ҚАЗАҚСТАН РЕСПУБЛИКАСЫНЫҢ ҒЫЛЫМ ЖӘНЕ ЖОҒАРЫ БІЛІМ МИНИСТРЛІГІ
МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РЕСПУБЛИКИ КАЗАХСТАН
MINISTRY OF SCIENCE AND HIGHER EDUCATION OF THE REPUBLIC OF KAZAKHSTAN

ХАЛЫҚАРАЛЫҚ АҚПАРАТТЫҚ ЖӘНЕ КОММУНИКАЦИЯЛЫҚ ТЕХНОЛОГИЯЛАР ЖУРНАЛЫ

МЕЖДУНАРОДНЫЙ ЖУРНАЛ ИНФОРМАЦИОННЫХ И КОММУНИКАЦИОННЫХ ТЕХНОЛОГИЙ

INTERNATIONAL JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGIES

2025 (21) 1

ақпан - наурыз

ISSN 2708–2032 (print)
ISSN 2708–2040 (online)

БАС РЕДАКТОР:

Исахов Асылбек Абдиашимович — басқарма төрағасы, Халықаралық ақпараттық технологиялар универси тетінің ректоры, есептеу теориясы саласындағы математика бойынша PhD докторы, “Компьютерлік ғылымдар және информатика”бағыты бойынша қауымдастырылған профессор (Қазақстан)

БАС РЕДАКТОРДЫҢ ОРЫНБАСАРЫ:

Колесникова Катерина Викторовна — техника ғылымдарының докторы, Халықаралық ақпараттық технологиялар университеті, «Ақпараттық жүйелер» кафедрасының проректоры (Қазақстан)

ҒАЛЫМ ХАТШЫ:

Ипалакова Мадина Тулегеновна — техника ғылымдарының кандидаты, қауымдастырылған профессор, «Халықаралық ақпараттық технологиялар университеті» АҚ, Ғылыми-зерттеу жұмыс департаментінің директоры (Қазақстан)

РЕДАКЦИЯЛЫҚ АЛҚА:

Разак Абдул — PhD, Халықаралық ақпараттық технологиялар университетінің профессоры (Қазақстан)

Лучио Томмазо де Паолис — Саленто университетінің (Италия) инновациялар және технологиялық инженерия департаменті AVR зертханасының зерттеу және әзірлеу бөлімінің директоры

Лиз Бэкон — профессор, Абертей университеті вице-канцлердің орынбасары (Ұлыбритания)

Микеле Пагано — PhD, Пиза университетінің профессоры (Италия)

Отелбаев Мухтарбай Отелбаевич — физика-математика ғылымдарының докторы, ҚР ҰҒА академигі, Халықаралық ақпараттық технологиялар университеті, «Математикалық және компьютерлік модельдеу» кафедрасының профессоры (Қазақстан)

Рысбайұлы Болатбек — физика-математика ғылымдарының докторы, Халықаралық ақпараттық технологиялар университеті, «Математикалық және компьютерлік модельдеу» кафедрасының профессоры (Қазақстан)

Дайнеко Евгения Александровна — PhD, қауымдастырылған профессор, Халықаралық ақпараттық технологиялар университетінің Жаһандық серіктестік және қосымша білім беру жөніндегі проректоры (Қазақстан)

Дузбаев Нуржан Токкужаевич — PhD, Халықаралық ақпараттық технологиялар университетінің Цифрландыру және инновациялар жөніндегі проректоры (Қазақстан)

Синчев Бахтгерей Куспанович — техника ғылымдарының докторы, Халықаралық ақпараттық технологиялар университетінің «Ақпараттық жүйелер» кафедрасының профессоры (Қазақстан)

Сейлова Нұргүл Абдуллаевна — техника ғылымдарының кандидаты, Халықаралық ақпараттық технологиялар университетінің «Компьютерлік технологиялар және киберқауіпсіздік» факультетінің деканы (Қазақстан)

Мухамедиева Ардақ Габитовна — экономика ғылымдарының кандидаты, Халықаралық ақпараттық технологиялар университетінің «Цифрлық трансформациялар» факультетінің деканы (Қазақстан)

Ыдырыс Айжан Жұмабайқызы — PhD, Халықаралық ақпараттық технологиялар университетінің «Математикалық және компьютерлік модельдеу» кафедрасының меңгерушісі (Қазақстан)

Шильдибеков Ерлан Жаржанович — PhD, Халықаралық ақпараттық технологиялар университетінің «Экономика және бизнес» кафедрасының меңгерушісі (Қазақстан)

Аманжолова Сауле Токсановна — техника ғылымдарының кандидаты, Халықаралық ақпараттық технологиялар университетінің «Киберқауіпсіздік» кафедрасының меңгерушісі (Қазақстан)

Ниязгулова Айгүл Аскарбековна — филология ғылымдарының кандидаты, Халықаралық ақпараттық технологиялар университетінің «Медиакоммуникациялар және Қазақстан тарихы» кафедрасының меңгерушісі (Қазақстан)

Айтмагамбетов Алтай Зуфарович — техника ғылымдарының кандидаты, Халықаралық ақпараттық технологиялар университетінің «Радиотехника, электроника және телекоммуникация» кафедрасының профессоры (Қазақстан)

Алмисреб Али Абд — PhD, Халықаралық ақпараттық технологиялар университетінің қауымдастырылған профессоры (Қазақстан)

Мохамед Ахмед Хамада — PhD, Халықаралық ақпараттық технологиялар университетінің «Ақпараттық жүйелер» кафедрасының қауымдастырылған профессоры (Қазақстан)

Янг Им Чу — PhD, Гачон университетінің профессоры (Оңтүстік Корея)

Тадеуш Валлас — PhD, Адам Мицкевич атындағы университеттің проректоры (Польша)

Мамырбаев Өркен Жұмажанұлы — Ақпараттық жүйелер саласындағы техника ғылымдарының (PhD) докторы, ҚР БҒМ ҚҰО ақпараттық және есептеу технологиялары институты директорының ғылым жөніндегі орынбасары (Қазақстан)

Бушуев Сергей Дмитриевич — техника ғылымдарының докторы, профессор, Украинаның «УКРНЕТ» жобаларды басқару қауымдастығының директоры, Киев ұлттық құрылыс және сәулет университетінің «Жобаларды басқару» кафедрасының меңгерушісі (Украина)

Белощицкая Светлана Васильевна — техника ғылымдарының докторы, доцент, Астана IT университетінің деректер жөніндегі есептеу және ғылым кафедрасының профессоры (Қазақстан)

ЖАУАПТЫ РЕДАКТОР:

Мрзабаева Раушан Жәлиқызы — «Халықаралық ақпараттық технологиялар университеті» АҚ (Қазақстан)

Халықаралық ақпараттық және коммуникациялық технологиялар журналы
ISSN 2708–2032 (print)
ISSN 2708–2040 (online)
Меншіктенуші: «Халықаралық ақпараттық технологиялар университеті» АҚ (Алматы қ.).
Қазақстан Республикасы Ақпарат және әлеуметтік даму министрлігінің Ақпарат комитетінде – **20.02.2020** жылы берілген.
№ KZ82VPY00020475 мерзімдік басылым тіркеуіне қойылу туралы куәлік.
Тақырыптық бағыты: ақпараттық технологиялар, әлеуметтік-экономикалық жүйелерді дамытудағы цифрлық технологиялар, ақпараттық қауіпсіздік және коммуникациялық технологияларға арналған.
Мерзімділігі: жылына 4 рет.
Тиражы: 100 дана
Редакцияның мекенжайы: 050040, Алматы қ-сы, Манас к-сі, 34/1, 709-кабинет, тел: +7 (727) 244-51-09.
E-mail: ijict@iitu.edu.kz Журнал сайты: https://journal.iitu.edu.kz
© Халықаралық ақпараттық технологиялар университеті АҚ, 2025
© Авторлар ұжымы, 2025

ГЛАВНЫЙ РЕДАКТОР:

Исахов Асылбек Абдиашимович — кдоктор PhD по математике в области теории вычислимости, ассоциированный профессор по направлению "Компьютерные науки и информатика", Председатель Правления – Ректор АО «Международный университет информационных технологий» (Казахстан)

ЗАМЕСТИТЕЛЬ ГЛАВНОГО РЕДАКТОРА:

Колесникова Катерина Викторовна — доктор технических наук, профессор, проректор по научно-исследо ва тель-ской деятельности Международного университета информационных технологий (Казахстан)

УЧЕНЫЙ СЕКРЕТАРЬ:

Ипалакова Мадина Тулегеновна — кандидат технических наук, ассоциированный профессор, директор департа мента по научно-исследовательской деятельности Международного университета информационных технологий (Казахстан)

РЕДАКЦИОННАЯ КОЛЛЕГИЯ:

Разак Абдул — PhD, профессор кафедры кибербезопасности Международного университета информационных тех нологий (Казахстан)

Лучио Томмазо де Паолис — директор отдела исследований и разработок лаборатории AVR департамента инноваций и технологического инжиниринга Университета Саленто (Италия)

Лиз Бэкон — профессор, заместитель вице-канцлера Университета Абертей (Великобритания)

Микеле Пагано — PhD, профессор Университета Пизы (Италия)

Отелбаев Мухтарбай Отелбайулы — доктор физико-математических наук, профессор, академик НАН РК, профессор кафедры математического и компьютерного моделирования Международного университета информационных технологий (Казахстан)

Рысбайулы Болатбек — доктор физико-математических наук, профессор, профессор кафедры математического и компьютерного моделирования Международного университета информационных технологий (Казахстан)

Дайнеко Евгения Александровна — PhD, ассоциированный профессор, проректор по глобальному партнерству и дополнительному образованию Международного университета информационных технологий (Казахстан)

Дузбаев Нуржан Токкужаевич — PhD, ассоциированный профессор, проректор по цифровизации и инновациям Международного университета информационных технологий (Казахстан)

Синчев Бахтгерей Куспанович — доктор технических наук, профессор, профессор кафедры информационных систем Международного университета информационных технологий (Казахстан)

Сейлова Нургуль Абадуллаевна — кандидат технических наук, декан факультета компьютерных технологий и кибербезопасности Международного университета информационных технологий (Казахстан)

Мухамедиева Ардак Габитовна — кандидат экономических наук, декан факультета цифровых трансформаций Международного университета информационных технологий (Казахстан)

Ыдырыс Айжан Жумабаевна — PhD, ассистент профессор, заведующая кафедрой математического и компьютерного моделирования Международного университета информационных технологий (Казахстан)

Шилдибеков Ерлан Жаржанович — PhD, заведующий кафедрой экономики и бизнеса Международного университета информационных технологий (Казахстан)

Аманжолова Сауле Токсановна — кандидат технических наук, заведующая кафедрой кибербезопасности Международного университета информационных технологий (Казахстан)

Ниязгулова Айгуль Аскарбековна — кандидат филологических наук, доцент, заведующая кафедрой медиакоммуникаций и истории Казахстана Международного университета информационных технологий (Казахстан)

Айтмагамбетов Алтай Зуфарович — кандидат технических наук, профессор кафедры радиотехники, электроники и телекоммуникаций Международного университета информационных технологий (Казахстан)

Алмисреб Али Абд — PhD, ассоциированный профессор кафедры кибербезопасности Международного университета информационных технологий (Казахстан)

Мохамед Ахмед Хамада — PhD, ассоциированный профессор кафедры информационных систем Международного университета информационных технологий (Казахстан)

Янг Им Чу — PhD, профессор университета Гачон (Южная Корея)

Тадеуш Валлас – PhD, проректор университета имен Адама Мицкевича (Польша)

Мамырбаев Оркен Жумажанович — PhD, заместитель директора по науке РГП Института информационных и вычислительных технологий Комитета науки МНВО РК (Казахстан)

Бушуев Сергей Дмитриевич — доктор технических наук, профессор, директор Украинской ассоциации управления проектами «УКРНЕТ», заведующий кафедрой управления проектами Киевского национального университета строительства и архитектуры (Украина)

Белощицкая Светлана Васильевна — доктор технических наук, доцент, профессор кафедры вычислений и науки о данных Astana IT University (Казахстан)

ОТВЕТСТВЕННЫЙ РЕДАКТОР:

Мрзабаева Раушан Жалиевна — АО «Международный университет информационных технологий» (Казахстан).

Международный журнал информационных и коммуникационных технологий
ISSN 2708–2032 (print)
ISSN 2708–2040 (online)
Собственник: АО «Международный университет информационных технологий» (г. Алматы).
Свидетельство о постановке на учет периодического печатного издания в Министерство информации и общественного развития Республики Казахстан **№ KZ82VPY00020475,** выданное от **20.02.2020 г.**
Тематическая направленность: информационные технологии, информационная безопасность и коммуникационные технологии, цифровые технологии в развитии социо-экономических систем**.**
Периодичность: 4 раза в год.
Тираж: 100 экземпляров.
Адрес редакции: 050040 г. Алматы, ул. Манаса 34/1, каб. 709, тел: +7 (727) 244-51-09.
E-mail: ijict@iitu.edu.kz Сайт журнала: https://journal.iitu.edu.kz
© АО Международный университет информационных технологий, 2025
© Коллектив авторов, 2025

EDITOR-IN-CHIEF:

Isakhov Asylbek Abdiashimovich — PhD in Mathematics specializing in Computability Theory and Associate Professor in Computer Science and Informatics, Chairman of the Board, Rector of International Information Technology University (Kazakhstan)

DEPUTY CHIEF DIRECTOR:

Kolesnikova Katerina Viktorovna — Doctor of Technical Sciences, Vice-Rector of Information Systems Department, International Information Technology University (Kazakhstan)

SCIENTIFIC SECRETARY:

Ipalakova Madina Tulegenovna — Candidate of Technical Sciences, Associate Professor, Director of the Research Department, International University of Information Technologies (Kazakhstan)

EDITORIAL BOARD:

Razaq Abdul — PhD, Professor of International Information Technology University (Kazakhstan)

Lucio Tommaso de Paolis — Director of Research and Development, AVR Laboratory, Department of Innovation and Process Engineering, University of Salento (Italy)

Liz Bacon — Professor, Deputy Director, and Deputy Vice-Chancellor of the University of Abertay. (Great Britain)

Michele Pagano — Ph.D., Professor, University of Pisa (Italy)

Otelbaev Mukhtarbay Otelbayuly – Doctor of Physical and Mathematical Sciences, Academician of the National Academy of Sciences of the Republic of Kazakhstan, Professor of the Department of Mathematical and Computer Modeling of International Information Technology University (Kazakhstan)

Rysbayuly Bolatbek — Doctor of Physical and Mathematical Sciences, Professor of the Department of Mathematical and Computer Modeling, International Information Technology University (Kazakhstan)

Daineko Yevgeniya Alexandrovna — PhD, Associate Professor, Vice-Rector for Global Partnership and Continuing Education, International Information Technology University (Kazakhstan)

Duzbaev Nurzhan Tokkuzhaevich — Candidate of Technical Sciences, Vice-Rector for Digitalization and Innovations, International Information Technology University (Kazakhstan)

Sinchev Bakhtgerey Kuspanuly — Doctor of Technical Sciences, Professor of the Department of Information Systems, International Information Technology University (Kazakhstan)

Seilova Nurgul Abdullaevna — Candidate of Technical Sciences, Dean of the Faculty of Computer Technologies and Cybersecurity, International Information Technology University (Kazakhstan)

Mukhamedieva Ardak Gabitovna – Candidate of Economic Sciences, Dean of the Faculty of Digital Transformations, International Information Technology University (Kazakhstan)

Idyrys Aizhan Zhumabaevna — PhD, Head of the Department of Mathematical and Computer Modeling, International Information Technology University (Kazakhstan)

Shildibekov Yerlan Zharzhanuly — PhD, Head of the Department of Economics and Business, International Information Technology University (Kazakhstan)

Amanzholova Saule Toksanovna — Candidate of Technical Sciences, Head of the Department of Cyber Security, International Information Technology University (Kazakhstan)

Niyazgulova Aigul Askarbekovna — Candidate of Philology, Head of the Department of Media Communications and History of Kazakhstan, International Information Technology University (Kazakhstan)

Aitmagambetov Altai Zufarovich — Candidate of Technical Sciences, Professor of the Department of Radioengineering, Electronics and Telecommunication, International Information Technology University (Kazakhstan)

Almisreb Ali Abd — PhD, Associate Professor, International Information Technology University (Kazakhstan)

Mohamed Ahmed Hamada — PhD, Associate Professor, Department of Information systems, International Information Technology University (Kazakhstan)

Young Im Choo — PhD, Professor, Gachon University (South Korea)

Tadeusz Wallas — PhD, University of Dr. Litt Adam Miskevich in Poznan (Poland)

Mamyrbayev Orken Zhumazhanovich — PhD in Information Systems, Deputy Director for Science, Institute of Information and Computing Technologies CS MSHE RK (Kazakhstan)

Bushuyev Sergey Dmitriyevich — Doctor of Technical Sciences, Professor, Director of Uдоктор технических наук, профессор, директор Ukrainian Association of Project Management UKRNET, Head of Project Management Department, Kyiv National University of Construction and Architecture (Ukraine)

Beloshitskaya Svetlana Vasilyevna — Doctor of Technical Sciences, Associate Professor, Professor of the Department of Computing and Data Science, Astana IT University (Kazakhstan)

EXECUTIVE EDITOR

Mrzabayeva Raushan Zhalievna — International Information Technology University (Kazakhstan)

«International Journal of Information and Communication Technologies»
ISSN 2708–2032 (print)
ISSN 2708–2040 (online)
Owner: International Information Technology University JSC (Almaty).
The certificate of registration of a periodical printed publication in the Ministry of Information and Social Development of the Republic of Kazakhstan, Information Committee No. KZ82VPY00020475, issued on 20.02.2020.
Thematic focus: information technology, digital technologies in the development of socio-economic systems, information security and communication technologies
Periodicity: 4 times a year.
Circulation: 100 copies.
Editorial address: 050040. Manas st. 34/1, Almaty. +7 (727) 244-51-09. E-mail: ijict@iitu.edu.kz
Journal website: https://journal.iitu.edu.kz
© International Information Technology University JSC, 2025
© Group of authors, 2025

МАЗМҰНЫ

ӘЛЕУМЕТТІК-ЭКОНОМИКАЛЫҚ ЖҮЙЕЛЕРДІ ДАМЫТУДАҒЫ ЦИФРЛЫҚ ТЕХНОЛОГИЯЛАР

АҚПАРАТТЫҚ ТЕХНОЛОГИЯЛАР

АҚПАРАТТЫҚ ҚАУІПСІЗДІК ЖӘНЕ КОММУНИКАЦИЯЛЫҚ ТЕХНОЛОГИЯЛАРҒА АРНАЛҒАН

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

СОДЕРЖАНИЕ

ЦИФРОВЫЕ ТЕХНОЛОГИИ В РАЗВИТИИ СОЦИО-ЭКОНОМИЧЕСКИХ СИСТЕМ

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

ИНФОРМАЦИОННАЯ БЕЗОПАСНОСТЬ И КОММУНИКАЦИОННЫЕ ТЕХНОЛОГИИ

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

CONTENT

DIGITAL TECHNOLOGIES IN THE DEVELOPMENT OF SOCIO-ECONOMIC SYSTEMS

INFORMATION TECHNOLOGY

INFORMATION SECURITY AND COMMUNICATION TECHNOLOGIES

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

INTERNATIONAL JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGIES
ISSN 2708–2032 (print)
ISSN 2708–2040 (online)
Vol. 6. Is. 1. Number 21 (2025). Pp. 226–244
Journal homepage: https://journal.iitu.edu.kz
https://doi.org/10.54309/IJICT.2025.21.1.015

UDC 004.056.5

RECONSTRUCTING ATTACK VECTORS USING GENETIC PROGRAMMING

Y. Churakova[1], O. Novikov[2], O. Baranovskyi[2], T.V. Babenko[3], N.Y. Askarbekova[3*]

[1]Royal Institute of Technology, Stockholm, Sweden;
[2]Blekinge Institute of Technology, Kalskrona, Sweden;
[3]International Information Technology University, Almaty, Kazakhstan.
E-mail: n.askarbekova@iitu.edu.kz

Yekatierina Churakova — PhD student, Department of Computer Science, Royal Institute of Technology
E-mail: yech22@student.bth.se. ORCID: 0009-0004-0657-095X;
Oleksii Novikov – PhD student, Department of Computer Science, Blekinge Institute of Technology
E-mail: olno22@student.bth.se. ORCID: 0000-0001-5629-5205;
Oleksii Baranovskyi – PhD, senior lecturer, Department of Computer Science, Blekinge Institute of Technology
E-mail: oleksii.baranovskyi@bth.se ORCID: 0000-0001-5629-5205;
Tatiana Babenko – Doctor of Sciences, professor, Department of Cybersecurity, International Information Technology University
E-mail: t.babenko@iitu.edu.kz. ORCID: 0000-0003-1184-9483;
Nessibeli Askarbekova – Master of Science, senior lecturer, Department of Cybersecurity, International Information Technology University
E-mail: n.askarbekova@iitu.edu.kz. ORCID: 0009-0006-6230-5063.
© Y. Churakova, O. Novikov, O. Baranovskyi, T.V. Babenko, N.Y. Askarbekova, 2025

Abstract. This paper presents a novel approach to detecting and predicting attack vectors based on genetic programming. The proposed method utilizes a genetic algorithm to evolve a set of rules that predict attack vectors over the system based on caught indicators of compromise. The generated rules are then used to identify potential attack vectors and predict how it started and how it will develop in the future. This research aims to enhance the accuracy and efficiency of existing attack reconstruction methods. The proposed approach is evaluated using real-world attack data.

Keywords: MITTRE ATT&CK, genetic programming, genetic algorithm, attack vectors, attack prediction

For citation: Y. Churakova, O. Novikov, O. Baranovskyi, T.V. Babenko, N.Y. Askarbekova. RECONSTRUCTING ATTACK VECTORS USING GENETIC PROGRAMMING//INTERNATIONAL JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGIES. 2025. Vol. 6. No. 1. Pp. 226–244 (In Eng.). https://doi.org/10.54309/IJICT.2025.21.1.015.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

ГЕНЕТИКАЛЫҚ БАҒДАРЛАМАЛУ ӘДІСІМЕН ШАБЫЛ ВЕКТОРЛАРЫН ҚАЙТА ҚҰРУ

Е. Чуракова[1], О. Новиков[2], О. Барановский[2], Т.В. Бабенко[3], Н.Е. Асқарбекова[3,*]

[1]Корольдік технология институты, Стокгольм, Швеция;
[2]Блекинге технологиялық институты, Калскрона, Швеция;
[3]Халықаралық ақпараттық технологиялар университеті, Алматы, Қазақстан.
E-mail: n.askarbekova@iitu.edu.kz

Екатерина Чуракова — PhD докторанты, Корольдік технологиялық институтының информатика кафедрасы
E-mail: yech22@student.bth.se. ORCID: 0009-0004-0657-095X;
Олексей Новиков – PhD докторанты, Блекинге технологиялық институтының информатика кафедрасы
E-mail: olno22@student.bth.se. ORCID: 0000-0001-5629-5205;
Олексей Барановский – PhD докторы, Блекинге технологиялық институтының информатика кафедрасының аға оқытушысы
E-mail: oleksii.baranovskyi@bth.se ORCID: 0000-0001-5629-5205;
Татьяна Бабенко – ғылым докторы, профессор, Халықаралық ақпараттық технологиялар университетінің киберқауіпсіздік кафедрасы
E-mail: t.babenko@iitu.edu.kz. ORCID: 0000-0003-1184-9483;
Несібелі Асқарбекова – ғылым магистрі, Халықаралық ақпараттық технологиялар университеті киберқауіпсіздік кафедрасының аға оқытушысы
E-mail: n.askarbekova@iitu.edu.kz. ORCID: 0009-0006-6230-5063.

© Е. Чуракова, О. Новиков, О. Барановский, Т.В. Бабенко, Н.Е. Асқарбекова, 2025

Аннотация. Бұл жұмыс генетикалық бағдарламалау негізінде шабуыл векторларын анықтау және болжау үшін жаңа тәсілді ұсынады. Ұсынылған әдіс ымыраға келу индикаторлары негізінде жүйедегі шабуыл векторларын болжайтын ережелер жинағын әзірлеу үшін генетикалық алгоритмді пайдаланады. Содан кейін жасалған ережелер ықтимал шабуыл векторларын анықтау және оның қалай басталғанын және болашақта қалай дамитынын болжау үшін пайдаланылады. Зерттеу шабуылдарды қайта құрудың қолданыстағы әдістерінің дәлдігі мен тиімділігін арттыруға бағытталған. Ұсынылған тәсіл нақты әлемдегі шабуыл деректері арқылы бағаланады.

Түйін сөздер: MITTRE ATT&CK, генетикалық бағдарламалау, генетика-лық алгоритм, шабуыл векторлары, шабуылды болжау

Дәйексөз үшін: Е.Чуракова, О.Новиков, О.Барановский, Т.В.Бабенко, Н.Е.Асқарбекова. ГЕНЕТИКАЛЫҚ БАҒДАРЛАМАЛАР ӘДІСІМЕН ШАБУЫЛ ВЕКТОРЛАРЫН ҚАЙТА ҚҰРУ//ХАЛЫҚАРАЛЫҚ АҚПАРАТТЫҚ ЖӘНЕ КОММУНИКАЦИЯЛЫҚ ТЕХНОЛОГИЯЛАР ЖУРНАЛЫ. 2025. Том. 6. № 1. Б. 226–244 (ағыл.тілінде). https://doi.org/10.54309/IJICT.2025.21.1.015.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

РЕКОНСТРУКЦИЯ ВЕКТОРОВ АТАК МЕТОДОМ ГЕНЕТИЧЕСКОГО ПРОГРАММИРОВАНИЯ

Е. Чуракова[1], О. Новиков[2], О. Барановский[2], Т.В. Бабенко[3], Н.Е. Аскарбекова[3*]

[1]Королевский технологический институт, Стокгольм, Швеция;
[2]Технологический институт Блекинге, Кальскруна, Швеция;
[3]Международный университет информационных технологий, Алматы, Казахстан.
E-mail: n.askarbekova@iitu.edu.kz

Екатерина Чуракова — PhD студент кафедры Компьютерных наук Королевского технологического института
E-mail: yech22@student.bth.se. ORCID: 0009-0004-0657-095X;
Олексей Новиков – PhD студент кафедры Компьютерных наук Технологического института Блекинге
E-mail: olno22@student.bth.se. ORCID: 0000-0001-5629-5205;
Олексей Барановский – PhD, старший преподаватель кафедры Компьютерных наук Технологического института Блекинге
E-mail: oleksii.baranovskyi@bth.se ORCID: 0000-0001-5629-5205;
Татьяна Бабенко – доктор наук, профессор кафедры Кибербезопасность Международного университета информационных технологий
E-mail: t.babenko@iitu.edu.kz. ORCID: 0000-0003-1184-9483;
Нессибели Аскарбекова – магистр наук, старший преподаватель кафедры Кибербезопасность Международного университета информационных технологий
E-mail: n.askarbekova@iitu.edu.kz. ORCID: 0009-0006-6230-5063.

© Е. Чуракова, О. Новиков, О. Барановский, Т.В. Бабенко, Н.Е. Аскарбекова, 2025

Аннотация. В данной статье представлен новый подход к обнаружению и прогнозированию векторов атак на основе генетического программирования. Предлагаемый метод использует генетический алгоритм для разработки набора правил, которые прогнозируют векторы атак в системе на основе обнаруженных индикаторов компрометации. Сгенерированные правила затем используются для выявления потенциальных векторов атак и прогнозирования того, как они начались и как будут развиваться в будущем. Целью исследования является повышение точности и эффективности существующих методов реконструкции атак. Предлагаемый подход оценивается с использованием реальных данных об атаках.

Ключевые слова: MITTRE ATT&CK, генетическое программирование, генетический алгоритм, векторы атак, прогнозирование атак

Для цитирования: Е. Чуракова, О. Новиков, О. Барановский, Т.В. Бабенко, Н.Е. Аскарбекова. РЕКОНСТРУКЦИЯ ВЕКТОРОВ АТАК МЕТОДОМ ГЕНЕТИЧЕСКОГО ПРОГРАММИРОВАНИЯ//МЕЖДУНАРОДНЫЙ ЖУРНАЛ ИНФОРМАЦИОННО-КОММУНИКАЦИОННЫХ ТЕХНОЛОГИЙ. 2025. Т. 6. № 1. С. 226–244 (на англ. яз.). https://doi.org/10.54309/IJICT.2025.21.1.015.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

Introduction

In recent years, the frequency and sophistication of cyberattacks have increased, presenting a significant threat to the security of organizations' information. Traditional cybersecurity measures, such as firewalls and antivirus software, have become less effective in protecting against the constantly evolving threat landscape. Cyberattacks have grown more sophisticated as adversaries leverage advanced technologies and exploit the increasing volume of data, making it challenging to identify these attacks within a vast flow of information.

Existing intrusion prevention and detection methods rely on predefined patterns, such as correlation rules used in Security Information and Event Management (SIEM) systems. However, network latency can significantly impact detection efficiency. Correlation rules are consequences of indicators (events in logs, file names, network packets, etc.) named as Indicators of Compromise (IoC). But standard approach of using predefined attacks' patterns is incapable of detecting or predicting novel attack vectors that have not been encountered before therefore absent in correlation rules set. In case of novel or complex attack detection systems SIEM or IDS can fix subset of IoCs, but are not able to generate notification due the absence of attack scenario in the rule set. Thus, there is a necessity for an on the fly generation of potential attack vector patterns based on subset of detected/discovered IoCs. Moreover, since attacks may deviate from existing patterns, the construction method should involve exploring new approaches for attack implementation and identifying the most likely ones. Predicting of the attack vectors potentially could provide several key advantages. Firstly, it could help to identify new/zero-day vulnerabilities that attackers exploit to compromise a system. Secondly, understanding the techniques employed by hackers aids in developing defense strategies that can proactively prevent attacks and mitigate risks. Additionally, predicting attack vectors facilitates swift response to emerging threats. By promptly recognizing new attack methods, SIEM or IDS could be proactively improved to prevent systems from being compromised. By anticipating potential attack methods, organizations can implement measures to mitigate damage should an attack succeed.

Any attack vectors could be represented as sequence of steps/nodes with targeting specific goal. In this case sequence generation requires satisfying constraints, such as consistency and compatibility of neighborhood nodes. General assumption that there are two main approaches to generate a list of potential attack vectors/sequences: brute force and selective generation. Brute force is not applicable due to detection time and computational constraints of any system. For selective generation, the most promising approach involves using various Machine Learning (ML) methods. Currently, ML methods are used to analyze network traffic and identify anomalies or suspicious activities, analyze malware and determine its characteristics, scan applications and systems for vulnerabilities or to identify and block unwanted email and websites, including those used for phishing. Table 1 shows the summary of machine-learning methods and cybersecurity tasks they commonly solve.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

Table 1. Various Machine Learning Methods and Their Applications

Method	Solving Tasks
Naive Bayes	Spam filtering
Support Vector Machine SVM	Intrusion detection
Random forest	Anomaly detection
Gradient Boosting	Phishing page detection
Convolutional Neural Networks CNN	Malware detection
Recurrent Neural Networks RNN	Threat prediction
Long-Short Term Memory LSTM	Attack detection
Reinforcement learning	Penetration testing

According to Table 1, Neural Networks has a wide potential of implementation for solving cybersecurity tasks by analyzing substantial amounts of data in real-time and identifying potential threats before they can cause harm. However, even these models are not without drawbacks. For example, machine-learning models require training data, and often that means using historical data. If attacks are highly diverse and change over time, models may have difficulty detecting new and previously unknown attacks, for which there is no information in the training data. In addition, when used for anomaly detection, machine-learning models can have accuracy issues, especially when attacks are difficult to distinguish from normal behavior or when there is not enough attack data for reliable training. This can lead to false positives where legitimate processes are classified as attacks or missing real attacks. It should also be noted that these solutions are able to act on a specific attack step that has already been detected, but without reconstructing its further development, or pointing to a potential «entry point» (in case the attack is not detected at the initial stage). Moreover, machine learning models are prone to the vanishing and exploding gradient problems, which can make training difficult. When reconstructing attack vectors, it is crucial to have models that can learn from both the most probable and the least probable events. MLs may struggle with this due to their sensitivity to gradient-related issues.

The problem of generation attack vectors (reconstructing the complete attack picture based on knowledge of some stages of the attacks) is a complex task because attacks can be subtle and camouflaged, and data can be incomplete or inaccurate. Considering the aforementioned factors, this paper proposes a novel method for attack vectors generation based on genetic programming.

Literature Review

Genetic programming (GP) currently has broad potential for application in cybersecurity. This is the focus of O'Reilly and Toutouh's article (Nafie et al., 2019: 311–325). The authors point to such areas of application of genetic algorithms as network defense investigation, self-adapting cyber defenses, anomaly detection, vulnerability testing, malware detection, intrusion detection, which is currently the subject of most works. The authors note that the genetic programming method is used to solve multiclass classification problems, namely, separating legitimate traffic from

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

botnets or denial of service (e.g., the article (Kayacık et al., 2009: 512–530)).

In the context of vulnerability testing, GP can be applied to create and optimize software tools and algorithms that help detect vulnerabilities in software. For example, Kayacık in the paper ”Generating mimicry attacks using genetic programming: A benchmarking study” (Laroche et al., 2009: 104–120) shows the method of applying evolutionary programming for simulating attacks and discovering vulnerabilities. Other authors, like Laroche and Zincir-Heywood propose a genetic-algorithm based method to evolve the headers of TCP-packets and modify them in a way to perform the best results for vulnerability testing (Rovito et al., 2022: 391–405).GP can be used in malware analysis to create and optimize algorithms and models that can help detect and analyze malware. In the study of Rovito, Bonin and Manzoni genetic algorithm is used as a basis for developing a method, which allows to find a bot account in Twitter social network. Some methods, proposed by scientists, focused on malware analysis for selecting executable file potentially malicious features, thus performing dynamic prepossessing of data, which later incomes to machine learning classifiers (Al-Harahsheh et al., 2021: 303–317).

Al-Harahsheh, Al-Shraideh, Al-Sharaeh are exploring a similar paradigm (Al-Sahaf et al., 2019: 487–499) in their research ”Performance of Malware Detection Classifier Using Genetic Programming in Feature Selection”.

Researchers Al-Sahaf and Welch show that genetic programming method can successfully be used to solve the well-known problem of detection phishing websites, alongside with ransom ware and spam detection. After comparison of GP based methods with other machine learning techniques (LaRoche et al., 2006: 145–158), authors show, that on the same data sets genetic algorithm performs better and significantly improves the performance of commonly used approaches. The paper (Suhaimi et al., 2019) highlights the threat of phishing attacks on the internet and the effectiveness of using a blacklist approach for detecting new attacks. The paper proposes a solution to this problem by using GP for phishing detection. The researchers conducted experiments on a dataset of phishing and legitimate sites collected from the internet and compared the performance of Genetic Programming with other machine learning techniques. The results showed that Genetic Programming was the most effective solution for detecting phishing attacks.

However, the most common application of genetic programming is to generate detection rules in intrusion detection systems: GP can be used to create intrusion detection rules that define normal and abnormal behavior on a network or host. GP creates and optimizes logical expressions or rules that consider various parameters such as network traffic, event logs and system metrics. This study (Lu et al., 2014: 579–593) discussed the ideology of genetic algorithm evolution used to create desirable network intrusion detection solutions. The fitness value indicates the quality of a chromosome (candidate solution) that can detect a set of predefined attack linkage data during the training process. The proposed method uses a combination of genetic operators, which are the processes of cloning, crossover, and mutation to generate

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

new chromosomes. Based on the presented results, the proposed method can detect any network connection intrusions and has proven to be a good mechanism for improving the security of computer networks. The paper (Mane et al., 2020: 723–738) presents and evaluates a genetic programming-based approach to detect known or emerging network attacks. The proof of concept shows that the new rules generated by the genetic algorithm have the potential to detect new forms of attacks. The paper (LaRoche et al., 2006: 3204–3215) explores the application of genetic programming to intrusion detection. The Modern DDoS dataset is used for this study. This dataset contains modern threats collected from different environments. The proposed genetic algorithm model detects DDoS attacks with 98.67 % accuracy when compared with six established classifier models. The paper (Pozi et al., 2016: 279–290) demonstrates that genetic-based IDS can be applied to attacks that are unique to the domain of WiFi networks. The results show that GP can be trained on one of the attacks on 802.11 networks with a result of 100 % detection rate with 0.529 % false positive rate. Also, the article describes that the genetic algorithm, even after being trained for such a specific type of attack, can provide a sufficiently generalized solution to detect such attacks. The paper (Qureshi et al., 2020: 654–670) discusses the issue of rare attack detection rate in intrusion detection systems (IDS) using the NSL-KDD dataset. The problem is that some rare attacks are not recognized due to their patterns being absent from the training set, which reduces the rare attack detection rate. The authors propose a new classifier, GPSVM, based on support vector machine (SVM) and genetic programming (GP) to address this problem. Experimental results demonstrate that GPSVM achieves a higher detection rate for anomalous rare attacks without significantly reducing overall accuracy. GPSVM is designed to balance accuracy between classes without reducing the generalization property of SVM.

In addition, applications of genetic programming on the Internet of Things (IoT) and attacks on wireless networks are being explored, as supported by articles on these topics. The paper (LaRoche et al., 2006: 3204–3215) proposes a novel and secure framework using genetic programming to detect security threats in RPL based IoT and IIoT networks. The proposed framework can detect various attacks and has been evaluated for attack detection accuracy, true positive rate, false-positive rate, throughput, and end-to-end delay. The results suggest that the proposed framework is the best choice for RPL based IIoT environments. The paper (LaRoche et al., 2006: 3204–3215) explores the use of Parallel Genetic Program ming (Karoo GP) in wireless attack detection to improve detection rates and processing time. Experiments showed that the processing time of Karoo GP was significantly improved compared to standard GP.

Nevertheless, genetic programming methods have the potential to detect and reconstruct attacks, however, existing implementations have limitations in terms of historical data, computational complexity, and retraining flexibility. However, this is true in relation to identifying a specific attack step. Moreover, if the goal is to reconstruct an attack, then each step can be considered as a separate node, and their

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

sequence is connected by edges. The task is to identify the nodes at each step of the attack and find their sequence. Thus, the attack vector is reduced to the form of a directed graph, and the reconstruction task itself consists in constructing it and finding the most probable path from the initial stage to the final one.

Attack Representation

Further, to build the most universal and widely implemented model, it is necessary to use a universal attack classifier as input. The source must provide fully structured, detailed and up-to-date information about the attacks, as well as distribute the stages of the attack in accordance with the phases of the kill chain. In addition, since the main task of this study is to find the path in the graph represented by all the actions of intruders in a compromised system, the main data source should contain information about all existing actions.

The MITRE ATT&CK Matrix is the most detailed and complete source of attack data and provides many opportunities for use in research. MITRE ATT&CK is a unique project, widely recognized by professionals; which is a real-world knowledge base of tactics, techniques, and methods used by cybercriminals. The main goal of creating this knowledge base is to compile a structured matrix of techniques used by cybercriminals to simplify the task of responding to cyber incidents. For the reconstruction of attacks, the matrix has several undoubted advantages. It provides a systematic and comprehensive approach to classify the tactics and methods used by attackers, which allows to structure information about the attack and break it into separate stages. As it is based on actual observations of attackers' behavior in real attacks, it is more practical and useful for attack analysis and reconstruction, especially when the attacks are complex and multi staged. That is why it also allows analysts and researchers to use the matrix to identify patterns of behavior, identify links between different methods, and predict possible next attack moves. In addition, matrix tactics correspond to the kill chain phases, respectively, this information can be used to build an attack graph in such a way as to reduce the computational complexity of the algorithm, discarding direct connections between techniques from those tactics that do not have a consistent relationship with each other.

Since the research objective is to reconstruct the attack graph, and each stage is represented by a node in the graph, this node can be represented as a MITTRE ATT&CK matrix technique. In this case, any attack can be decomposed into existing matrix techniques and represented as a vector, as in the example Black Energy. Black Energy is a malware toolkit that has been used by both criminal and APT actors. It dates to at least 2007 and was originally designed to create botnets for use in conducting Distributed Denial of Service (DDoS) attacks, but its use has evolved to support various plug-ins.

The list of techniques, according to MITTRE ATT&CK, includes the following, which are presented in Table 2.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

Table 2. Black energy software attack techniques

ID	Name
T1548.002	Bypass User Account Control
T1047	Windows Management Instrumentation
T1555.003	Credentials from Web Browsers
T1070	Indicator Removal
T1113	Screen Capture
T1055.001	Dynamic-link Library Injection
T1553.006	Code Signing Policy Modification
T1057	Process Discovery
T1083	FileandDirectory Discovery
T1046	NetworkService Discovery
T1021.002	SMB/Windows Admin Shares
T1049	System Network Connections Discovery
T1120	Peripheral Device Discovery
T1547.009	Shortcut Modification
T1552.001	Credentials In Files
T1056.001	Keylogging
T1543.003	Windows Service
T1070.001	Clear Windows Event Logs
T1547.001	Registry Run Keys/Startup Folder
T1485	DataDestruction
T1574.010	Services File Permissions Weakness
T1016	System Network Configuration Discovery
T1082	System Information Discovery
T1008	Fallback Channels
T1071.001	Web Protocols

Thus, based on this data, an attack graph can be constructed, as shown in Figure 1.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

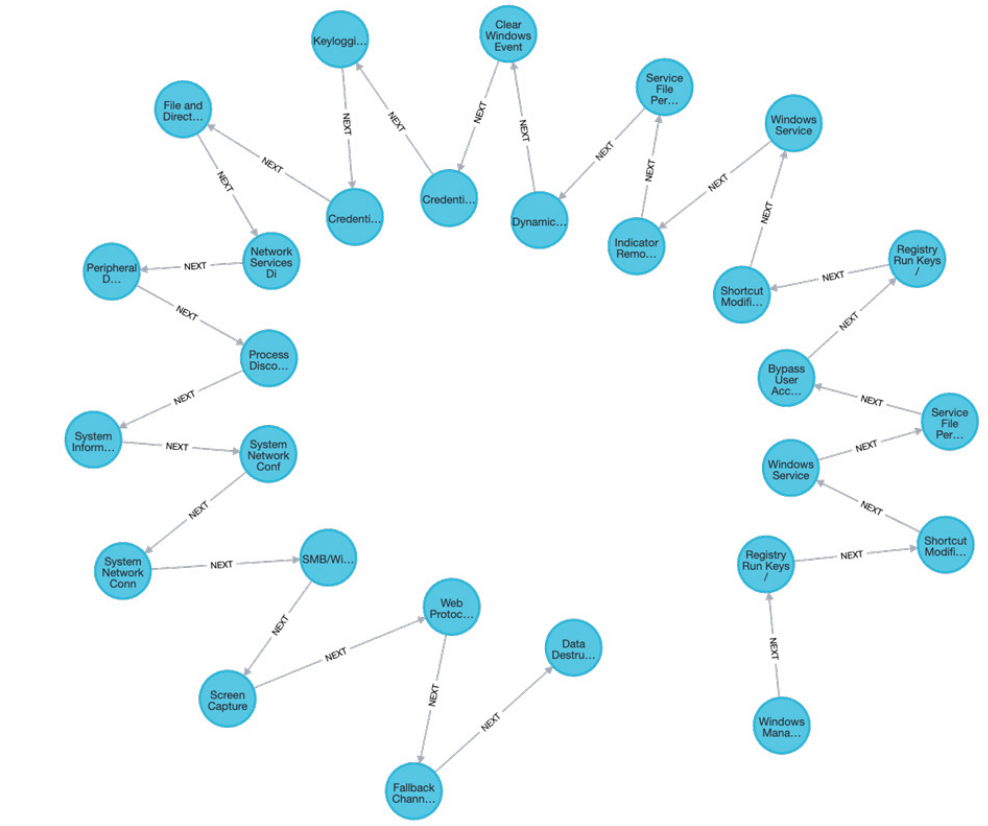

Figure 1 - Techniques Used in the Black Energy Sequence

Such a representation of the attack techniques clearly visualizes the connection edges between nodes. And the next section will describe the method, which was used to predict previous and subsequent steps.

Materials and Methods

1. Algorithm Structure

In this research, our primary objective is to explore the relationships between various techniques in the MITRE ATT&CK matrix. To accomplish this, we have created a dataset that links each technique to others based on their connections to various groups, campaigns, data sources, and software. The central idea is to identify the most probable association between a given technique, which is presumed to be detected as an indicator of compromise, and other techniques that are likely to be used in combination with it. By scrutinizing these links, we aim to uncover patterns and insights into how different techniques can be used together in actual cyberattacks. After that, we calculate the number of occurrences for each technique and create final dataset, which is a two-dimensional numpy array consisting of technique and calculated number of its occurrences among all objects that use the originally specified technique

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

(indicator of compromise).

It should be mentioned, that among all advantages of MITRE ATT&CK matrix it provides mapping of tactics to Cyber Kill-Chain phases. Cyber Kill-Chain is a security model that describes the stages of a cyber attack, which covers all stages of network hacking, from early planning and espionage to the hacker's final goal. This model supposes that blocking hackers actions at any stage breaks the entire attack chain and hackers must always go through the basic steps in Table 3 to commit their crimes.

Table 3. Cyber kill-chain phases

Phase	Impact
Reconnaissance	Researching potential targets
Weaponization	Searching for suitable malware
Delivery	Infiltrating to target network
Exploitation	Exploiting target vulnerabilities
Installation	Malware installation
Command and	Control Malware execution
Actions on Objectives	Reaching final goal

This structure of attack view will be used in this work to identify successive techniques in the attack graph assuming that phases (and mapped tactics) follow each other in the time in a strictly defined order. But techniques belonging to one tactic can be used in parallel or sequentially in free order.

Thus, the incoming data for genetic algorithm is vector consisting of number of occurrences for specific technique in relation to indicator of compromise. Further, for each of the techniques in the matrix, own separate occurrence matrix can be created. In this scenario, where we are dealing with the representation of the relationship between different techniques in the form of a graph, a hierarchical tree structure is a fitting approach. However, as we move through each layer of the tree, the number of interconnections between the techniques increases rapidly and exponentially. This can pose a significant challenge in terms of analyzing and visualizing the graph, as the number of nodes and edges grows exponentially with each additional layer.

Therefore, it is crucial to implement efficient algorithms and techniques for graph analysis, such as genetic programming, to effectively identify and analyze patterns and connections within the graph. By leveraging genetic programming, we can automate the process of identifying the most optimal paths and connections within the graph, making it easier to identify potential attack vectors and improve the overall security posture of the system.

Overall, genetic algorithm, developed in this research, has the structure, presented at Figure 2.

The first step in applying genetic algorithms to detect and predict attack vectors is to create an initial population. This initial population consists of one or more techniques that are indicators of compromise detected on one of the assets of the

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

system. Once the initial population has been established, the next step is to define the input vector for the genetic algorithm model. This input vector is obtained from the weight matrix for the detected indicator of compromise, and it represents the features or attributes that will be used to evaluate the fitness of the population.

With the initial population and the input vector in place, the genetic algorithm can begin to create branches from the root node using the incoming weight vector. Each branch represents a path or sequence of techniques that could be used in a real-world cyberattack. The algorithm evaluates each branch by computing its fitness score, which is based on how it will matches the desired attack vector. The fitness score is then used to select the fittest branches for further optimization through crossover and mutation operations.

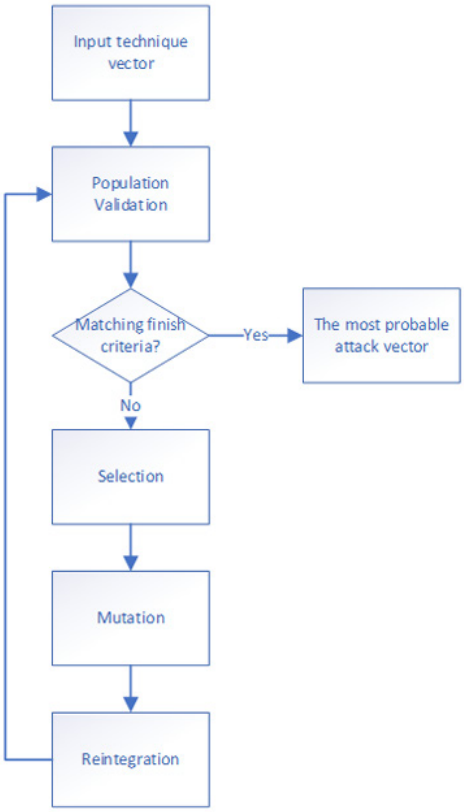

Fig.2 - Structure of the genetic algorithm

Additionally, the techniques already present in the previously created paths are eliminated to avoid duplication. Next, the techniques that do not fit the predicted kill-chain phase are filtered out. For example, while building a graph in the past to identify the entry point of the attack, the techniques that belong to the subsequent phases of the kill-chain are discarded. Similarly, for future predictions, all techniques

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

that belong to earlier phases of the kill-chain than the given one is filtered out.

Once the filtering is done, the techniques are ranked according to their popularity. The most popular techniques, which account for 67 % of the total, are selected, and the remaining 33 % are discarded. This ensures that the model focuses only on the most relevant techniques.

After the ranking, the incoming vector is normalized to obtain a weighted estimate of the frequency of occurrence of each technique relative to its global popularity. Normalization involves dividing each value in the vector by the maximum number of occurrences. This step helps to obtain a uniform scale for the frequency of occurrence of the techniques. The finish criteria in the developed model are a crucial component that ensures the effectiveness and efficiency of the genetic algorithm. The criterion's primary purpose is to maintain a balance between computational resources and the accuracy of the results. The delta score, which is calculated for each step of the algorithm, is a measure of the change in the weight vector of the technique. The threshold value of 0.01 is set to ensure that the delta score over two steps is not significant.

The algorithm takes the delta from the previous step and calculates a new delta for the next step. If the two deltas are below the threshold of 0.01, the algorithm will not create a new branch for this technique. This is a crucial step that eliminates unnecessary computation and optimizes the algorithm's overall performance. Additionally, this ensures that the algorithm con verges within a reasonable time frame, while also maintaining a high degree of accuracy in its results.

In cases where it is not possible to find a technique to jump to, the algorithm considers the creation of the path complete. This step ensures that the algorithm is comprehensive and explores all paths. The threshold value of 0.01 was chosen as the most optimal for the available computing resources, considering both the accuracy and speed of the algorithm.

Selection Phase

The selection process is based on two criteria: local fitness and global fitness. Local fitness is a measure of the effectiveness of the technique within its own path, while global fitness considers the performance of the technique across all paths. To select the one hundred genes, we follow a two-step process. First, we identify the top 50 % of techniques based on their local fitness scores. These techniques have the highest weighted score within their respective paths and are thus the most effective at achieving their intended goal. For the remaining 50 % of genes, we look at the 20 % best techniques across the entire population, except for those in the top 50 %. This ensures that we maintain a diverse population and do not simply replicate the best-performing individuals from the previous generation.

The selection phase is a critical component of the genetic algorithm, as it determines which individuals will be passed on to the next generation and shapes the trajectory of the algorithm's search for optimal solutions. By balancing local and global fitness criteria, we can maintain a diverse and effective population that can

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

achieve the desired goals.

In the genetic algorithm we are using, a mutation rule is employed to change the coefficient value of the graph path that has been constructed. This means that a mutation can occur in the graph at any point during the evolutionary process. By changing the coefficient value, we can explore new paths and potentially find more optimal solutions. The mutation rule is just one aspect of the genetic algorithm, which relies on principles of natural selection and genetic recombination to create novel solutions to a problem. The algorithm uses a population of candidate solutions, or individuals, and iteratively evolves them to create better and better solutions over time.

Graph Construction and Weight Calculation

In our research, we employed a series of steps to construct the attack graph for a given system based on the MITRE ATT&CK framework. We started by parsing the data from version 13 of the MITRE ATT&CK matrix, which involved compiling a list of technique popularity relative to a given one using related groups, campaigns, software variants, and date components. This information was used to fill in the global frequency matrix for each technique. We then determined the direction of the prediction (future or past) and the StixID of the technique for which the prediction would be built. The phase of the kill-chain was also determined based on the direction of the prediction. If the construction of the graph was directed towards the past, then the phase was considered final, but if it was directed towards the future, then it was initial. The depth step of the tree being built was also set as a parameter. The given technique (indicator of compromise) was represented as a single path vector and thrown into the recursive path building function. This function included several parameters such as the phase ID of the last element in the path vector, the reconstruction type, the count delta, the count factor, the chum count, and the depth to stop. The path creation function initially checked if the depth had been exceeded, and if it had, the execution ended with the return of all incoming parameters. Variable restrictions on the frequency of techniques were discarded, and the frequency vector was normalized by the maximum value in the range from 0 to 1. For each technique remaining in the vector, the search for scores began. To do this, the normalized score calculated at the previous stage was multiplied by the score coefficient, and the delta of the score and the delta calculated at this stage were checked to ensure that they exceeded the threshold value.

One important consideration was the need to avoid local minima and maxima in the data, which could lead to inaccurate or incomplete conclusions. To achieve this, we developed a multi-stage approach to calculating score coefficients, which consider the relative frequency and depth of each technique in the matrix. The first stage of the coefficient calculation involves multiplying the previous coefficient by the normalized frequency of the next technique. This allows us to weigh the importance of each technique in the overall analysis, based on its observed prevalence in the matrix. The second stage of the calculation involves applying a depth penalty multiplier to the coefficient. The depth penalty considers the current depth of the technique in the

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

the path vector, which is represented as a constant rise to the power of the depth. This helps to account for the fact that techniques that are further down the path may be less significant than those closer to the beginning.

The third step was to find the average frequency of the new technique to which the branch is being built. We used 80% of the upper frequencies to avoid negatively affecting the accuracy of the overall statistics. If the number of appearances of the new technique in the current technique is greater than the average value, then the coefficient is increased by the next value. The number of spawns of new techniques is divided by the average value and raised to the power of the depth penalty multiplied by the weight of the technique (Formula 1). If the number of appearances is less than the average value, then the score coefficient is multiplied by the quotient of the frequency of the technique by the average frequency of the technique (Formula 2).

$$new\_score\_coef \mathrel{*}= \frac{technique\_frequency^{depth\_penalty * weigh\_tecch\_freq}}{avg\_tech\_freq} \tag{1}$$

$$new\_score\_coef \mathrel{*}= \frac{technique\_frequence}{avg\_tech\_freq} \tag{2}$$

$$new_{score_{coef}} \mathrel{*}= \\ = 1 + \frac{technique_{frequency} - avg_{tech_{freq}}}{\max\_tech\_freq - avg\_tech\_freq} * depth_{p}\text{enalty} * weigh_{t}\text{otal}_{t}\text{\_ech}_{f} frequency \tag{3}$$

To calculate the score coefficient for each technique, we first multiplied the previous coefficient by the normalized frequency of the next technique. Then, we applied a depth penalty multiplier, which is a constant raised to the power of the existing current depth. This helped to avoid local minima and maxima and ensure that the graph accurately represented the relationships between techniques.

In the fourth step, we divided the global frequency of the new technique by the global average occurrence of technique. If the result was greater than one, then the new coefficient was multiplied by formula three, otherwise, the quotient of the given technique by its global frequency was used. By prioritizing relationships over global popularity, we were able to more accurately identify the most optimal paths for a particular system in the attack graph.

For the fifth step, it was essential to ensure that the new phase was calculated accurately, depending on whether the reconstructing was made to the past or the future. For the case of a prediction to the past, it was necessary that the new phase be no higher than the current phase, and in the optimal case, be close to or equal to it. On the other hand, for the case of a predictor into the future, it was necessary that the new phase is not lower than the current phase, and in the optimal case, it should be close to

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

or equal to it. For the sixth step the recursive function was used to create a new path consisting of the current path and the new technique, a new phase identifier, a prediction type, a new score delta, a new score factor, and its sum with the new delta, and the depth value to stop. When the loop was exited, if the length of the pathwas0, then the current path was returned, otherwise all paths created by the recursive function were returned. Then algorithms enter a while loop and continue until the maximum length among all paths reaches the specified depth. The depth is increased by a fixed amount (delta) in each iteration. We select the most probable paths and use them to create new paths iteratively with increased depth. The iteration process follows a specific strategy to avoid being stuck in local minimums. After testing different strategies, we found the optimal approach to be alternating the depth delta between 3 and 4 and using a mirrored interpretation. Values less than 3 are not suitable because they generate too few paths, while values greater than five increase the computational complexity and time of solving the problem. Using a depth greater than four reduces the reconstruction accuracy, which negatively affects the algorithm's results. To predict the possible attack vector for a particular system, it is crucial to have a clear understanding of the security posture of all assets in the network. For this purpose, our research represents and describes each asset through the corresponding MITRE ATT&CK matrix, such as Windows, Linux, and Mac OS. By doing so, we can identify the specific vulnerabilities and techniques that are applicable to each system, as well as the potential attack paths that an adversary may follow to exploit these weaknesses. This approach enables us to create a comprehensive attack graph that accurately reflects the attack surface of the entire network and helps us to develop effective defensive strategies to mitigate the risks. The results of the algorithm performance, representing reconstructed graphs for past and future, are presented below.

Table 4. Predicting past techniques for attack pattern: DNS Server T1584.002

№	Technique name	Technique code
1	DNS Server	T1584.002
2	Domains	T1584.001
3	Domains	T1583.001
4	Tool	T1588.002
5	Malware	T1587.001
6	SocialMediaAccounts	T1585.001
7	EmailAccounts	T1585.002

Table 5. Predicting past techniques for attack pattern: DNS Server T1584.002

№	Technique name	Technique code
1	DNS Server	T1584.002
2	Domains	T1584.001
3	Domains	T1583.001
4	Tool	T1204.002
5	Malware	T1105

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

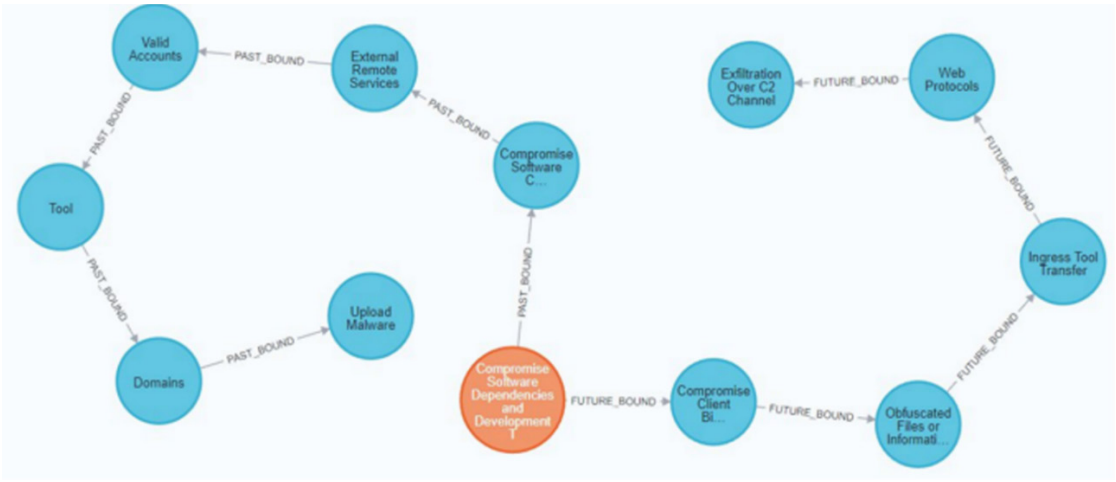

Fig. 5 – Full predicted chain for: Compromise Software Dependencies and Development Tools T1195.001

Results and Discussion

There are some limitations to the proposed approach. The effectiveness of the algorithm may be affected by the quality and completeness of the MITRE ATT&CK Matrix. Additionally, the algorithm may require significant computational resources, making it challenging to implement in large-scale systems. Also, one of the limitations of our study is that we only used the MITRE ATT&CK Enterprise Matrix to build our model. We did not include data from the ICS and Mobile matrices. This limitation is significant because, in today's world, the use of mobile devices and industrial control systems in corporate networks is increasing. The concept of BYOD and MDM has made it easier for employees to access corporate data through their personal devices. This means that there is a higher likelihood of these devices being compromised, and it is essential to consider these devices in any predictive model for attack vectors. Industrial control systems are also a critical aspect that needs to be considered. The increasing number of these systems being used in various industries has made them a potential target for cyberattacks. The absence of these matrices from our study is a significant limitation that needs to be addressed in future research. Another limitation is the use of data from only one source, the MITRE ATT&CK Matrix. While this matrix is a widely recognized and respected resource in the field of cybersecurity, it is important to note that it may not capture all attack vectors or techniques. There are likely to be many attacks that have not been included in the matrix, as they may not have been reported or discovered yet. Therefore, the findings of the study may not be comprehensive and could potentially miss some important attack vectors.

Conclusion

Several potential future research directions stem from this study on detecting and predicting attack vectors using genetic programming: Enhancing system accuracy: The current system can be further refined to improve its predictive accuracy. For

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

example, the system may need to be trained on a larger dataset to increase its predictive power, or new features could be added to the system to help it better detect attack vectors. Also, two other types of MITRE ATT&CK Matrices should be included as a data source, so the model could be able to build attack vectors for systems with mobile devices and industrial control systems. Extending the system to new types of attacks: The current system may be tailored to a specific type of attack, but it could be extended to detect and predict other types of attacks. For example, the system could be adapted to detect phishing attacks or malware attacks. Evaluating the system in a real-world scenario: The current system may have been tested in a controlled environment, but it would be valuable to evaluate its performance in a real-world scenario. This could involve working with a company or organization to test the system against real-world attacks and evaluate its effectiveness.

Developing a user-friendly interface: The current system may be complex and difficult to use for non-experts. Developing a user-friendly interface could help make the system more accessible to a wider range of users and increase its adoption. Conducting a comparative study: The current system could be compared to other methods for detecting and predicting attack vectors, such as rule-based systems or machine learning models. This would help to determine the strengths and weaknesses of different approaches and identify areas for further improvement. Adding a language model: Since the input to the algorithm is already detected indicators of compromise related to a known technique, the language model can help isolate a particular technique based on a linguistic description of the attack in progress.

REFERENCES

Nafie M.S., Abounaser H. & Khaled Adel A.B. (2019). Hybrid genetic FSM technique for detection of high-volume DoS attack. — *International Journal of Advanced Computer Science and Applications.* — 2019. — Pp. 311–325.

Kayacık G., Zincir-Heywood A.N., Heywood M.I. & Burschka S. (2009). Generating mimicry attacks using genetic programming: A benchmarking study. — *IEEE Symposium on Computational Intelligence in Cyber Secu-rity (*CICS). — 2009. — Pp. 512–530.

Laroche P., Zincir-Heywood A.N., Heywood M.I. & Burschka S. (2009). Evolving TCP/IP packets: A case study of port scans. — *IEEE Symposium on Computational Intelligence in Cyber Security* (CICS). — 2009. — Pp. 104–120.

Rovito L., Bonin L., Manzoni L. & Lorenzo A.D. (2022). An evolutionary computation approach for Twitter bot detection. — *Applied Sciences (Switzerland).* — 2022. — Pp. 391–405.

Al-Harahsheh H., Al-Shraideh M. & Al-Sharaeh S. (2021). Performance of malware detection classifier using genetic programming in feature selection. — *Informatica (Slovenia).* — 2021. — Pp. 303–317.

Al-Sahaf H. & Welch I. (2019). A genetic programming approach to feature selection and construction for ransomware, phishing and spam detection. — *2019 Genetic and Evolutionary Computation Conference.* — 2019. — Pp. 487–499.

LaRoche P. & Zincir-Heywood A.N. (2006). Genetic programming-based WiFi data link layer attack detection. — *Communication Networks and Services Research Conference.* — 2006. — Pp. 145–158.

Suhaimi H., Suliman S.I. & Musirin I. (2019). Network intrusion detection system by using genetic algorithm. — *Indonesian Journal of Electrical Engineering and Computer Science.* — 16 (3). — 1593. — 2019.

Lu W. & Traore I. (2014). Detecting new forms of network intrusion using genetic programming. — *Evolutionary Computation*. — 2014. — Pp. 579–593.

Mane N., Verma A. & Arya A. (2020). A pragmatic optimal approach for detection of cyber attacks using genetic programming. — *International Symposium on Computational Intelligence and Informatics.* — 2020. — Pp.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

LaRoche P. & Zincir-Heywood A.N. (2006). 802.11 de-authentication attack detection using genetic programming. — *European Conference on Genetic Programming.* — 2006. — Pp. 3204–3215.

Pozi M.S.M., Sulaiman M.N. & Mustapha N. (2016). Improving anomalous rare attack detection rate for intrusion detection system using support vector machine and genetic programming. — *Neural Processing Letters*. — 44(2). — 2016. — Pp. 279–290.

Qureshi K.N., Rana S.S. & Ahmed A. (2020). A novel and secure attacks detection framework for smart cities industrial internet of things. — *Sustainable Cities and Society.* — 2020. — Pp. 654–670.

LaRoche P. & Zincir-Heywood A.N. (2006). 802.11 de-authentication attack detection using genetic programming. — *European Conference on Genetic Programming.* — 2006. — Pp. 3204–3215.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

ХАЛЫҚАРАЛЫҚ АҚПАРАТТЫҚ ЖӘНЕ КОММУНИКАЦИЯЛЫҚ ТЕХНОЛОГИЯЛАР ЖУРНАЛЫ

МЕЖДУНАРОДНЫЙ ЖУРНАЛ ИНФОРМАЦИОННЫХ И КОММУНИКАЦИОННЫХ ТЕХНОЛОГИЙ

INTERNATIONAL JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGIES

Правила оформления статьи для публикации в журнале на сайте:
https://journal.iitu.edu.kz
ISSN 2708–2032 (print)
ISSN 2708–2040 (online)
Собственник: АО «Международный университет информационных технологий» (Казахстан, Алматы)

ОТВЕТСТВЕННЫЙ РЕДАКТОР
Мрзабаева Раушан Жалиқызы

КОМПЬЮТЕРНАЯ ВЕРСТКА
Асанова Жадыра

Подписано в печать 15.03.2025.
Формат 60x881/8. Бумага офсетная. Печать - ризограф. 9,0 п.л. Тираж 100
050040 г. Алматы, ул. Манаса 34/1, каб. 709, тел: +7 (727) 244-51-09).

Издание Международного университета информационных технологий
Издательский центр КБТУ, Алматы, ул. Толе би, 59