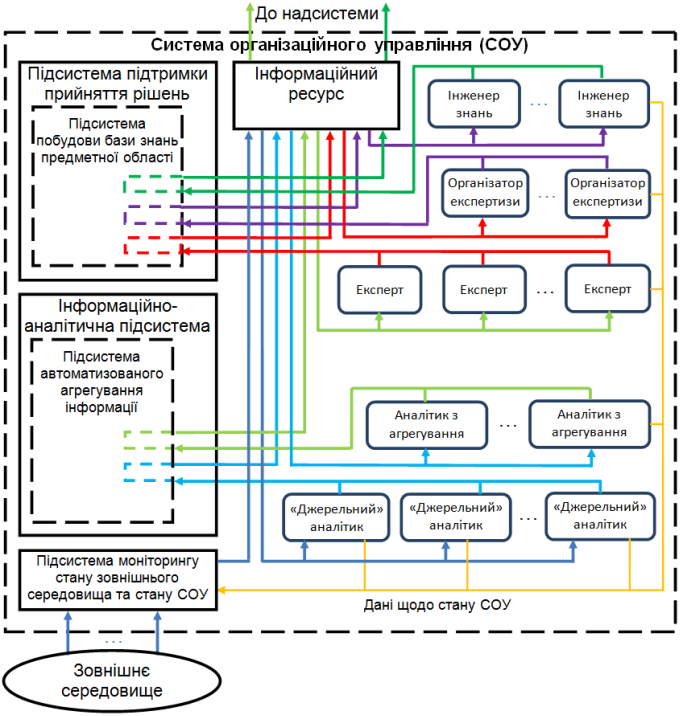

Рис. 1. Схема формування інформаційного ресурсу СОУ

«Джерельний» аналітик – це аналітик, що здійснює свою аналітичну діяльність у межах одного елемента СОУ, яка базується на даних безпосередніх спостережень за показниками, що характеризують стан відповідного елемента системи, та/або даних моніторингу факторів зовнішнього середовища. У результаті такий аналітик формує судження щодо стану елемента системичи стану окремих процесів та явищ, які спостерігаються у зовнішньому середовищі функціонування СОУ, спираючись на власні уявлення та досвід. Як правило, такі судження формуються шляхом визначення одного з можливих станів елементу (або складової зовнішнього середовища) із множини заздалегідь визначених станів. Судження «джерельних» аналітиків переважно базуються на об'єктивних даних спостереження (моніторингу), однак вихідні дані містять суб'єктивні викривлення інформації у результаті когнітивної діяльності.

До другої групи можна віднести ***аналітиків з агрегування***, які здійснюють аналітичну діяльність шляхом узагальнення (інтеграції) даних (інформації) від «джерельних» аналітиків. Вони формують судження щодо стану різних підсистем або груп елементів СОУ і, відповідно, здебільшого базуються на суб'єктивних даних. Оскільки у рамках різних СОУ можуть виділятися різні підсистеми та можлива наявність ієрархічних зв'язків між ними, очевидно, що діяльність таких аналітиків формує ієрархічну процедуру агрегування інформації із отриманням суджень щодо стану підсистем (стану зовнішнього середовища) різного ступеня узагальнення. У результаті, суб'єктивні викривлення, які містяться у вихідних судженнях (результатах аналітичної діяльності), можуть зростати за рахунок накладання власних когнітивних викривлень на когнітивні спотворення «джерельних» аналітиків. Як підсумок, інформаційна невизначеність щодо дійсного стану явищ, що аналізуються, може зрости.

Зниження когнітивних викривлень інформації, що характерні для двох визначених груп фахівців, може забезпечуватися максимально досяжною формалізацією процесу її отримання та агрегування, а також спільною обробкою дублюючої інформації від декількох аналітиків. Для такої обробки інформації може ефективно застосовуватися інструментарій, розроблений в рамках теорії нечітких множин та нечіткої логіки, який спеціально створювався для роботи із судженнями експертів. Другим шляхом є зменшення навантаження на одного аналітика з урахуванням психофізіологічних обмежень людини через раціональний розподіл функціональних завдань між аналітиками.

Інженери знань та організатори експертизи перед побудовою БЗ мають попередньо знайомиться з відповідною предметною областю. Для цього переважно використовуються доступні відкриті джерела (Інтернет, статті, лекції), а також спеціалізовані засоби, наприклад, системи контент-моніторингу. При цьому може спрацювати когнітивне упередження «ефект ілюзії правди» [2]. Таким чином, в СОУ недоцільно формувати експертні оцінки, базуючись безпосередньо на даних популярних джерел. Більш обґрунтованим при здійсненні експертного оцінювання є залучення формалізованих методів обробки неповної експертної інформації. Крім того, організатор експертизи проводить підбір групи експертів, що мають достатній рівень компетентності у предметній області, а в подальшій роботі з побудови БЗ враховується рівень компетентності експерта у кожному з питань експертизи. При цьому може спрацювати когнітивне упередження - ефект Даннінга-Крюгера [2].

Експерти здійснюють декомпозицію предметної області, розділяють цілі на підцілі, визначають критерії та фактори, які безпосередньо впливають на результат експертизи. Цьому процесу характерні наступні когнітивні викривлення: ефект Рінгельмана, ефект «фокусування», «помилка того, хто вижив», ефект «велосипедного сараю» [2]. Для їх уникнення при роботі експертної групи доцільно використовувати системи розподіленого збору та обробки експертної інформації. Експертне оцінювання має за мету достовірно визначити ступені переваги між альтернативами/критеріями. Можливе проведення безпосереднього оцінювання в балах та парні порівняння. При цьому слід уникати наступних когнітивних упереджень: помилка розрізнення, ефект прив'язки, ефект контрасту [2]. Також слід враховувати дослідження Джорджа Міллера, присвячені обмеженості короткочасної пам'яті людини.

Висновки

У роботі здійснено аналіз особливостей процесу отримання та обробки експертних знань в системах організаційного управління. Виділено п’ять характерних для СОУ видів аналітичної діяльності, розглянуто їх відмінності та зазначено джерела виникнення когнітивних викривлень даних та знань, притаманні кожному виду діяльності. Запропоновано декілька можливих шляхів зниження когнітивних викривлень та зменшення їх впливу на якість інформаційного ресурсу СОУ.

Перелік посилань

1. Volodymyr Yuzefovych, Yevheniia Tsybulska, Oleh Andriichuk One Approach to Formation of Common Information Space Information Resource in Organizational Management Systems / CEUR Workshop Proceedings (ceur-ws.org). Vol. 3241 urn:nbn:de: 0074-3241-0. Selected Papers of the XXI International Scientific and Practical Conference "Information Technologies and Security" (ITS 2021), Kyiv, Ukraine, December 9, 2021. - pp. 215-224. [http://ceur-ws.org/Vol-3241/paper20.pdf]

2. О.В. Андрійчук, В.В. Циганок, С.В. Каденко, Я.В. Порпленко, Мінглей Фу, О.О. Власенко Підходи до врахування когнітивних упереджень експертів в системах підтримки прийняття рішень / Інформаційні технології і безпека. Матеріали XXI Міжнародної науково-практичної конференції ІТБ-2021, 09 грудня 2021, Київ, Україна. - К.: ООО "Инжиниринг", 2021. - С. 225-229.

3. O. Andriichuk, V. Tsyganok, D. Lande, O. Chertov Y. Porplenko Usage of Decision Support Systems for Modelling of Conflicts During Recognition of Information Operations // In: Tagarev T., Atanassov K.T., Kharchenko V., Kacprzyk J. (eds) Digital Transformation, Cyber Security and Resilience of Modern Societies. Studies in Big Data, vol 84. Springer, Cham. 2021. https://doi.org/10.1007/978-3-030-65722-2_30

4. Oleh Andriichuk, Vitaly Tsyganok, Sergii Kadenko, Reduction of the Number of Expert Pair-wise Comparisons During Decision Support Using AHP / Proceedings of the 16th International Symposium on the Analytic Hierarchy Process, Pitsburg, USA (web conference), 3 – 6 December, 2020. pp. 1-3.

5. В.Г. Тоценко Методы и системы поддержки принятия решений. Алгоритмический аспект – К.: Наук. думка, 2002. – 382 с.

КІЛЬКІСНЕ ОЦІНЮВАННЯ РІВНЯ КРИТИЧНОСТІ ЕЛЕМЕНТІВ ДЛЯ СТІЙКОГО ФУНКЦІОНУВАННЯ ОРГАНІЗАЦІЙНОЇ СИСТЕМИ

Гнатієнко Григорій Миколайович[1], Бабенко Тетяна Василівна[1]

[1] Київський національний університет імені Тараса Шевченка, вул. Володимирська, 64/13, 01601 Київ, Україна

Вступ

Критичною частиною будь-якої інфраструктури або критичним елементом (вузлом, ланкою, підсистемою) (КЕ) будь-якої системи є такі елементи, від функціонування яких залежить стан та функціональність системи. Оскільки схема взаємодії елементів може бути представлена багатозв'язним графом, то критичними вузлами такого графа є ті вузли графа, вихід яких з ладу призведе до зниження функціональної стійкості (ФС) або до втрати зв'язності значної кількості вузлів графа.

Критичні елементи відіграють важливу роль на усіх етапах забезпечення ФС організаційної системи (ОС) та у всіх аспектах прийняття рішень щодо управління ФС ОС. Для визначення рівня впливу КЕ на функціонування системи слід ввести поняття рівня критичності (РК). РК – це числовий показник, який відображує інтегральний вплив КЕ на якість функціонування системи в цілому.

Аспекти системи

Для аналізу даних та прийняття рішення щодо рівня (класу, показників) критичності елементів слід здійснити моніторинг характеристик, які можуть бути розділені на кілька аспектів [1] (груп, напрямків, шарів, блоків). Доцільно розглядати такі аспекти елементів складної слабкоструктурованої системи, які є вузлами графа, що моделює конкретну систему у деякій предметній області: α_1 – вплив на ресурси системи; α_2 – потоки, які контролює елемент; α_3 – вплив на прийняття рішень в системі; α_4 – сукупність управлінських впливів на елементи системи; α_5 – необхідність реагування на множину запитів, які обробляються системою; α_6 – управління доступами до важливих аспектів діяльності системи; α_7 – множина ребер графа, які забезпечують інформування вершин; α_8 – сукупність

ребер графа, які моделюють процедури погодження прийняття рішень у ОС.

Шкали вимірювання

Особливістю моделювання критичності елементів складної системи є те, що дані, які використовуються для аналізу та порівняння критичності, вимірюються або оцінюються у різних шкалах. Зокрема через це, агрегування даних, об'єднаних у аспекти щодо кожного елемента системи, може бути здійснене у різних формах: визначено деякі дискретні рівні критичності; обчислено функцію належності критичності елемента нечіткій множині; визначено інтервали значень критичності кожного елемента; обчислено фіксовані значення критичності елементів у метризованих шкалах.

Ресурси системи

Розглянемо докладніше показники (характеристики, параметри), які можуть бути включені до деяких аспектів. При визначенні впливу елемента системи на її ресурси, слід здійснити поглиблений аналіз ресурсів системи: α_{11} – людські (трудові) ресурси; α_{12} – фінансові ресурси; α_{13} – матеріальні ресурси; α_{14} – нематеріальні ресурси; α_{15} – інформаційні ресурси.

Потоки у системі

Слід також здійснити аналіз потоків, які контролює елемент, або має на них суттєвий вплив: α_{21} – фінансові потоки; α_{22} – матеріальні потоки; α_{23} – інформаційні потоки; α_{24} – сервісні потоки.; α_{25} – управлінські впливи.

Визначення рівня критичності

Для визначення РК елементів слід розглядати інтегровані потоки, до яких мають відношення елементи системи. Кількісні показники, що характеризують потоки, на які має вплив кожен елемент системи, обчислюються формулою $f_2\left(\alpha_{21}^i,...,\alpha_{25}^i\right)$, $i \in I$. Зрозуміло, що значення цієї функції для деяких елементів може бути нульовим: $\exists i,\ i \in I:\ f_2\left(\alpha_{21}^i,...,\alpha_{25}^i\right)=0$.

Загальний вплив кожного елемента на прийняття рішення у ОС (аспект α_3), інтенсивність сукупності управлінських впливів (аспект α_4

), необхідність реагування на множину запитів, які обробляються системою, (аспект α_5) та рівень повноважень елемента щодо управління доступами до важливих аспектів діяльності системи (аспект α_6) можуть бути визначені, наприклад, у порядкових шкалах шляхом експерного оцінювання. Характеристики ребер графа, які забезпечують інформування вузлів, (аспект α_7) та характеристики погодження прийняття рішень у системі, (аспект α_8) можуть бути визначені у кількісних шкалах.

Інтегральне (агреговане, узагальнене) значення РК кожного $i-$го елемента системи $i \in I,$ є суперпозицією функцій, аргументами яких є зазначені аспекти:

$$F^i = F\left(f_j\left(\alpha_{jl}^i\right)\right),$$

де $i \in I,\ \ j = 1,...,8,\ \ l = 1,...,s_j,$ (1)

s_j – кількість критеріїв кожного аспекта $\alpha_j,\ j = 1,...,8.$

Висновки

З огляду на описану модель виду (1), можна зробити висновок, що вплив КЕ на ФС ОС має бути щонайменше на один чи на кілька порядків більшим від впливу лінійних елементів. Цей факт підтверджується відповідними обчислювальними експериментами.

Список літератури

1. Гнатієнко Г.М., Снитюк В.Є. Експертні технології прийняття рішень: Монографія. – К.: ТОВ «Маклаут», 2008. – 444 с.

THE USE OF STOCHASTIC CODES TO ENSURE SECURITY OF INFORMATION TRANSMISSION

Bohdan Zhurakovskyi[1][0000-0003-3990-5205], Mykhailo Klymash[2] [0000-0003-2867-1482], Liubov Berkman[3][0000-0002-6772-1596], Sergei Otrokh [4][0000-0001-9008-0902], Oleksandr Chumak [5][0000-0003-3876-8149], Kristina Pravdokhina [6][0000-0001-7376-0156]

[1, 4, 6]National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine
[2]Lviv Polytechnic National University, Lviv, Ukraine
[3]State University of Telecommunications, Kyiv, Ukraine
[5]Military Diplomatic Academy named after Yevheniy Bereznyak, Kyiv, Ukraine
zhurakovskiybyu@tk.kpi.ua, mykhailo.m.klymash@lpnu.ua, berkmanlubov@gmail.com, 2411197@ukr.net, a_ch_i@ukr.net, kristpravda@gmail.com

All known algorithms of cryptographic systems, which have the property of interference resistance, are based on codes that detect and correct errors. The study of stochastic codes for their use in algorithms of cryptographic systems are proposed in this work. For stochastic codes, there is a "copy" decoding algorithm, when for two or more values of a code block of a stochastic code, including (n, n - 1) - a code with the detection of errors that are the same during their transmission, it is possible to carry out joint decoding of the extended code with bug fixes. At the same time, the number of errors corrected in a block of extended code significantly exceeds the number of errors corrected in total in each block. To simplify the comparative analysis, a recalculation was made from the given value Pq to the value of the probability of twisting the binary symbol P_0 for different degrees of error grouping, which is estimated according to the Portov model with the coefficient a.

Keywords: stochastic code, cryptographic protection, probabilities of distortion in the channel, error-correcting codes, error bursts, decoding mode

Introduction

The introduction of modern information technologies into the everyday life of society has caused problems in ensuring information security. One of the solutions to this problem is the widespread use of cryptography. At

the moment, strict technological requirements are imposed on cryptographic algorithms not only in terms of stability, but also in terms of speed and ease of implementation [1].

Increased speed requirements are associated with the need to maintain high performance of automated systems after some protection mechanisms are built into them. The simplicity of hardware implementation is necessary to reduce the cost of encryption tools, which will contribute to their mass application and wider possibilities of embedding in portable equipment. Due to the specificity of information presentation in digital devices, block ciphers are of greatest interest.

All known algorithms of cryptographic systems, which have the property of interference resistance, are based on codes that detect and correct errors [2].

Thus, interference-resistant crypto-algorithms have high requirements for hardware implementation, operation speed, memory, crypto-security and jam-resistance, which directly depend on the properties of the applied code algorithms that use artificial redundancy.

Statement of research problem

The construction and properties of error-correcting stochastic codes

In the 1980s, work was started on the creation of a new design of codes that fit into the structure of existing data transmission networks, with the aim of increasing the technical and economic effect when transmitting information through communication channels of different quality. As a result of the work, designs and algorithms for coding (decoding) of q-th stochastic codes with error correction based on forming binary codes for communication channels of different quality were created [3].

The basis of the code is $q = 2^{32}$, it means that the binary length of the q-symbol is 32 bits, the number of such symbols in the block is n and n i k [3].

The probability of error [3] decoding stochastic q-codes does not depend on the type and nature of twists and is mainly related to the value of q as in the error detection mode (n, n - 1) - ode (with one redundant symbol), and in error correction mode. With the selected base q, the probability of an error after decoding does not exceed any type of twists

$$P_{errors} < q^{-1} = 2^{-32} < 10^{-9} \qquad (1)$$

The number of corrected errors t is related to the code distance d of the original binary code by the ratio
t = d - 2 and approximately corresponds to the number of corrected errors of the Reed-Solomon code with the same parameters n i k.

Decoding (and encoding) uses only binary operations with q- th symbols [4].

Furthermore, for the code (16,15) at $q = 2^{32}$ he number of binary encoding (decoding) operations is 16 per block of length 16*32 = 512 bits [5].

The value of the probability of successful decoding of the code block $P_r(1)$ rom the first transmission and the effective speed, calculated by the formula:

$$R_{ef} = \frac{k \times N_r}{n \times N_t} \qquad (3)$$

N_r i N_t – are the number of received and transmitted blocks, respectively.

For stochastic codes, there is a "copy" decoding algorithm, when for two or more values of a code block of a stochastic code, including (n, n - 1) - a code with the detection of errors that are the same during their transmission, it is possible to carry out joint decoding of the extended code with error correction. At the same time, the number of errors corrected in the block of extended code significantly exceeds the number of errors corrected in total in each block, for example, if the source code corrects t = 2 errors , then with 2-x repetitions of the source block in the extended block at least 6 twisted q-ic symbols are corrected, with three repetitions - at least 10 symbols, etc. At the same time, the guarantee of the reliability of the decoded information is preserved.

The copy decoding mode is most promising in simplex radio channels, especially with low quality of the communication channel, including with intense radio interference, as well as in duplex channels with joint decoding of not previously decoded and repeated blocks.

Temporal ("pace") characteristics depend both on the effective speed of the R_{ef}, and, at the same speed, on the probability of the block being delivered from the first (second, etc.) transmission [3].

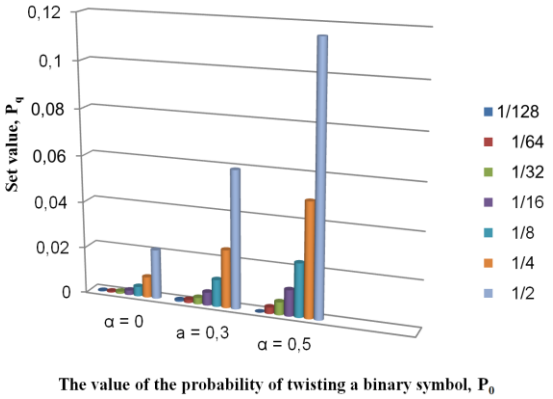

Fig.1. Channel quality for different degrees of error grouping with coefficient α

Comparative characteristics of stochastic codes with error correction, and obtained results of hardware and software tests

During the tests, the values of the probability of twisting in the channel of the q-th symbol P_q [were set twisted randomly on average every second (1/2), every fourth (1/4), etc. q-th symbol]. To simplify the comparative analysis, the calculation from the given value P_q o the value of the probability of twisting the binary symbol P_0 for different degrees of grouping of errors is estimated according to the Purtov model with the coefficient a (0 – independent errors, 0.3 – weak grouping in the leading channel, 0.5 – strong grouping in the radio channel.

Conclusions

The obtained results reveal the possibility of using codes with natural redundancy in information systems of various purposes, in which there are strict requirements for the security of the processed information in conditions of noise of communication channels, as well as for the hardware in terms of minimizing its size, cost and energy consumption.

References

1. Mustafa Al-Bassam, Alberto Sonnino, Vitalik Buterin, Ismail Khoffi, Fraud and Data Availability Proofs: Detecting Invalid Blocks in Light Clients. Financial Cryptography and Data Security: 25th

International Conference, FC 2021, Virtual Event, March 1–5, 2021, Revised Selected Papers, Part II, Mar 2021, Pages 279–298. https://doi.org/10.1007/978-3-662-64331-0_15

2. 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.

3. Mykyta Moshenchenko, Bohdan Zhurakovskyi, Vadym Poltorak, Andrii Bondarchuk, and Nataliia Korshun, Optimization Algorithms of Smart City Wireless Sensor Network Control, Paper Proceedings of the Cybersecurity Providing in Information and Telecommunication Systems II. Volume II co-located with International Conference on Problems of Infocommunications. Science and Technology (PICST 2021), Kyiv, Ukraine, October 26, 2021, Vol-3188, pp. 32–42. 2021. - Resource access mode:http://ceur-ws.org/Vol-3188/paper4.pdf

4. S. Toliupa, L. Berkman, S. Otrokh, B. Zhurakovskyi, V. Kuzminykh, H. Dudarieva. Formation of shift index vectors of ring codes for information transmission security, CEUR Workshop Proceedings, 2021, 3241, pp. 248–257.

5. Richard E. Blahut. Theory and Practice of Error Control Codes Hardcover – January 1, 1984

ІНФОРМАЦІЙНА БЕЗПЕКА У ХМАРНИХ ТА ГІБРІДНИХ СЕРЕДОВИЩАХ ДЛЯ РЕАЛІЗАЦІЇ СЕРВІСУ ІНТЕРНЕТ-ГОЛОСУВАНЬ

Володимир Фльонц
Фізико-технічний інститут, Національний технічний університет України "КПІ ім. Ігоря Сікорського"

Вступ

Створення Національного антикорупційного бюро України, а згодом,Указ Президента України від 15 травня 2015 року № 272/2015 вперше ввели в правове поле термін інтернет-голосування як офіційнийспосібпроведення конкурсу з формування органів цивільного контролю.

Перша спроба провести інтернет-голосування 26 травня 2015 року завершилась передчасно, внаслідок численних повідомлень про злам системи голосування, що в підсумку призвело до відміни його результатів. Однією з ключових проблемзастосування інструментів інтернет-голосувань залишається забезпечення інформаційної безпеки.Реалізація сервісівінтернет-голосувань з використанням фізичних серверів потребує значних ресурсів, що може бути не виправдано з урахуванням відносно невеликої кількості таких голосувань. Використання для цього хмарних та гібрідних середовищ може значно спростити реалізацію сервісів голосувань, без шкоди забезпечення вимог інформаційної безпеки[1].

Актуальність

Інтернет-голосування як офіційний інструмент для проведення конкурсу з формування громадських рад використовують п’ять міністерств та інші центральніорган виконавчої влади. Станом на жовтень 2022 року проведено понад 20 офіційних інтернет-голосувань.Найближчим часом кількість таких голосувань лише зростатиме.

Аналіз проблеми

Не дивлячись на типовість задачі, які виконує інтернет-голосування,досі не існує спільних вимог до побудови систем інтернет-голосувань.Міністерства та інші центральні органи виконавчої владивикористовують різні положення при формуванні

громадських рад.Відсутністьцентралізованого рішення для проведення інтернет-голосувань призводить до необхідності кожному органу влади самостійно підтримувати власнусистему голосувань. Враховуючи необхідність в голосуваннях в середньому лише раз на рік, очевидним стає використання тимчасової оренди хмарних сервісів для розгортання системи голосування. Водночас, закон України «Про хмарні послуги» [2] набрав чинності лише в серпні 2022 року. Більшість підзаконних актів в сфері інформаційної безпеки для хмарних середовищ наразі лише заплановані для прийняття. Дослідження ризиків інформаційної безпеки в хмарних середовищах виносить на перший план додаткові ризики, пов'язані з паралельним існуванням значної кількості віртуальних машин [3].

Реалізація сервісуінтернет-голосувань

Нова версія сервісу інтернет-голосувань, що проходить тестову експлуатацію в Національному антикорупційному бюро України, спроектована із урахуванням попереднього досвіду та виявлених ризиків інформаційної безпеки. Сервіс інтернет-голосувань реалізований на мікросервіснійбагаторівневій архітектуріз використанням інтегрованої системи електронної ідентифікації тарознесенням серверів, що обробляють персональні дані громадян та ведуть підрахунок голосів.

Відкритий протокол, що забезпечує можливість незалежної перевірки ходу та результатів голосування захищено з використанням ланцюжку геш-сум, подібному до того, що використовується при створенні ланцюжків блокчейн.Сервіс інтернет-голосувань спроектований та реалізований з урахуванням вимог інформаційної безпеки щодо ізоляції сервісів в хмарному середовищі, розгортання та експлуатації всієї інформаційної системи на хмарній інфраструктурі.

Висновки

В результаті аналізу виявлено такі проблеми забезпечення інформаційної безпеки при створенні сервісу інтернет-голосувань у хмарних та гібрідних середовищах:

- нормативно-правова база, що регулює питання забезпечення інформаційної безпеки для хмарних середовищнедостатньо розвинена, підзаконні акти, що передбачені Законом України «Про хмарні послуги», напрацьовуються прямо зараз і заплановані до прийняття найближчим часом;

- реалізація сервісу онлайн голосувань у хмарному середовищі дозволяє забезпечити виконання поставлених вимог інформаційної безпеки із використанням меншого бюджету порівняно з реалізацією на фізичних (bare-meal) серверах.

Перелік посилань

1. NIST Референтна (еталонна) архітектура хмарних обчислень (англ.) [Електронний ресурс]: Режим доступу: https://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication500-292.pdf
2. Закон України «Про хмарні послуги», 17 лютого 2022 року № 2075-IX. [Електронний ресурс]: Режим доступу: https://zakon.rada.gov.ua/laws/show/2075-20
3. Д. Нікішин, О. Федюшин, Ризики інформаційної безпеки в хмарних сервісах, «GlobalCyberSecurityForum 2019», Харків, 2019, с. 80-81.

ЗАСТОСУВАННЯ СЕМАНТИЧНИХ WIKI-ТЕХНОЛОГІЙ ДЛЯ ІНФОРМАЦІЙНОЇ ПІДТРИМКИ ВІДКРИТОЇ НАУКИ В УКРАЇНІ

Рогушина Юлія Віталіївна
Інститут програмних систем НАНУ, Київ, Україна,
ladamandraka2010@gmail.com

Запропоновано використання семантичних Wiki-технологій для розширення засобів пошуку та навігації у відкритих наукових ресурсах на змістовному рівні. Проаналізовано відповідність такої технологічної платформи принципам FAIR.
Ключові слова: відкрита наука, семантичні Wiki-технології, FAIR, онтологія.

Вступ

Відповідно до розпорядження Кабінету Міністрів України від 8 жовтня 2022 р. затверджено Національний план щодо відкритої науки, в якому передбачаються заходи, що спрямовані на забезпечення відкритого доступу до наукових результатів та науково-технічної інформації, отриманої під час здійснення наукових досліджень.

Для розвитку і поширення політики відкритої науки в Європі, Європейська Комісія запропонувала створити Європейську хмару відкритої науки (European Open Science Cloud - EOSC), яка поєднує дослідницькі інфраструктури для підтримки наукових досліджень [1]. Ця мережа сховищ даних та сервісів дозволяє дослідникам знаходити, використовувати і комбінувати набори даних, забезпечуючи основу для створення нових інструментів з обробки даних, наприклад на основі штучного інтелекту.

До глобальних цілей EOSC входить розробка інструментів та сервісів, що дозволяють дослідникам знаходити, отримувати доступ, повторно використовувати та комбінувати результати наукової діяльності. Для досягнення таких цілей ставляться практичні завдання, що стосуються підтримки принципів FAIR у поданні результатів досліджень та забезпечення спільної платформи для доступу до цифрових об'єктів FAIR, які можуть містити як дані, так і програмне забезпечення та сервіси.

Для цього EOSC застосовуються спільна розробка стандартів для предметної галузі та впровадження практики Open Science шляхом

взаємодії з дослідницькими спільнотами, а також адаптація для використання науковими співтовариствами таких технічних компонентів екосистеми FAIR, структури метаданих для цифрових об'єктів FAIR.

Одним з напрямів розбудови відкритої науки є забезпечення відкритого доступу до наукових результатів та науково-технічної інформації шляхом використання хмарних платформ для проведення фундаментальних та прикладних досліджень широкими колами науковців. Крім широкого спектру спеціалізованих сервісів аналізу даних, які потрібні для досліджень, однією з базових функцій такої системи є *пошук інформації* на змістовному рівні як одна з базових складових інфраструктури відкритої науки. Це має забезпечити повторне використання результатів раніше виконаних досліджень в різних галузях, результати експериментів та розрахунків тощо.

Пошукові сервіси мають не тільки забезпечувати доступ до різних типів інформації (природномовних документів і мультимедійних даних) і підтримувати здобуття знань зі знайденої інформації, але й надавати можливість знаходження складних інформаційних об’єктів, що відповідають вимогам користувачів. Прикладами таких об’єктів є:

- різноманітні обчислювальні засоби та ресурси, що можуть бути отримані в тимчасове користування,
- наукові колективи територіально розподілених дослідників, які об’єднуються для виконання спільного проекту;
- створення колективних монографій або тематичних видань, публікація статей у найбільш релевантних виданнях.

Тому важливою часткою досліджень є створення системи метаданих для опису не тільки інформаційних ресурсів відкритої науки, але й усіх її об’єктів та суб’єктів. Така модель метаданих, що використовує знання із зовнішніх онтологій предметних областей, організаційних онтологій, онтологій професій та компетенцій (таких, як ESCO), а також різноманітні національні та міжнародні стандарти, має стати основою для семантичної розмітки та пошуку всіх типів інформації, що генерується, зберігається та передається інфраструктурою відкритої науки.

Передумовами для цього є створення портальних версій науково-технічних ресурсів з відкритим доступом – електронних бібліотек, галузевих довідників та енциклопедій, архівів та наукометричних баз даних. Але, крім наявності доступу та підтримки загальноприйнятих стандартів метаданих (таких як Dublin Core), ефективне використання таких ресурсів потребує також більш специфічних

засобів навігації та пошуку в цих ресурсах, які дозволили б задовольнити персоніфіковані інформаційні потреби дослідників, які описувалися б у терміносистемах відповідних галузей знань та наукових напрямків.

Великий потенціал для цього мають семантично розмічені Wiki-ресурси, які можуть інтегруватися із зовнішніми онтологіями тих предметних областей (ПрО), що цікавлять користувачів та відповідають напрямкам їх наукових досліджень. Онтології мають надати спільну терміносистему, однозначно визначивши зміст використаних понять та відношень між ними, та зробити спілкування членів наукового співтовариства більш ефективним. Одним з аспектів цієї роботи має бути побудова онтологічних моделей конкретних спільнот, проектів та задач, яка дозволить створювати та аналізувати релевантні набори Big Data.

Онтології як основа семантизації інформаційних ресурсів

Семантизація інформаційних ресурсів полягає у зв'язуванні елементів контенту з формально описаними поняттями певної ПрО. На сьогодні у Web-орієнтованих інтелектуальних система найчастіше для цього використовуються онтології, подання яких базується на відкритих стандартах Semantic Web. Створення семантичної розмітки наукових інформаційних ресурсів України на основі таких онтологій не тільки забезпечує інтеграцію у світове/європейський науковий інформаційний простір, але й забезпечує розширені сервіси семантичного пошуку у різних галузях науки (приклади такого пошуку демонструє портальна версія Великої української енциклопедії e-ВУЕ – vue.gov.ua) [2]. Саме на основі досвіду розробки цього порталу ми пропонуємо створити семантичний Wiki-довідник, в якому розробниками будуть забезпечені шаблони для введення типових об'єктів наукових досліджень (статті, монографії, проекти тощо), а контент зможуть вводити й уточнювати самі науковці.

У середовищі Semantic MediaWiki онтології можуть не тільки використовуватися як базис семантичної розмітки, але й поповнюватися на основі сторінок з такою розміткою відповідно до запитів користувачів. Це дозволяє генерувати персоніфіковані онтології, що відображають уявлення про світ різних наукових спільнот та окремих дослідників. Щоб інтегрувати ці персоніфіковані уявлення, доцільно розробляти методи та програмні

засоби співставлення та вирівнювання таких онтологій зі спільною онтологією наукового співтовариства. Це дозволить, з одного боку, явно представляти відмінності поглядів різних наукових груп та напрямків, а з іншого – забезпечить ефективний пошук потрібної інформації незалежно від того, яку саме терміносистему обирають для метаданих розробники окремих ресурсів.

Приклади використання такого підходу – інтеграція національних та міжнародних рамок кваліфікацій, формування багатонаціональних територіально розподілених дослідницьких колективів, пошук навчальних курсів дистанційної освіти та інтеграція ринку освітніх послуг із ринком праці.

Семантичні Wiki-технології та FAIR

Технологія Wiki та її семантичне розширення відповідають базовим принципам "FAIR Guiding Principles for scientific data management and stewardship" (Findable, Accessible, Interoperable, Reusable). Використання FAIR спрямоване на оптимізацію повторного використання даних та їх об'єднання в різних задачах. Для цього дані та метадані докладно описують із застосуванням набору однозначних і релевантних атрибутів, і ця інформація подається відповідно до стандартів певного тематичного наукового співтовариства. Важливо, що основні принципи FAIR не потребують стандартизації чи конкретної технології їхньої підтримки, але визначають умови для створення даних для озер даних з урахуванням функціональності їх пошуку, приступності, сумісності і повторного використання.

Згідно FAIR, функції пошуку, здобуття і представлення даних реалізують не користувачі, а інформаційна система. При цьому мова йде не тільки про самі дані і метадані, але і про алгоритми й інструменти керування ними. До розробки підходів до керування науковими даними залучаються всі зацікавлені сторони: науково-дослідні організації й окремі вчені; оператори баз даними і видання, що публікують наукові статті і результати експериментів; організації, що фінансують ці наукові дослідження; виробники програмного забезпечення й інструментів обробки даних; компанії, що надають послуги з аналізу й інтерпретації даних. Важливо, що в коло зацікавлених сторін також включаються самі обчислювальні системи (алгоритми обробки даних) як самостійний об'єкт — у залежності від їхнього рейтингу приймається рішення про включення обчислювального методу до конфігурації [3].

Семантичні розширення Wiki-технологій, наприклад, на основі Semantic MediaWiki, – це потужне рішення для спільного редагування даних та їх метаописів, створення різних довільних наборів властивостей в шаблонах цих метаописів, з одночасним поданням їх як в машинно-оброблюваній формі, так і формі придатній для розуміння людиною, що в результаті надає можливість оперувати цими даними, автоматизовано керувати, проводити аналіз, публікувати.

Вбудовані можливості Semantic MediaWiki забезпечують завантаження файлів різного формату і додавання до них метаданих з різним набором атрибутів, які можливо змінювати, доповнювати. Інформаційні ресурси, які будуються в цьому середовищі, відповідають усім вимогам FAIR до відкритих даних великого обсягу [4].

Даний підхід апробовано в процесі створення бази знань портальної версії Великої української енциклопедії (vue.gov.ua), яка є джерелом інтероперабельних знань, що придатні для повторного використання в інших інтелектуальних застосуваннях.

Висновки

Важливою особливістю запропонованого підходу є те, що, хоча розробка моделі метаданих виконується спеціалістами з онтологічного інженірінгу у співпраці з експертами предметних областей, але використання технології Wiki дозволяє вільно використовувати цю модель широким колам користувачів (як для розмітки, так і для пошуку інформації) без спеціальних знань та без потреби у встановленні додаткового програмного забезпечення.

Створення семантичного Wiki-довідника наукових ресурсів України є не альтернативою забезпечення відкритого доступу до наукових публікацій та результатів досліджень, а його доповненням для розширення функціоналу пошуку в цих ресурсах з використанням знань щодо окремих областей досліджень

Список посилань

1. FAIR_data. https://en.wikipedia.org/wiki/FAIR_data.
2. Rogushina J.V., Grishanova I.J. Ontological methods and tools for semantic extension of the media WIKI. Proc. of the 12th International Scientific and Practical Conference of Programming (UkrPROG 2020),CEUR Workshoop Proceedings, 2021, Vol-2866, P.61-73. http://ceur-ws.org/Vol-2866/ceur_61-73Rogushina6.pdf

3. The FAIR Guiding Principles for scientific data management and stewardship. Available from: https://www.nature.com/articles/sdata201618.

4. Рогушина Ю.В., Гришанова І.Ю. Дослідження принципів, моделей та методів парадигми менеджменту наукових даних FAIR для аналізу метаданих big data // Проблеми програмування, 2021, №4. С.26-35. http://doi.org/10.15407/pp2021.04.026.

FINDING THE MINIMUM NUMBER OF SOURCES OF COMPROMISE INDICATORS TO COVER THE MAXIMUM SET

Sydorenko Kateryna Leonidivna
National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine
purplemikate@gmail.com

The paper considers the model of IOCs obtained from the open database OTX. A formal representation is proposed. The problem of minimizing IOC sources is solved as a set cover problem(SCP) and a maximum coverage problem(MCP).

Keywords: Indicators of compromise (IOCs), set cover problem, maximum coverage problem

Introduction

An indicator of compromise (IOC) is data that can help detect malicious activity, obtained after investigating a cyber incident. Alien Labs Open Threat Exchange (OTX), an open threat analysis database, is used. IOC sets, called "Pulses", are published on OTX. Each IOC has details and a list of all sources. This data is used as an input to the SIEM system. Dependence on input data, large amount of data and excessive maintenance requirements are current SIEM problems[1].

Problem statement

The task is to find such a minimal list of data sources to ensure both maximum coverage and a minimum amount of data.

All IOCs and Pulses must be represented as sets. The OTX is constantly updated, so sources are limited time. U is the set of all IOCs. $U = [C_1, C_2, \dots C_i]$, where $i \in \mathbb{N}$, C_i - every single IOC of any type. IOC sources are pulses P_i. $P_i \in U$. $P = [C_1, C_2, \dots , C_n]$, $n \in \mathbb{N}$. Each IOC C_i consists of characteristics X and Y. $C_i = [x_1, \dots , x_m, y_1, \dots , y_k]$, where $x_m \in X$, $y_k \in Y$, $m, k \in \mathbb{N}$. The characteristics of Y are such that: $y_i \in C_1$, $y_i \in C_2$, then $C_1 = C_2$. So when the Y-type characteristic values match, the IOCs are considered as the same. The characteristics of X are such that: $x_i \in C_1$, $x_i \in C_2$, then $C_1 \neq C_2$.

Problem solution

If characteristics of each individual IOC are ignored, the formalization of data obtained from OTX is suitable for solving the SCP and the MCP. Characteristics are needed only to find unique IOCs. The set X can be described for different types: *Domain = {domain_name, ip}. Hostname = {hostname, domain_name, ip}.* If *domain_name* $\in$ *Hostname* and *Domain*, then $C_1=C_2$, only hostname is stored. The domain can be normal, and malicious or suspicious activity will be performed from a specific hostname.

IPv4 = {ip, reverse_DNS}. If *reverse_DNS = hostname : hostname* $\in$ *Hostname* and *reverse_DNS* $\in$ *IPv4*, then $C_1=C_2$, only hostname is counted. IPv6 logic is similar to IPv4. *Filehash = {md5, sha1, sha256}*. If $a \in$ *Filehash2, Filehash1*, where a is the value of any field, then $C_1=C_2$.

First solution is IOCs filtering. One way to reduce the amount of data is to filter IOCs. The proposed metric is the number of pulses. First, the "normal" number of IOC pulses of a certain type is determined statistically and IOCs with a smaller number are not added in U. The task comes down to solving the SCP[2]. The main purpose of which is to return the smallest number of subsets (sources), so set U is fully covered.

If the number of subsets is defined at the beginning, the weighted version of MCP[3] will be used. The task is to find a coverage that has the biggest weight sum of selected IOCs. Each IOC has a weight. To find the smallest amount of bytes, weight should be inversely proportional to the amount of memory occupied by each IOC. To extract only relevant IOCs, the value should be proportional to the number of pulses.

If the maximum amount of memory that can be allocated for storing logs is used as a budget, the budgeted MCP[3] will be used. The price in this case is the amount of memory occupied by the source, and the weight is the amount of pulses. Each subset has a price, each element has a weight. The task is to maximize the weight's sum of all elements so that the price of all subsets does not exceed the budget.

Conclusions

The paper proves that the task can be represented as an NP problem and can be solved using the Set cover problem and the Maximum coverage problem. The future aim is to implement the described solutions and to make a comparison of time spent and coverage received.

References

1. J. Henderson, Why a SIEM Won't Solve All Your Problems: 5 Common Issues and How to Avoid Them, 2021. URL: https://redcanary.com/blog/common-siem-issues/.

2. V. Vijay, Approximation Algorithms, Springer-Verlag Berlin Heidelberg GmbH, Atlanta, USA, 2003.

3. C. Chekuri, A. Kumar, Maximum Coverage Problem with Group Budget Constraints and Applications, Springer, 2004.

URL: https://doi.org/10.1007/978-3-540-27821-4_7

IMPLEMENTATION OF A DUAL VOTING SYSTEM IN THE ELECTORAL SYSTEM OF UKRAINE

Novikov Oleksii Volodymyrovych
NTUU "KPI", Institute of physics and technologies, Kyiv, Ukraine
alexeyn08@gmail.com

Methods of implementing low-tech electronic systems for the conduct of elections are considered in work. The purpose of the article is to present a new solution based on the analysis of existing systems, which can be a superstructure of the already existing electoral system of Ukraine.
Keywords: dual voting system, elections, homomorphic cryptosystems, low-tech, CEC, polling stations

Introduction

The introduction of new election technologies that could prevent falsifications is an urgent problem. In his work, Ronald Rivest envisages a mathematical protocol as a technological add-on to a traditional paper bulletin. The solution proposed in this work is based on a preliminary analysis of the *Scantegrity* and *Wombat* methods and their main shortcomings [1].

Implementation of the dual system in Ukrainian realities

In the modern realities of Ukraine, the algorithm of using a mixed network will not be suitable due to the fact that some political parties are very interested in falsifications. The method of using scanning machines without additional security mechanisms is not reliable. Therefore, it was decided to create a layer using a partially homomorphic cryptosystem to protect the election results from statistical falsification.

Election settings

The CEC is creating a pair of RSA2048 RSAKEY cryptosystem keys that will be responsible for the electronic election process. A pair of EcDSA keys (Ed448). The set of simple numbers that will be responsible for a certain candidate (for the final part of the calculation). And also the database of ballots, where the following information will be stored for the ballot: two random prime numbers (510-512 bits): the first will be the main unique key to this ballot in the CEC database, the second will be the second unique number of the ballot; signature of the main unique number of the bulletin with the key of the CEC (Ed448); place under the voice; an

array of random numbers *0, 1, . . . , n*, where n is the number of candidates on the ballot. This array is necessary in order to assign a random number to each candidate on the ballot. So upon receiving the voting results from the polling station, it is possible to immediately check the correctness of the ballot entry; key (128 bits) to obtain the second prime number. It is necessary to encrypt the secondary unique number of the ballot – so that no one can recover it prematurely because this will provide a potential opportunity for selling votes.

After creating this structure, the CEC sends the entire structure to be printed, except for the reserved place under the vote and the second simple number of the ballot. Optical terminals that will be located at the polling station are also being prepared. Each is assigned its own pair of Ed448 system keys. Keys are also provided by a member of the precinct election commission, who usually signs paper documents.

Election day

During the elections, a couple of QR codes will be added to the regular ballot, the correctness of which can be checked during voting.

The detached part that the voter takes with him after casting his own vote. Also, after he has decided on a candidate, he will need to remember the corresponding number of the candidate on the ballot, or write it down on a removable part.

Closing of polling stations (end of elections)

After closing the polling stations, the terminal verifies the main unique number of the ballot with his signature. If they match, it adds the ballot to the file with the following structure: the main unique number of the ballot; the number of the candidate voted for; a separate statistical list of votes cast for candidates.

After checking the statistical data issued by the terminal and the commission, the file is signed with the keys of the commission members and transferred to the server, where it immediately becomes publicly accessible. Based on the main unique number of the ballot, the CEC retrieves the corresponding structure of the ballot from the database and decodes the value.

After decoding, the CEC creates a new structure: $x_{1i} * x_{2i} * y_i = Mi$, where x_{1i} and x_{2i} — the main and secondary unique numbers of the ballot, respectively; y_i — the candidate for whom a vote was cast; M_i — general message. For each polling station, a product of messages is formed, *Prod* $M = \prod_{i=1,\ n} M_i$, where n — the number of ballots. Then, from this

product, it is easy to select the number of votes for one or another candidate (it is enough to divide by the prime number of the candidate).

In order to find out whether the vote was counted — devide *Prod M* on the main unique number of the ballot. It is very easy to check the integrity of this course of elections, because the following rule is fulfilled due to the homomorphism of the cryptosystem:

$$Sign_{RSAKEY}(ProdM) = \prod_{i=1,n} Sign_{RSAKEY}(Mi).$$

Conclusions

This work proposes a method of introducing a dual system into the electoral processes of Ukraine with minimal changes to existing processes in the country.

References

1. Going from bad to worse: from Internet voting to blockchain voting / Park S., Specter M., Narula N., and Rivest R. L. // Journal of Cybersecurity. – 2021. – 02. – Vol. 7, no. 1. https://academic.oup.com/cybersecurity/article-pdf/7/1/tyaa025/36276521/tyaa025.pdf.

ПРОПОЗИЦІЯ МЕТОДУ СТВОРЕННЯ ВІДБИТКУ APT-ГРУПИ

Козленко Олег Віталійович Наконечна Юлія Володимирівна
НТУУ “КПІ імені Ігоря Сікорського”, Київ, Україна,
education.kozlenko@gmail.com, nakonechna.yu@gmail.com

Вступ

За останні роки можна побачити багато випадків компроментації корпоративних та державних інформаційних систем групами зловмисників, які діють на основі особистих мотивів або яких пов’язують з іноземними спецслужбами. Дослідження поведінки кіберзлочинців за роки спостережень були проаналізовані та скомпоновані у вигляді бази знань компанією MITRE, яка отримала назву MITRE ATT&CK. На даний час функції та/або застосунку визначення відбитків для ідентифікації зловмисних хакерських груп не існує і тому проблема є актуальною. Головним завданням данного дослідження є створення функції відбитку діяльності (fingerprint) активних на даний час груп з можливістю швидко порівнювати без допомоги програмних застосунків різницю між відбитками.

Створення відбитку групи

Для реалізації завдання були обрані таки алгоритми:

1. MD5 – для гешування секції domain;
2. SHA-224 – для гешування секції softwares;
3. SHA-256 – для гешування секції techniques;

Поле “name” не буде проходити гешування і буде використано як ідентифікатор відбитку.

Для прикладу отримання відбитку візьмемо інформацію щодо APT-16. Розіб’ємо її на частини:

- "name": "APT16"" – буде використаний як ідентифікатор
- domain": "enterprise-attack"
- "techniques": [{"techniqueID": "T1584.004", "title": "Compromise Infrastructure: Server"}]
- "softwares":[{"softwareID":"S0064", "title": "ELMER"}]

Отже результуюча функція відбитку групи APT-16 матиме вигляд:

3468f22d494c679d74f38e463221fb83fa8d0f27353f543a94d241e6676
26510bc69be316c3d223d614acd1ca8f3b5ff27f226b666c5e69e8375164f0
6931753b1ca343210e22227906e27f2

По частинам даний відбиток буде представлений як:

- enterprise-attack - 3468f22d494c679d74f38e463221fb83
- [{"techniqueID": "T1584.004", "title": "Compromise Infrastructure: Server"}] - fa8d0f27353f543a94d241e667626510bc69be316c3d223d614acd1c
- [{"softwareID":"S0064", "title": "ELMER"}] - a8f3b5ff27f226b666c5e69e8375164f06931753b1ca343210e22227906e27f2

Даний підхід допоможе реалізувати одну з цілей дослідження – можливість контролювати зміни та визначати наскільки схожі відбитки різних груп без використання додаткового програмного забезпечення.

У ході роботи над дослідженням також була поставлена мета зробити утіліту, яка буде реалізовувати поставлені завдання. Усі вищезазначені методи та функціонал були реалізовані та запрограмовані на мові Rust. Результуюча утіліта знаходиться на репозиторії GitHub за посиланням https://github.com/lxldx/A2PTF

Висновки

У даному дослідження автори провели аналіз існуючих алгоритмів створення функцій цифрових відбитків, аналіз APT-груп та інформацію з відкритих джерел, насамперед з MITRE ATT&CK фреймворка, яку можна використовувати для визначення відбитку діяльності груп. Результатом роботи є представлена функція відбитку, яка складається з 3 основних частин: MD5 – для гешування секції domain, SHA-224 – для гешування секції softwares, SHA-256 – для гешування секції techniques. Даний поділ може допомогти швидко визначати, що є спільного у діяльності двох груп. Також для більш конкретного порівняння було запропановано використовувати функцію SimHash для визначення рівня подібності двох відбитків. Результатом діяльності є програмний застосунок, який дозволяє генерувати з JSON файлу відбиток діяльності групи.

Перелік посилань

1. "What Is an Advanced Persistent Threat (APT)?". Cisco. Retrieved 11 August 2019.

2. MITRE ATT&CK framework - https://attack.mitre.org/

23 Gonzalez, Joaquin Jay, III; Kemp, Roger L. (16 January 2019). Cybersecurity: Current Writings on Threats and Protection. McFarland. p. 69. ISBN 9781476674407.

4. Kleppmann, Martin (2 April 2017). Designing Data-Intensive Applications: The Big Ideas Behind Reliable, Scalable, and Maintainable Systems (1 ed.). O'Reilly Media. p. 203. ISBN 978-1449373320.

5. "Proposed Revision of Federal Information Processing Standard (FIPS) 180, Secure Hash Standard". Federal Register. 59 (131): 35317–35318. 1994-07-11. Retrieved 2007-04-26

MODEL OF DETECTION OF PHISHING URLS BASED ON MACHINE LEARNING

Kateryna Burbela[1]
[1] National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", 37 Peremohy Avenue, Kyiv, 03056, Ukraine

Phishing is one of the most common ways of obtaining personal data. The search for a solution to the issue of detecting phishing is very relevant because to minimize the damage from a phishing attack, it is necessary to detect it as early as possible. Almost every type of phishing attack uses phishing URLs. Cybercriminals use phishing URLs to try to obtain sensitive information such as usernames, passwords, or bank details. Some conventional ways of phishing URL detection are based on whitelisting and blacklisting. However, it cannot detect Zero-day attacks. Machine learning methods are used to increase detection accuracy and reduce the misjudgment ratio.
The article describes a CNN–MHSA model which uses a Convolutional Neural Network (CNN) and the MHSA combined approach for highly precise.
Keywords: Phishing, URL, Deep learning, Convolutional layer, Multi-head self-attention

Introduction

There are various phishing attacks such as spear phishing, whaling, vishing, smishing, pharming, etc. Likewise, there are various phishing detection methods based on whitelisting, blacklisting, and machine learning.

Since URLs are a component of the majority of attacks, the detection of phishing URLs is relevant. There are several strategies you can implement to protect your users and business from phishing URLs:

- Safety training
  - Domain verification;
  - Link verification;
- Use of software tools
  - Based on lists: it is understood that URLs are divided into phishing sites and legitimate, they are stored in a database containing the "to open" and "to ban" links.

- Heuristics is an advanced list technique. In this technique, website characteristics such as URLs and content are extracted and used for comparison between different sites.
- Machine Learning: Based on different data sets obtained from different website features, machine learning trains models from these data sets and validates them with different machine learning classifiers.

Classifiers help predict phishing websites before they are created, so machine learning solves the problem of zero-day phishing attacks.

Machine learning approaches

Typical machine learning methods include naive Bayes classifier, Random Forest, decision tree, and K-nearest neighbors. These techniques can block zero-day attacks, but researchers must define the classification characteristics manually. This limits the improvement of the result, as no one can guarantee that the selected features are sufficient to detect phishing sites.

There are methods using neural networks, they consist of information-processing elements that imitate brain neurons. Neural networks are divided into two types depending on the number of layers: shallow and deep neural networks.

Convolutional neural network

Convolutional Neural Networks (CNN) is a deep learning technique that works well for discovering simple patterns in data, which will then be used to form more complex patterns in subsequent layers. It provides high accuracy compared to other machine learning classifiers.

Multi-head self-attention

Multi-Head Self-Attention (MHSA) is a kind of attention mechanism that is now widely used in machine translation. MHSA calculates weights to express different importance for each characteristic, which is more suitable for phishing website detection scenarios.

Methodology

Since MHSA can dig into the inner dependency relationships between different characters in URLs, it can be applied to URL analysis, which may outperform LSTM. In the meanwhile, CNN can automatically learn URL features without human intervention. To get a good result two technologies can be assembled to combine their merits to serve phishing website detection.

The CNN–MHSA model combines two methods to detect phishing sites. In the model, URLs are treated as sentences, and the model is divided into two layers, where the first layer, the CNN layer, is used to learn the characteristics of the URLs, and the second layer, the MHSA layer, is designed to calculate the weights of the corresponding features. After getting the characteristics and characteristic weights from the weight calculators, they will be combined to get the final result.

CNN–MHSA can produce highly-precise detection results for a URL object by integrating its features and their weights. The thorough experiments demonstrate that the method achieves 99.84% accuracy, which outperforms the most used method CNN–LSTM.

Conclusions

The most used methods of phishing use URLs, so the question of detecting phishing links is important. Methods of detecting phishing sites include lists, heuristics, and machine learning.

Machine learning provides the highest percentage of accuracy because machine learning trains models from data set obtained from different website features and validate them using different machine learning classifiers.

A model named CNN–MHSA combines Convolutional Neural Network (CNN) and MHSA to improve detection accuracy. CNN–MHSA first applies CNN to automatically study URLs' features without manual intervention. Then CNN–MHSA leverages MHSA to explore these relationships and to generate weights for CNN studied features.

References

1. P.Kalaharshaa, B. M. Mehtre, Detecting Phishing Sites - An Overview, URL: https://arxiv.org/pdf/2103.12739.pdf.

2. Ahmet Kara, Multi-scale deep neural network approach with attention mechanism for remaining useful life estimation, URL: https://doi.org/10.1016/j.cie.2022.108211.

3. Chenguang Wang, Yuanyuan Chen, Exploring hybrid transformer and convolutional neural network on phishing URL detection, URL: https://doi.org/10.1016/j.knosys.2022.109955.

DYNAMIC CLASS LOADING IN THE JVM

Zholob Tetiana [1]
[1] National Aviation University, 1 Liubomyra Huzara ave., Kyiv, 0358, Ukraine

Class loaders are a powerful mechanism for dynamically loading software components on the Java platform. They are unusual in that they suport all of the following features: laziness, type-safe binding, user-defined extensibility, and multiple binding namespaces. We introduce the concept of class loaders and demonstrate some of their interesting uses. Additionally, we discuss how to maintain type safety in the presence of user-defined dynamic class loading.
Keywords: Java, JVM, ClassLoader

Introduction

In this article, we explore an important feature of the Java Virtual Machine: dynamic class loading. This is the core mechanism that provides much of the power of the Java platform: the ability to install software components at runtime. An example of a component is an aplet that is loaded into a web browser. Although many other systems 0 0 also suport some form of dynamic loading and linking, the Java platform is the only known system that includes all of the following features:

- Delayed loading. Classes are loaded on demand. Class loadi ng is delayed as long as possible, reducing memory usage and improving system response time.
- Type-safe connection. Dynamic class loadi ng should not violate Java Virtual Machine type safety. Dynamic load ing should not require additional runtime checks to ensure type safety. Additional checks during connection are acceptable because these checks are performed only once.
- User-defined class loading policy. Class loaders are first-class objects. Programmers have full control over dynamic class loading. A user-defined class loader can, for example, specify a remote location from which classes are loaded, or assign apropriate security attributes to classes loaded from a specific source.
- Multiple namespaces. Class loaders provide separatenamespaces for different software components. Forexample, the HotjavaTM browser loads applets fromdifferent sources into separate class loaders.

Theseapplets may contain classes of the same name, but theclasses are treated as distinct types by the Java virtualmachine.

In contrast, existing dynamic linking mechanisms donot support all of these features. Although most operatingsystems support some formof dynamic linked libraries, suchmechanisms are targeted toward C/C++ code, and are nottype-safe. Dynamic languages such as Lisp [13], Smalltalk[6], and Self [21] achieve type safety through additionalrun-time checks, not link-time checks.The main contribution of this paper is to provide thefirst in-depth description of class loaders, a novel conceptintroduced by the Java platform. Class loaders existed inthe first version of the Java Development Kit. Theoriginal purpose was to enable applet class loading in theHotjava browser. Since that time, the use of class loadershas been extended to handle a wider range of softwarecomponents such as server-side components (servlets) [11],extensions [10] to the Java platform, and JavaBeans [8]components. Despite the increasingly important role of classloaders, the underlyingmechanismhas not been adequatelydescribed in the literature.A further contribution of this paper is to present asolution to the long-standing type safety problem [20] withclass loaders. Early versions of the JDKcontained a serious flaw in class loader implementation.Improperly written class loaders could defeat the type safetyguarantee of the Java virtual machine. Note that the typesafety problem did not impose any immediate security risks,because untrusted code (such as a downloaded applet) wasnot allowed to create class loaders. Nonetheless, applicationprogrammerswho had the need towrite customclass loaderscould compromise type safety inadvertently. Although theissue had been known for some time, it remained an openproblem in the research community whether a satisfactorysolution exists. For example, earlier discussions centeredaround whether the lack of type safety was a fundamental limitation of user-definable class loaders, and whether wewould have to limit the power of class loaders, give uplazy class loading, or introduce additional dynamic typecheckingat runtime.

ClassLoaders

The purpose of class loaders is to support dynamic loading of software components on the Java platform. The unit of software distribution is a class. Classes are distributed us-ing a machine-independent, standard, binary representation known as the class file format 0. The representation of an individual class is referred to as a class file. Class files are produced by Java compilers, a nd can be loaded into any Java

virtual machine. A class file does not have to be stored in an actual file; it could be stored in a memory buffer, or obtained from a network stream.

Overview of ClassLoading

The Java virtual machi ne uses class loaders to load class files and create class objects. Class loaders are special objects that are defined by Java code. These are examples of the ClassLoader class shown in **Ошибка! Источник ссылки не найден.**. We have omitted the methods that are not di rectly relevant to this presentation. The ClasLoader.loadClass method takes a class name as argument, and returns a Class object that is the run-time representation of a class type.

```
class ClassLoader {
   public Class loadClass(String name);
   protected final Class defineClass(String name,
                                     byte[] buf, int off, int len);
   protected final Class findLoadedClass(String name);
   protected final Class findSystemClass(String name);
   ...
}
```

Figure 1. The ClasLoader class

Multiple Class Loaders

A Java application may use several different kinds of class loaders to manage various software components. For example, Figure 2 shows how a web browser written in Java may use class loaders.

This example shows the use of two types of class loaders: user-defined class loaders and the system class loader supplied by the Java virtual machine. User-defined class loaders can be used to create classes that originate from user-defined sources. For example, it creates an applet loader that downloads a navigation program.0 We use a separate class loader for the web browser application itself. All system classes (such as java.langString) are loaded into the systemclass loader. The Java Virtual Machine basically supports the System class loader.

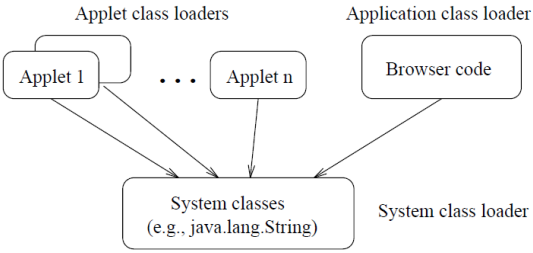

Figure 2. Class loaders in a web browser

The arrow s in the figure indicate the delegation relationship betwen class loaders. A class loader L1 can ask another loader L2 to load a class C on its behalf. In such case, L1 delegates C to Applets and application loaders, for e xample, delegate all systemclasses to the systemclass loader. All system classes, as a result, are shared among the application and the applets.0 This is desirable because type safety would be vi olated if, applet and system code, for example, had a different notion of what the type java.langString was.

Delegating class loaders allow us to ma intain namespace separation while still sharing a common set of classes.0 A class type is uniquely determined by the combination of the class loader and class name, in the Java virtual machine. Application class loaders and applet delegate to the system class loader. This guarantes that all system class types are unrepeatable, such as java.langString. A class named Cloaded in applet 1 is supposed to a different type from a class named Cin applet 2, n the other hand. They are defined by di fferent class loaders, though these two classes have the same name. These two classes can be absolutely unrelated, in fact. They may have di fferent fields or methods, for example.

Because of applets are loaded in separate class loaders, classes from one applet cannot interfere with classes in another. This is crutical in guaranteing Java platform security.0 Applets cannot acess the classes used to implement the browser, first of all bec ause the browser resides in a separate class loader. Applets are only allowed to acess, exposed in the system classes, the standard Java API.

The Java vi rtual machine starts up by using created the application class loader to load the initial browser class. Application execution starts in the public class method void main(String[]) of the initial class. The invocation of this method drives all further execution. Execution of instructions may cause loading of additional classes. The browser also

creates additional class loaders for downloaded applets, in this application.0

The garbage collector unloads applet classes that are no longer referenced. Each class object contains a reference to its defining loader; each class loader re fers to all the classes it defi nes. This means that, from the garbage collector's point of view, classes are strongly connected with their defining loader.0 Classes are unloaded when their defining loader is garbage-collected.

Aplications of ClassLoaders

Reloading Classes

It is often desirable to upgrade software components in a long-running application such as a server. The upgrade must not require the application to shut down and restart.

On the Java platform, this ability translates to reloading a subset of the classes already loaded in a running virtual machine. It corresponds to the schema evolution 0 problem, which could be rather difficult to solve in general. Here are some of the difficulties:

- There may be live objects that are instances of a class we want to reload. These objects must be m igrated to conform to the schema of the new class.0 For example, if the new version of the class conta ins a different set of instance fields, we must somehow map the existing set of instance field values to fields in the new version of the class.
- Simi larly, we may have to map the static field values to a different set of static fields in the reloaded version of the class.
- The application may be executing a method that belongs to a class we want to reload.

We show how it is sometimes possible to bypass them using class lo aders.0 By organi zing software components in separate class loaders, programmers can often avoid dealing with schema evolution. Instead, new classes are loaded by a separate loader.

Figure 3 illustrates how a Server class can dynamically redirect the service requests to a new version of the Service class. The key technique is to load the server class, old service class, and new service class into separate class loaders. For example, we can define Server using the MyClassLoader class introduced in the last section.

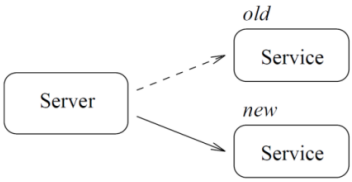

Figure 3. Class Server redirects to a new version of Service class

```
class Server {
   private Object service;

   public void updateService(String location) {
      MyClassLoader cl = new MyClassLoader(location);
      Class c = cl.loadClass("Service");
      service = c.newInstance();
   }

   public void processRequest(...) {
      Class c = service.getClass();
      Method m = c.getMethod("run", ...);
      m.invoke(service, ...);
   }
}
```

The Server.processRequest method redirects all incoming requests to a Service object stored in a private field. It uses the Java Core Reflection API 0 to invoke the “run” method on the service object. In addition, the Server.updateService method allows a new version of the Service class to be dynamically loaded, replacing the existing service object. Callers of update Service supply the the location of the new class files. Further requests will be redirected to the new object referenced to by service.

To make reloading possible, the Server class must not directly refer to the Service class:

```
class Server {
        private Service service;

        public void updateService(String location) {
                MyClassLoader cl = MyClassLoader(location);
                Class c = cl.loadClass("Service");
                service = (Service) c.newInstance();
        }
}
```

Once the Server class resolves the symbolic reference to a Service, it will contain a hard link to that class type. An already-resolved reference cannot be changed.0 The type conve rsion in the last line of the Server.updateService method will fail for new versions of Service returned from the class loader.

Reflection allows the Server class to use the Service class without a direct reference. Alternatively, Server and classes can share a common interface or superclass:

```
class Server {
        private ServiceInterface service;
        public void updateService(String location) {
                MyClassLoader cl = new MyClassLoader(location);
                Class c = cl.loadClass("Service");
                service = (ServiceInterface) c.newInstance();
        }
        public void processRequest( ... ) {
                service.run(...);
        }
}
```

Dispatching through an interface is typically more efficient than reflection. 0 The interface type itself must not be reloaded, because the Server class can refer to only one Ser viceInterface type. The get ServiceClass method must return a class that implements the same ServiceInterface every time.

After we call the updateService method, all future requests will be processed by the new Service class. The old Service class, however, may not have finished processing some of the earlier requests. Thus two Service classes may coexist for a while, until all uses of the old class are complete, all references to the old class are dropped, and the old class is unloaded.

Maintaining Type-safe Linkage

The loadClass method may return different class types for a given name at different times. To maintain type safety, the virtual machine must be able to consistently obtain the same class type for a given class name and loader. Consider, for example, the two references to class X in the following code:

```
class C {
    void f(X x) { … }
    ...
```

```
    void g() { f(new X()); }
}
```

If C's class loader were to map the two occurrences of X into different class types, the type safety of the method call to f inside g would be compromised.

The virtual machine cannot trust any user-defined loadClass method to consistently return the same type for a given name. Instead, it internally maintains a loaded class cache. The loaded class cache maps class names and initiating loaders to class types. After the virtual machine obtains a class from the loadClass method, it performs the following operations:

- The real name of the class is checked against the name passed to the loadClass method. An error is raised if loadClass returns a class that does not have the requested name.
- If the name matches, the resulting class is cached in the loaded class cache. The virtual machine never invokes the loadClass method with the same name on the same class loader more than once.

Temporal Namespace Consistency

According to the analysis of foreign publications [21-27], the most commonly used passwords at present are passwords associated with personal data, which include birthdays, passport numbers, phone numbers, e-mail addresses, etc. Using the proposed approach, we will establish how the dependence ofpersonal information on the number of categories used in the password.

The study poses and solves the problem of determining an effective password based on the theory of planning and processing the results of experiments.

This analysis of the results obtained allows us to conclude that password efficiency is primarily affected by password duration. The effect of mutual influence, based on the initial conditions of the experiment, practically does not have a significant effect on the listed condition of reliability, which made it possible to exclude this regressor from the equation.

Solution

A straightforward solution to the type-safety problem is to uniformly use both the class's name and its defining loader to represent a class type in the Java virtual machine. The only way to determine the defining

loader, however, is to actually load the class through the initiating loader. In the example in the previous section, before we can determine whether C.f's call to Delegated.g is type-safe, we must first load Spoofed in both L1, and L2, and see whether we obtain the same defining loader. The shortcoming of this approach is that it sacrifices lazy class loading.

Our solution preserves the type safety of the straightforward approach, but avoids eager class loading. The key idea is to maintain a set of loader constraints that are dynamically updated as class loading takes place. In the above example, instead of loading Spoofed in L1 and L2, we simply record a constraint that Spoofed L1 = Spoofed L2. If Spoofed is later loaded by L1, or L2, we will need to verify that the existing set of loader constraints will not be violated.

What if the constraint Spoofed L1 = Spoofed L2 is introduced after Spoofed is loaded by both L1 and L2. It is too late to impose the constraint and undo previous class loading.

We must therefore take both the loaded class cache and loader constraint set into account at the same time. We need to maintain the invariant: Each entry in the loaded class cache satisfies all the loader constraints. The invariant is maintained as follows:

- Every time a new entry is about to be added to the loaded class cache, we verify that none of the existing loader constraints will be violated. If the new entry cannot be added to the loaded class cache without violating one of the existing loader constraints, class loading fails.
- Every time a new loader constraint is added, we verify that all loaded classes in the cache satisfy the new constraint.If a new loader constraint cannot be satisfied by all loaded classes, the operation that triggered the addition of the new loader constraint fails.

Let us see how these checks can be applied to the previous example. The first line of the C.f method causes the virtual machine to gnerate the constraint Spoofed L1 = Spoofed L2.

If L1 and L2, have already loaded the Spoofed class when we generate this constraint, an exception will immediately be raised in the program. Otherwise, the constraint will be successfully recorded. Assuming Delegated.g loads Spoofed L2 first, an exception will be raised when C.f tries to load Spoofed L2 later on.

Conclusions

We have presented the notion of class loaders in the Java platform. Class loaders combine four desirable features: lazy loading, type-safe linkage,

multiple namespaces, and user extensibility. Type safety, in particular, requires special attention. We have shown how to preserve type safety without restricting the power of class loaders.

Class loaders are a simple yet powerful mechanism that proven to be extremely valuable in managing software has components.

References

1. E. Gilles Barbedette. Schema modifications in the LISP O2 persistent object-oriented language. In European in ts a of ng Conference on Object-Oriented Programming

2. Sonya E. Keene. Object-Oriented Programming in Common Lisp. Addison-Wesley

3. Tim Lindholm and Frank Yellin. The Java Virtual Machine Specification. Addison-Wesley, Reading, Massachusetts.

4. Oberon Microsystems, Inc. Component Pas- hell. cal Language Report.

5. Drew Dean. Private communication.

6. Drew Dean. The security of static typing with dynamic ne ger linking. In Fourth ACM Conference on Computer and Communications Security, pages 18-27.

7. A. Goldberg and D. Robson. Smalltalk-80: the Language and Its Implementation.

8. James Gosling, Bill Joy, and Guy Steele. The Java Language Specification. Addison-Wesley, Reading, Mas- sachusetts.

9. JavaSoft, Sun Microsystems, Inc. JavaBeans Components API for Java.

10. JavaSoft, Sun Microsystems, Inc. Reflection.

11. JavaSoft, Sun Microsystems, Inc. The Java Extensions Framework.

12. JavaSoft, Sun Microsystems, Inc. Servlet

13. Jim Roskind. Private communication. Matrix design notes. Saraswat.

14. Vijay Saraswat. Java is not type-safe.

15. David Ungar and Randall Smith. SELF: The power of simplicity. In Proc. of the ACM Conf. on Object-Oriented Programming, Systems, Languages and Applications

ЗМІСТ

Національна академія наук України
Інститут проблем реєстрації інформації НАН України

ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ ТА БЕЗПЕКА

Матеріали XXII Міжнародної науково-практичної конференції

Випуск 22

Підп. до друку 20.12.2022. Формат 60x84 1/16. Папір офс. Гарнітура Times.
Спосіб друку – ризографія. Ум. друк. арк. 6,5. Обл.-вид. арк. 24,36. Наклад 100 пр.
Зам. № 15-200.

ТОВ "Інжиніринг" ISBN 978-966-2344-85-1