Міністерство освіти і науки України

Харківський національний університет радіоелектроніки

САПРИКІН ОЛЕКСАНДР СЕРГІЙОВИЧ

УДК 658:512.011: 681.326: 519.713

МОДЕЛІ АВТОМАТИЗОВАНОГО АНАЛІЗУ ТА ДІАГНОСТУВАННЯ ПОЛІМОРФНИХ ВІРУСІВ У КОМП'ЮТЕРНИХ СИСТЕМАХ ТА МЕРЕЖАХ

05.13.05 – комп'ютерні системи та компоненти

Автореферат дисертації на здобуття наукового ступеня

кандидата технічних наук

Харків – 2021

Дисертацією є рукопис

Робота виконана в Харківському національному університеті радіоелектроніки, Міністерство освіти і науки України

Науковий керівник: доктор технічних наук, професор
Чумаченко Світлана Вікторівна, Харківський національний університет радіоелектроніки, завідувач кафедри автоматизації проектування обчислювальної техніки.

Офіційні опоненти: доктор технічних наук, професор
Мірошник Марина Анатоліївна, Український державний університет залізничного транспорту Міністерства освіти і науки України, України, професор кафедри спеціалізованих комп'ютерних систем;

доктор технічних наук, професор
Леонов Сергій Юрійович, Національний технічний університет «Харківський політехнічний інститут» Міністерства освіти і науки України, професор кафедри обчислювальної техніки та програмування.

Захист відбудеться "29" вересня 2021 р. о 15-00 годині на засіданні спеціалізованої вченої ради Д64.052.01 в Харківському національному університеті радіоелектроніки за адресою: 61166, місто Харків, пр. Науки, 14.

З дисертацією можна ознайомитись в бібліотеці Харківського національного університету радіоелектроніки за адресою: 61166, місто Харків, пр. Науки, 14.

Автореферат розісланий "19" серпня 2021 року.

Вчений секретар
спеціалізованої вченої ради Є.І. Литвинова

ЗАГАЛЬНА ХАРАКТЕРИСТИКА РОБОТИ

Актуальність теми дослідження. Запропоноване дослідження присвячене науково-практичній задачі розпізнавання поліморфних мутаторів, що автоматично змінюють синтаксис і логіку роботи при кожній активації. Актуальність обгрунтована високим рівнем ринкової затребуваності технологій, моделей, методів та інфраструктури для розвитку Cyber Security, що підтверджено тенденціями, визначеними Gartner Analytics в сфері інформаційної безпеки та управління кіберризиками на 2021-2022 роки. Тема роботи націлена на створення і впровадження федеративної архітектури cloud-edge комп'ютингу на основі ML-sandbox і векторно-логічних методів пошуку zero-day шкідливих кодів з метою захисту інфраструктури кіберфізичного простору, істотного зменшення часу і вартості розпізнавання поліморфних мутаторів. Іншими словами, інтеграція нейромереж, федеративного машинного навчання, сигнатурних методів пошуку AI-malware створює на сьогодні ефективний захист в формі FLT-комп'ютингу, який покликаний протистояти сотням тисяч шкідливих програм від кіберкримінального світу. Істотний внесок в наукові дослідження, що стосуються інформаційної безпеки та захисту від комп'ютерних загроз, внесли вчені та фахівці: J. von Neumann, L. Penrose, F. Cohen, Крис Касперськи, Brian Krebs, Брюс Шнайер, Bruce Dang, Alexandre Gazet, Elias Bachaalany, Michael Sikorski, Andrew Honing, Abhijit Mohanta, Anoop Saldanha, Є. Касперський, С. Новиков, І.Д. Горбенко, В.С. Харченко.

Зв'язок роботи з науковими програмами та темами. Розробка теми дисертації здійснювалася відповідно до планів аспірантської підготовки, держбюджетних НДР і міжнародних договорів, виконуваних на кафедрі Автоматизації проектування обчислювальної техніки ХНУРЕ в період з 2007 року, у тому числі: 1) Прикладна держбюджетна НДР № 216 «Енергозберігаючі інформаційні технології на основі паралельних обчислювальних процесів, безпровідних систем і мереж», 2007-2008, № ДР 0107U001540. 2) Договір про дружбу і співробітництво між ХНУРЕ та компанією «Aldec Inc.», USA, 2001 – 2020. 3) Фундаментальна держбюджетна НДР № 232 «Теорія й проектування енергозберігаючих цифрових обчислювальних систем на кристалах, що моделюють і підсилюють функціональні можливості людини», 2009-2011, № ДР 0109U001646. Автор дисертаційної роботи брав участь у виконанні зазначених договорів і програм як розробник і програміст кіберфізичної інфраструктури захисту кіберпростору у вигляді програмних засобів перевірки, діагностування шкідливих програм (поліморфних мутаторів), що характеризуються механізмом управління логічною і синтаксичною модифікацією коду шкідливої програми для її маскування від детектування існуючими антивірусними сервісами. Автор також брав участь у конкурсах інноваційних проектів та розробок, серед яких Міжнародна студентська конференція «IT Security For New Generation» 2008, 2009 як доповідач та розробник навчального курсу "Computer threats: detection and analysis methods".

Науково-практична задача – розпізнавання поліморфних мутаторів, що автоматично змінюють синтаксис і логіку роботи при кожній активації.

Сутність дослідження – розробка і впровадження федеративної архітектури cloud-edge комп'ютингу на основі ML-sandbox і векторно-логічних методів пошуку zero-day шкідливих кодів з метою захисту інфраструктури кіберфізичного простору, істотного зменшення часу і вартості розпізнавання поліморфних мутаторів.

Об'єкт дослідження – технології комп'ютингу для розпізнавання деструктивних програм і прийняття рішення по їх усуненню.

Предмет дослідження – архітектури, моделі і методи пошуку шкідливих програм, що мутують, на основі federated machine learning.

Мета дослідження – істотне зменшення часу і вартості розпізнавання поліморфних мутаторів шляхом розробки і впровадження федеративної архітектури cloud-edge комп'ютингу на основі ML-sandbox і векторно-логічних методів пошуку zero-day шкідливих кодів для захисту інфраструктури кіберфізичного простору.

Поліморфний мутатор – механізм управління логічною і синтаксичною модифікацією коду шкідливої програми для її маскування від детектування існуючими антивірусними сервісами.

Задачі дослідження:

1) Розробити федеративну ML-архітектуру sandbox комп'ютингу для пошуку шкідливих кодів в об'єктах і додатках кіберпростору, синтезу та аналізу таблиці істинності з метою паралельного виконання логічних операцій пошуку шкідливих кодів.

2) Удосконалити матрично-логічний метод діагностування malware для паралельного виконання логіки алгоритму з використанням сигнатурного аналізу.

3) Розробити методи: вирішення задачі детектування модифікованих шкідливих кодів на основі евристичного аналізатора шкідливого коду за допомогою нейронної мережі і ML-таблиць; публічного і локального мультисканера; Yara правил.

4) Виконати тестування розробленого federated cloud–edge сервісу для детектування шкідливих програм на основі «пісочниць» з відкритим вихідним кодом і MAS.

Наукова новизна результатів дисертаційної роботи:

1) Вперше запропоновано федеративну ML-архітектуру sandbox комп'ютингу, яка характеризується федеративним розподілом у просторі терміналів машинного навчання на основі «пісочниць», що дає можливість істотно знизити навантаження на канали передачі даних шляхом локальної обробки підозрюваних кодів, підвищити продуктивність сукупного cloud-edge computing, зменшити час навчання і тестування глобальної хмарної

«пісочниці», підвищити якість розпізнавання шкідливих кодів, зберегти цілісність і конфіденційність даних на терміналах користувачів.

2) Удосконалено структурну модель ML-комп'ютингу, яка відрізняється від відомих синтезом та аналізом ML-таблиці істинності на основі характеристичного рівняння тестування T⊕F⊕L = 0, що дає можливість паралельно виконувати логічні операції пошуку шкідливих кодів у локальному кіберпросторі.

3) Удосконалено матрично-логічний метод діагностування шкідливих кодів, який відрізняється від відомих технологій паралельним виконанням логіки алгоритму над рядками і стовпцями попередньо синтезованої таблиці функцій деструктивних компонентів, що дозволяє підвищити продуктивність edge-комп'ютингу користувача.

4) Удосконалено векторно-матричний метод діагностування шкідливих кодів, який відрізняється від відомих технологій векторним поданням координат матриці деструктивних компонентів, що дає можливість підвищити продуктивність алгоритму аналізу багатозначних або сигнатурних даних шляхом паралельного виконання трьох логічних операцій.

5) Вперше запропоновано методи:

– детектування модифікованих шкідливих кодів, що заснований на реалізації евристичного аналізатора шкідливого коду за допомогою штучної нейронної мережі;

– детектування досліджуваного зразка заздалегідь встановленими антивірусними рішеннями, які дозволяють в окремому потоці проводити статичне сканування досліджуваного об'єкта без обмежень на кількість запитів на хвилину, підвищувати швидкість обробки об'єктів і обмежувати публічний доступ до конфіденційних файлів;

– діагностування поліморфних шкідливих програм за допомогою Yara правил, що дозволяє детектувати нові модифікації, які не виявляються доступними рішеннями;

– створення URL сигнатур нового покоління, що дозволяють успішно детектувати нові шкідливі URL на скомпрометованих легальних серверах, точково блокуючи доступ до шкідливого об'єкта, при цьому не блокуючи весь ресурс, це дає можливість скоротити розмір бази даних на 75%.

Практичне значення одержаних результатів досліджень полягає у:

– тестуванні, верифікації і впровадженні розроблених програмних засобів перевірки, діагностування шкідливих програм (поліморфних мутаторів), що характеризуються механізмом управління логічною і синтаксичною модифікацією коду шкідливої програми для її маскування від детектування існуючими антивірусними сервісами. Розроблені методи аналізу шкідливих програм успішно виявляли атаки під час їх моделювання, а також підозрілі файли, URL-адреси та електронні листи та зменшували час реакції на інциденти;

– розробці хмарного сервісу, що об'єднує MAS Sandbox і модифікований розподілений аналізатор шкідливих програм Cuckoo, який дозволяє швидко

реагувати на zero-day загрози, зберігати базу знань для кореляції артефактів між поліморфними зразками шкідливих програм, активно шукати нові зразки шкідливих програм та інтегруватися з програмно-апаратними комплексами захисту кіберпростору, що підтримують Cuckoo API. Запропонований хмарний сервіс дозволяє активно застосовувати нейромережні технології виявлення шкідливих об'єктів, збирати артефакти для створення Yara правил і детектувати поліморфні шкідливі програми, а також горизонтально масштабуватися в гібридних хмарах.

Обґрунтованість наукових положень. Отримані в процесі виконання досліджень наукові висновки і практичні результати є достовірними, що підтверджується достатньою кількістю проведених експериментів, точністю детектування, апробацією результатів на міжнародних науково-практичних конференціях, впровадженням результатів у виробничий та освітній процеси.

Впровадження результатів дисертації. Результати дисертаційного дослідження у складі моделей, методів та інфраструктури впроваджені у навчальний процес Харківського національного університету радіоелектроніки (акт про впровадження від 27.06.2021); у виробничу діяльність IT компанії EPAM, USA (довідка від 10.06.2021), а саме: критичні компоненти "Системи автоматичного аналізу шкідливих програм" в рамках проекту внутрішньої безпеки EPAM інтегровано в системи безпеки EPAM (Cofence Антифішинг, TrapX Honeypot, Cisco Umbrella DNS-фільтрація).

Особистий внесок здобувача. Всі наукові і практичні результати отримані автором особисто. У роботах, опублікованих зі співавторами, здобувачеві належать: [1] – метод детектування модифікованих шкідливих кодів, заснований на реалізації евристичного аналізатора шкідливого коду за допомогою штучної нейронної мережі; [2] – методика оцінки впливу шкідливих програм на діяльність підприємства, структурна модель інформаційної безпеки підприємства; [3] – метод діагностування поліморфних шкідливих програм за допомогою Yara правил, що дозволяє детектувати нові модифікації, які не виявляються доступними рішеннями; [4] – метод детектування досліджуваного зразка заздалегідь встановленими антивірусними рішеннями; метод діагностування поліморфних шкідливих програм за допомогою Yara правил; [5] – федеративна ML-архітектура sandbox комп'ютингу, удосконалена структурна модель ML-комп'ютингу, удосконалений матрично-логічний метод діагностування шкідливих кодів, удосконалений векторно-матричний метод діагностування шкідливих кодів; [6] – сучасні тенденції розвитку технологій пошуку вірусів, федеративні алгоритми та архітектури; [7] – структурна модель загроз; [8] – метод детектування модифікованих шкідливих кодів, заснований на реалізації евристичного аналізатора шкідливого коду за допомогою штучної нейронної мережі; [9] – метод діагностування шкідливого коду у виконуваних файлах; [10] – аналіз підходів до захисту персонального кіберпростору; [11] – методи захисту від шкідливого ПЗ; [12] – модифікована методика оцінки

збитків та програмна реалізація; [13] – модель інформаційної безпеки, її програмна реалізація та тестування; [14] – методи детектування для захисту комп'ютерів; [15] – евристичний аналізатор шкідливого коду на основі штучної нейронної мережі; [16] – хмарний сервіс аналізу поліморфних шкідливих програм.

Апробація результатів дисертації. Результати роботи були представлені та обговорені на таких конференціях: IEEE East-West Design and Test Symposium 2010 (Saint Petersburg, Russia), 2007 (Yerevan, Armenia), Міжнародний молодіжний форум «Радіоелектроніка та молодь у XXI столітті», 2007 (Харків, Україна); International Conference «The Experience of Designing and Application of CAD Systems in Microelectronics», CADSM 2009 (Lviv-Polyana, Ukraine); Міжнародна студентська конференція та конкурс наукових робіт з питань інформаційної безпеки «IT Security for the Next Generation», Kaspersky Office, 2008 (Москва, РФ); II International Student Conference on Computer Security Issues "It Security For The New Generation", Kaspersky Office, 2009, (Moscow, RF); XXIX International Scientific and Practical Conference "Science, theory and practice", 2021 (Tokyo, Japan); LXVIII Міжнародна інтернет-конференція «Літні наукові дискусії», 2021 (Україна, Чернівці).

Автор також брав участь у конкурсах інноваційних проектів та розробок із здобуттям призових місць, серед яких Міжнародна студентська конференція та конкурс наукових робіт з питань інформаційної безпеки «IT Security for the Next Generation», Kaspersky Office, 2008, 2009, Kaspersky Lab.

Публікації. Результати дисертаційної роботи відображені в 16 друкованих працях, серед яких 6 статей у міжнародних науково-метричних базах: 2 статті в міжнародних наукових журналах за кордоном, 4 – у наукових журналах, включених до «Переліку наукових фахових видань України»; а також 10 тез доповідей у матеріалах міжнародних наукових конференцій (з них 2 входять до науково-метричної бази Scopus).

Структура дисертації. Дисертація складається з 186 сторінок (з них 140 представляють основний текст) і містить: 5 розділів, 73 рисунка, список джерел з 89 назв (на 10 с.), 4 додатки (на 23 с.), анотації (на 15 с.).

ОСНОВНИЙ ЗМІСТ РОБОТИ

У **вступі** обгрунтовано актуальність завдань, які вирішуються в дисертаційній роботі, сформульована мета дослідження, а також викладені наукова новизна і практична цінність отриманих результатів.

У **першому розділі** наводиться аналітичний огляд існуючих моделей, методів і технологій, підходів до аналізу шкідливих програм, які включають статичний, динамічний, гібридний і заснований на візуалізації аналіз. Показуються сучасні технології пошуку вірусів, cloud-edge computing, а також федеративні алгоритми і архітектури машинного навчання. Визначаються переваги

і недоліки найбільш затребуваних моделей і методів, опублікованих в спеціальній літературі: матеріалах конференцій і наукових журналах. На основі проведеного аналізу формулюється мета і задачі дослідження, орієнтовані на усунення проблемних місць і недоліків існуючих моделей і методів у контексті їх реалізації в інфраструктурі аналізу поліморфних програм.

У **другому розділі** розглядається федеративна інфраструктура cloud-edge computing на основі «пісочниці». Показується використання сигнатурного аналізу і механізму асерцій для пошуку malware. Пропонується парадигма LTF-комп'ютингу пошуку деструктивних компонентів у програмних додатках. Розробляється матрично-логічний метод діагностування шкідливих компонентів, що має високу швидкодію за рахунок паралельних операцій. Подається векторно-матричний метод пошуку malware в CPS, який використовує реакції функціональності і malware на тестові набори даних. Всі моделі і алгоритми верифіковані в програмному коді і готові до імплементації в систему хмарного федеративного машинного навчання для пошуку malware. Алгоритми, запропоновані в дослідженні, мають лінійну або квадратичну обчислювальну складність. Завдяки паралельному виконанню регістрових логічних операцій процедури аналізу матриць або таблиць перевершують аналоги від двох до 10 разів по продуктивності. Пропонується оригінальна архітектура cloud-edge комп'ютингу (рис. 1) для алгоритму федеративного навчання, який працює ітеративно з чотирма фазами: локальне навчання (Training), завантаження параметрів (Upload) в хмарну «пісочницю», агрегування (Aggregating) параметрів на хмарі і повернення їх до терміналів (Download).

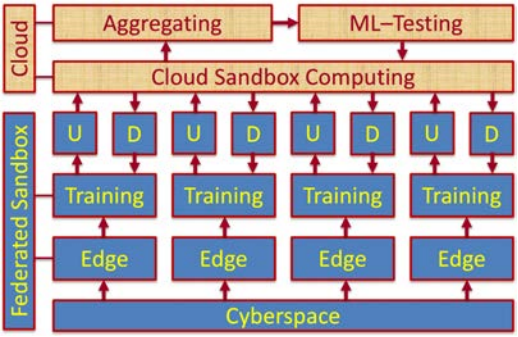


Рис. 1. Cloud-sandbox malware detection computing

Пропонується федеративна ML-архітектура sandbox комп'ютингу, яка характеризується федеративним розподілом в просторі терміналів машинного навчання на основі «пісочниць». Це дає можливість істотно знизити навантаження на канали передачі даних шляхом локальної обробки підозрюваних кодів, підвищити продуктивність сукупного cloud-edge computing, зменшити час навчання і тестування глобальної хмарної «пісочниці», підвищити якість розпізнавання шкідливих кодів, зберегти цілісність і конфіденційність даних на терміналах користувачів.

Комп'ютинг для пошуку та усунення деструктивних кодів використовує базове характеристичне рівняння тестування F⊕T⊕L = 0, яке дає можливість визначати будь-який з компонентів за двома іншими відомими. Тут F – функціональний код, T – код з можливими порушеннями, L – шкідлива програма, що підлягає усуненню. Архітектура комп'ютингу для online моніторингу та усунення шкідливих програм (ШП) подана на рис. 2.

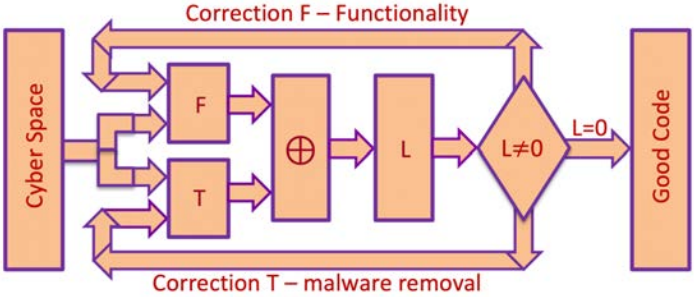
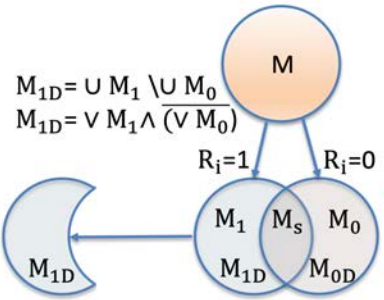


Рис. 2. Комп'ютинг пошуку та усунення malware

Отже, удосконалено структурну модель ML-комп'ютингу, яка відрізняється від відомих синтезом та аналізом ML-таблиці істинності на основі характеристичного рівняння тестування T⊕F⊕L = 0, що дає можливість паралельно виконувати логічні операції пошуку шкідливих кодів у локальному кіберпросторі.

Пропонується метод пошуку шкідливих кодів шляхом паралельного виконання трьох логічних операцій на матричних структурах даних. Але перед тим слід розглянути декілька корисних фрагментів теорії. Для аналізу таблиць шкідливих кодів (ШК) існує: алгоритм аналізу стовпців (метод наближення), а також алгоритм аналізу рядків (метод виключення). Останній зводиться до виконання одного рядка коду, яка визначає деструктивний стан шляхом виконання двох паралельних логічних операцій and-or над одиничними і нульовими рядками таблиці (рис. 3).



M_{ij}	M_1	M_2	M_3	M_j	M_m	R_i
T_1	M_{11}	M_{12}	M_{13}	M_{1j}	M_{1m}	0
T_2	M_{21}	M_{22}	M_{23}	M_{2j}	M_{2m}	1
T_3	M_{31}	M_{32}	M_{33}	M_{3j}	M_{3m}	0
T_i	M_{i1}	M_{i2}	M_{i3}	M_{ij}	M_{im}	1
T_n	M_{n1}	M_{n2}	M_{n3}	M_{nj}	M_{nm}	0

Mj – Malware, Ti – test, Ri – response

Рис. 3. Матрично-тестовий метод пошуку malware

Формули для визначення деструктивних станів по таблиці (матриці) мають такий вигляд:

$$L_s = \bigwedge_{R_i=1} M_i \wedge \overline{\bigvee_{R_i=0} M_i}; L_m = \bigvee_{R_i=1} M_i \wedge \overline{\bigvee_{R_i=0} M_i}.$$

Слід зазначити, що від таблиці ВК можна легко перейти до ML-таблиці істинності, а далі вирішувати питання ідентифікації ВК-станів методами машинного навчання, що дуже ефективно, особливо для складних CPS. Крім того, таблиця ШК, як матрична форма, є технологічно зручною структурою даних для еквівалентування деструктивів з метою синтезу дерева пошуку ШК. З огляду на те, що класи еквівалентності, як правило, мають неоднакову потужність, то розумне додавання додаткових тестів або асерційних точок до вузьких місць кіберструктури може привести дерево до зваженого вигляду.

Отже, вектор, таблиця і матриця є найбільш технологічними структурами даних, до яких слід приводити великі дані для їх подальшої тривіальної обробки. Тому ML-Computing, який оперує таблицями, є більш перспективним, ніж структурно складні нейромережі.

Таким чином, удосконалено матрично-логічний метод діагностування шкідливих кодів, який відрізняється від відомих технологій паралельним виконанням логіки алгоритму над рядками і стовпцями попередньо синтезованої таблиці функцій деструктивних компонентів, що дає можливість підвищити продуктивність edge-комп'ютингу користувача.

Пропонується векторний метод пошуку кратних деструктивностей D_m на основі обчислення теоретико-множинної різниці двох векторів-рядків матриці, що відповідають об'єднанню одиничних і нульових реакцій спостережуваних виходів на вхідному тесті перевірки ШП:

$$D_m = \bigcup_{\forall R_i=1} M_i \setminus \bigcup_{\forall R_i=0} M_i = \bigvee_{\forall R_i=1} M_i \wedge \overline{\bigvee_{\forall R_i=0} M_i}.$$

Алфавіт опису деструктивних станів CPS має вигляд: A = {0,1, X = {0,1}, ∅}, де коди символів складають множину: K (A) = {10,01,11,00}. Структури даних подані матрицею ШП на декартовому добутку множини тестових наборів і множини деструктивів в об'єкті діагностування, де кожна клітинка є двохбітовий (багатобітовий – сигнатура) код: перший з них ідентифікує деструктивність A, що перевіряється, а другий – B. Суперпозиція ШП (дві одиниці на одній лінії-осередку) дає можливість істотно мінімізувати структури даних для зберігання інформації з метою подальшого пошуку ШП при виконанні діагностичного експерименту в режимі online. Для перевірки векторного методу пошуку ШП пропонується таблиця, яка формує технологічні матричні структури даних, а також виконання діагностичного експерименту на основі об'єднання сукупності ШП-рядків з векторами в осередках, які формують некоректні стани виходів CPS на тестових наборах {T1-R10; T5-R11; T6- (R10, R11); T8-R11}:

M=<T,F>	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	R10	R11
T1	0	1	1	0	0	1	0	0	0	1	0	1	0
T2	1	0	0	0	1	0	0	0	0	1	0	1	0
T3	0	0	0	1	0	1	0	0	0	0	1	1	0
T4	1	0	0	0	1	0	0	1	0	0	1	1	0
T5	0	0	1	0	0	0	1	0	0	1	0	0	1
T6	0	1	1	0	0	0	0	1	0	0	1	1	1
T7	0	1	0	0	0	1	0	0	0	1	0	0	0
T8	0	0	1	0	1	0	1	1	0	0	0	0	1
$F_m^1=\bigvee_{\forall R_i=1} T_i$	0	1	1	0	1	1	1	1	0	1	1	1	1
$F^0=\bigvee_{\forall R_i=0} T_i$	1	1	0	1	1	1	0	1	1	1	1	0	0
$D_m=F^1 \wedge \overline{F^0}$	0	0	1	0	0	0	1	0	0	0	0		
$D_m=$	.	0	.	1	.	.	.	0	.	.	.		

Тут паралельне виконання операції диз'юнкції для рядків T1, T5, T6, T8 формує вектор F_m^1, який збирає всі можливі malware, що перевіряються на тестових наборах. Вектор F^0, отриманий за допомогою паралельної операції диз'юнкції над рядками T2, T3, T4, T7, об'єднує всі неможливі, непровірювані на тестових наборах malware. Віднімання всіх неможливих з усіх можливих malware дає шуканий результат у вигляді трьох ШП, закодованих в таблиці як F2 = 10; F4 = 01; F8 = 10. Таким чином, паралельне виконання двох реєстрових or-операцій на основі результатів проведеного діагностичного експерименту дозволило визначити три можливих деструктивних компонента, кожен з яких має місце в CPS: $D_m = \{F10, F41, F80\}$.

Таким чином, удосконалено векторно-матричний метод діагностування шкідливих кодів, який відрізняється від відомих технологій векторним поданням координат матриці деструктивних компонентів, що дає можливість підвищити продуктивність алгоритму аналізу багатозначних або сигнатурних даних шляхом паралельного виконання трьох логічних операцій.

В **третьому розділі** пропонується метод детектування модифікованих шкідливих кодів, заснований на реалізації евристичного аналізатора шкідливого коду за допомогою штучної нейронної мережі. Для організації методу детектування потрібно виконати такі дії: 1) провести запуск під API-монітором або емулятором шкідливого об'єкта; 2) отримати лог-файл, який містить послідовність виконання функцій і передані аргументи; 3) виконати процедуру аналізу на "схожість" функціональної поведінки програмного об'єкта з відомими даними про функціональну поведінку інших об'єктів, що зберігаються в бібліотеці; 4) виділити поведінкові групи серед програмних об'єктів, що мають подібну шкідливу поведінку, і зробити висновок про

належність / неналежність розглянутого зразка до деякого сімейства шкідливих програм. Розробляється система прийняття рішень на основі багатошарового персептрона, що дозволяє з високою часткою ймовірності виконувати завдання виявлення поліморфного модифікованого шкідливого коду в програмних об'єктах. Експериментальні дані були проведені на родинах поштових черв'яків Email-Worm.Win32.Warezov і на троянцях, призначених для крадіжки паролів Trojan-PSW.Win32.LdPinch. Практична реалізація запропонованого методу показала його високу точність і швидкість. Система здатна детектувати нові модифікації шкідливого коду, грунтуючись на даних, отриманих з навчальної вибірки. При появі абсолютно нових модифікацій потрібно додати відповідні ознаки і перенавчити нейронну мережу.

У **четвертому розділі** розглядаються моделі та методи діагностування zero-day загроз в кіберпросторі для підвищення ефективності виявлення складних шкідливих комплексів, що використовують поліморфні мутатори. Пропонується метод детектування досліджуваного зразка антивірусними рішеннями за допомогою публічного і локального мультисканера. Розробляється метод діагностування поліморфних шкідливих програм за допомогою Yara правил. Описується багатокомпонентний сервіс, що дозволяє організувати безкоштовне рішення аналізу шкідливих програм з гібридною архітектурою розгортання в публічних і приватних хмарах. Виконується проектування хмарного сервісу для детектування шкідливих програм на основі «пісочниць» з відкритим вихідним кодом і MAS, що дозволяє горизонтально масштабуватися в гібридних хмарах і показує високу пропускну здатність при обробці потоку шкідливих і нешкідливих об'єктів. Основним завданням сервісу є збір артефактів після динамічного і статичного аналізу досліджуваного об'єкта для детектування zero-day загроз. Показується ефективність запропонованих рішень.

Для вирішення завдань детектування поліморфних вірусів пропонується хмарний сервіс, який містить три основних модуля:

1) Сервер, управляє вхідним і вихідним потоком даних, проводить статичний аналіз досліджуваних об'єктів, сканування за допомогою Yara правил і сервісу Virustotal, аналізує артефакти після динамічного аналізу, виконує сканування зібраних зразків пам'яті і трафіка за допомогою Yara правил та формує вердикт про наявність деструктивної складової. Всі зібрані артефакти зберігаються в базі даних MySQL.

2) Мультисканер дозволяє просканувати файл за допомогою встановлених антивірусних рішень, заздалегідь сконфігурованих не відправляти зразки на аналіз в антивірусні лабораторії, тим самим дозволити перевірку сигнатурними сканерами чутливих файлів підприємства, які не можна відправляти в публічні мультисканери.

3) «Пісочниця», віртуальне чи реальне середовище з попередньо встановленою операційною системою, агентом для запуску шкідливої програми і

плагінами для динамічного аналізу. Плагіни збирають дампи пам'яті з усіх процесів, що запускаються, і коду, запровадженого в адресний простір процесів у системі, записують мережеву, файлову активність Windows реєстру, шукають аномалії в host файлі, які містять базу даних доменних імен і використовують при їх трансляції в мережеві адреси вузлів, аналізують встановлені перехоплювачі windows hooks, збирають інформацію про створених унікальних ідентифікаторів у системі, зберігають знімки екрану при роботі шкідливої програми.

Основні компоненти сервісу для аналізу шкідливих програм MAS забражені на рис. 4.

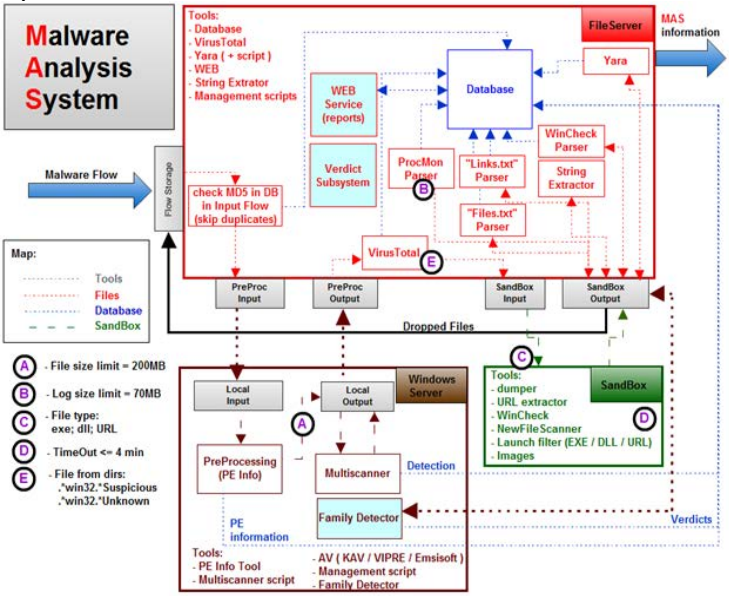


Рис. 4. Основні компоненти хмарного сервісу MAS

«Пісочниця» може використовувати інтернет емулятор, якщо пряме підключення до мережі Інтернет блокує шкідливий трафік на стороні провайдера. За наявності доступу до мережі Інтернет на сервері встановлена система детектування вторгнень Suricata, яка дозволяє детектувати аномалії в мережевому трафіку.

Основною перевагою цього модуля є те, що агент і плагіни можна запустити в ізольованому реальному середовищі і створити спеціальний примірник для аналізу високотехнологічних шкідливих комплексів, які блокують свій запуск у віртуальному середовищі для обходу виявлення. Модуль дозволяє налаштувати гібридну архітектуру на базі реальних і віртуальних машин як публічній хмарі, так і приватній. При цьому реальна машина поміщається в

повністю ізольований VLAN, а зв'язок з хмарою налаштовується через Site-to-Site VPN.

Хмарний сервіс аналізу шкідливих програм CUCKOO. Cuckoo Sandbox – це система автоматичного дослідження шкідливих програм, експлойтів, скриптів, документів, архівів і посилань з відкритим вихідним кодом [8]. У спеціально підготовленій віртуальній системі встановлюється агент Cuckoo і додається в автозавантаження, який буде взаємодіяти з «пісочницею». Спеціально налаштовуються інтерфейси мережі для перехоплення і подальшого аналізу мережевого трафіку. Після всіх маніпуляцій робиться знімок файлової системи Snapshot. «Пісочниця» завантажує тестований файл, визначає його тип і відповідно до нього виробляє необхідні маніпуляції. Всі зміни всередині «пісочниці» фіксуються в звіт. Після роботи система відновлює Snapshot, і віртуальне середовище повертається до свого початкового стану (рис. 5).

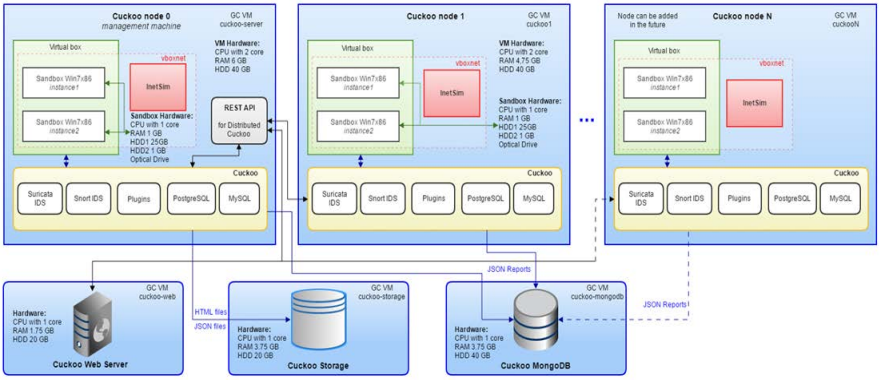


Рис. 5. Основні компоненти розподіленого хмарного сервісу Cuckoo

Для Cuckoo системи були розроблені додаткові плагіни, що значно розширюють функціональні можливості вихідного рішення при виявленні zero-day шкідливих програм.

Додані плагіни взяті з системи MAS і адаптовані для роботи в Cuckoo, виконують пошук аномалій в hosts файлі, зберігають дампи пам'яті шкідливих процесів і коду, впровадженого в адресний простір запущених процесів у системі. Моніторинг шкідливої rootkit активності виконується в модулі Wincheck.

Для детектування шкідливої активності використовуються Yara правила для дампів пам'яті, мережевого трафіка і самого виконуваного файлу. Yara сигнатури для дампів пам'яті дозволяють успішно декатувати шкідливі програми Rbot, Blazebot, Nrgbot, а також їх нові поліморфні модифікації, які слабо детектуються антивірусними рішеннями.

Пропонуються методи для аналізу і діагностування zero-day загроз:

1) Розроблено автоматичну систему прийняття рішень Verdictor, що дозволяє сформувати висновок про належність досліджуваного зразка до шкідливого об'єкта, грунтуючись на знайдених аномаліях у фрагментах трафіка, зразках пам'яті і виконуваному файлі за допомогою Yara правил. Метод включає: підготовку зразків мережевого трафіку, рядків, витягнутих з дампів пам'яті шкідливого об'єкта; підготовку унікального сету рядків, які аналізують CharsRate, TotalTrashRate, і кількість символів у рядку; виключення рядків, що входять в словник Exclusion; формування Yara правила; тестування правила на базі артефактів MAS для пошуку поліморфних об'єктів; рекурсивний пошук однакових рядків, виявлених у вибірці самплів після першої ітерації за допомогою difflib; оновлення Yara правила; інтеграцію правила у вхідний набір.

2) Розроблено систему для статичного аналізу шкідливого зразка за допомогою вбудованого мультисканера в MAS. Вона дозволяє додавати доступні на ринку Endpoint Protection антивірусні програми для локального детектування файлів і містить такі компоненти: інстанси з операційною системою Windows; встановлені Endpoint антивірусні рішення, максимум 3 на одному екземплярі; керуючий python скрипт для відправки на аналіз файлу і отримання вердикту антивірусу; збереження результатів у av_detects_table в MySql базі артефактів MAS.

3) Розроблено багатофункціональний Sanbox модуль на базі операційної системи Winows 7 і vmware VIX API, що дозволяє виконувати аналіз шкідливих і нешкідливих об'єктів з використанням модульної (плагін) архітектури для розширення функціональності на мовах програмування C ++, Perl, Python, Autoit. Модуль включає в себе наступні компоненти: Dumper – дампи нових процесів у системі і дампи пам'яті інжектів у процеси; String from dumps – зберігає терміни з дампів пам'яті та інжект; URL extractor – витягує з Wireshark логів усі знайдені URL адреси; ProcMon – плагін, що легує зміну файлової системи і реєстру Windows; Microsoft Sysinternals компонент; WinCheck – опенсорс антируткіт утиліта, що детектує аномалії в недокументованих або не повністю документованих Windows internal structures. Модуль дозволяє збирати інформацію про user-mode і kernel-mode руткітів; Screenshoter – модуль робить скріншоти вікон шкідливих програм; File Handler – зберігає нові витягнуті з тіла шкідливих програм файли або завантажені файли під час роботи; Status Handler – створює файл про розмір з MD5 хешем аналізованого об'єкта в каталозі SandboxOutput, що є індикатором успішного завершення аналізу: \\192.168.50.163\sandboxoutput\2021-03-03\0f041ac5d7d1acb3a9280743a909c330_done_15-35; Hosts file anomalies checker – збирає аномалії в hosts файлі операційної системи Windows; Mutexes checker – є компонентом Microsoft Sysinternals. Модуль збирає інформацію про створювані унікальні ідентифікатори шкідливої програми в операційній системі Windows; Launcher – запускає на аналіз такі файли: EXE, DLL, URL, DOC, PPT, PDF.

Порівняння результатів. Система аналізу шкідливих програм MAS дозволяє побудувати власну вірусну лабораторію для виявлення нових деструктивних об'єктів, що не детектуються антивірусними рішеннями, збирає URL адреси з досліджуваних файлів для автоматичного блокування трафіка до керуючих CnC серверів на IDS / IPS рішеннях. Дозволяє автоматично генерувати опис для шкідливих програм з покроковими інструкціями для видалення. База артефактів шкідливих програм дозволяє написати власне рішення для автоматичного видалення нової загрози при активному зараженні кіберпростору, коли антивірусні рішення не мають відповідних сигнатур.

Перевага системи MAS перед існуючими аналогами: 1) MAS не має фактичного обмеження для аналізу шкідливого файлу великого розміру 1GB і більше. Sandbox інстанси легко вертикально масштабуються і додають потрібну кількість ресурсів. За замовчуванням розмір файлу обмежений 200 MB, але при необхідності ці обмеження знімаються. Ліміти на розмір файлів у існуючих рішеннях, навіть в платних підписках, обмежені 100 MB. 2) Час аналізу шкідливого файлу в системі MAS становить 4 хвилини на одному інстансі. Для збільшення швидкості обробки вхідного потоку додаються нові Sandbox істанси. При цьому навіть платні аналоги мають обмеження на 20 реквестів на хвилину. 3) MAS містить локальний мультисканер, при цьому існуючі аналоги використовують тільки Virustotal результати з наявними обмеженнями 4 запити за хвилину. Основна особливість мультисканера полягає у спеціальних налаштуваннях системи, які не дозволяють відправляти чутливі дані в громадські організації.

Таким чином, запропоновано новий метод детектування досліджуваного зразка заздалегідь встановленими антивірусними рішеннями, які дозволяють в окремому потоці проводити статичне сканування досліджуваного об'єкта без обмежень на кількість запитів за хвилину і, тим самим, підвищити швидкість обробки об'єктів та обмежити публічний доступ до конфіденційних файлів;

Запропоновано новий метод діагностування поліморфних шкідливих програм за допомогою Yara правил, що дозволяє детектувати нові модифікації, які не виявляються доступними рішеннями.

Запропоновано гібридну архітектуру системи, яка дозволяє проводити ретроспективний пошук за сімействами, відстежуючи зміни в деструктивних компонентах, збирати базу шкідливих URL адрес для блокування трафіка до керуючих серверів, збирати витягнуті завантажені файли, аналізувати вкладення в фішингові листи, інтегруватися з SIEM, IDS, IPS, антифішинг і Honeypot системами, поліпшити якість роботи SOC аналітика і час реакції на інциденти, упереджувати атаки зловмисників і блокувати нові загрози, що не детектуються доступними антивірусними рішеннями.

Практичне значення одержаних результатів полягає в розробці хмарного сервісу, що об'єднує MAS Sandbox і модифікований розподілений аналізатор

шкідливих програм Cuckoo, який дозволяє швидко реагувати на zero-day загрози, зберігати базу знань для кореляції артефактів між поліморфними зразками шкідливих програм, активно шукати нові зразки шкідливих програм та інтегруватися з програмно-апаратними комплексами захисту кіберпростору, що підтримують Cuckoo API. Запропонований хмарний сервіс дозволяє активно застосовувати нейромережні технології виявлення шкідливих об'єктів, збирати артефакти для створення Yara правил і детектувати поліморфні шкідливі програми.

Розроблено багатокомпонентний сервіс, що дозволяє організувати безкоштовне рішення аналізу шкідливих програм з гібридною архітектурою розгортання в публічних і приватних хмарах. MAS може працювати з доступом до мережі Інтернет та без, використовуючи емулятор, показуючи високу продуктивність при аналізі шкідливого потоку у порівнянні з розглянутими рішеннями, детектуючи складні шкідливі комплекси, що використовують поліморфні мутатори, і zero-day уразливості для свого поширення.

У **п'ятому розділі** пропонується метод створення URL сигнатур нового покоління, що дозволяють успішно детектувати нові шкідливі URL на скомпрометованих легальних серверах, точково блокуючи доступ до шкідливого об'єкта, при цьому не блокуючи весь ресурс, що дає можливість скоротити розмір бази даних на 75%. Основна особливість полягає у створенні URL сигнатур, витягнутих зі шкідливих самплів, а також збиранні нових шкідливих URL артефактів за допомогою Heretrix краулер.

Головна мета при створенні URL сигнатур – відфільтрувати нешкідливі артефакти, щоб мінімізувати False Positive виявлення. Для виявлення і видалення легітимних сайтів з основного потоку URL адрес в MAS додані найбільш відвідувані URL з бази даних Alexa. При цьому система MAS включає в себе: 1) краулер Heretrix, який приймає на вхід 100% шкідливі URL адреси і дозволяє збирати нові об'єкти; 2) локальний мультисканер для URL, який дозволяє аналізувати контент і отримувати вердикт від антивірусних рішень; 3) Google safebrowsing verdictor; 4) Projecthoneypot verdictor.

Замкнутий цикл аналізу шкідливих самплів, завантаження нових самплів за зібраними URL адресами, збір нових URL за допомогою краулерів, детектування URL за допомогою вбудованих в MAS систем дозволяє організувати постійний потік нових шкідливих URL для попереджувального детектування невідомих загроз. Виникає нова проблема: база даних шкідливих URL артефактів стає дедалі більшою і уповільнює роботу сканерів. При цьому в потоці існують шкідливі програми, які використовують для свого поширення легітимні рішення, наприклад файлове сховище Dropbox, в ботнетах Rbot, Blazebot, Nrgbot, що ускладнює детектування і додає False negative спрацьовування. Детектування URL за допомогою вбудованих в MAS систем дозволяє організувати постійний потік нових шкідливих URL для попереджувального детектування невідомих загроз.

Вирішити задачу виявлення шкідливих об'єктів на легітимних ресурсах, зменшити розмір бази даних шкідливих URL, скоротити False Positive і False negative виявлення, збільшити якість детектування нових шкідливих об'єктів допоможе гібридний підхід, що складається з URL сигнатур 1:1 і 1:many.

Для досягнення мети – істотного підвищення ефективності виявлення шкідливих URL адрес при атаках складних загроз, що не декатуються, вирішені такі задачі: 1) розробка методу детектування URL адреси з використанням сигнатур 1:many для URL path, URL domain, URL subdomain, URL parameter, URL file; 2) створення 1:many сигнатур для сканування URL за допомогою регулярних виразів; 3) створення сканера URL з базою шкідливих URL адрес нового покоління. З метою поліпшення роботи з базою даних нового покоління 1:many URL сигнатур був створений модуль для автоматичної генерації регулярного виразу. Регулярні вирази можуть використовуватися для пошуку певного патерну в рядку і на відміну від сигнатур зі знаком "*" не вимагають складної логічної обробки. Для кожного 1:many правила модуль створює регулярний вираз. URL сигнатури нового покоління були протестовані на базі шкідливих URL з системи MAS і дозволили скоротити розмір бази даних на 75% (рис. 6). База даних меншого розміру з проактивним детектуванням дозволяє швидко і якісно відхиляти сучасні загрози та мінімізувати ризики зазнати збитків компаніям від діючих вірусних технологій.

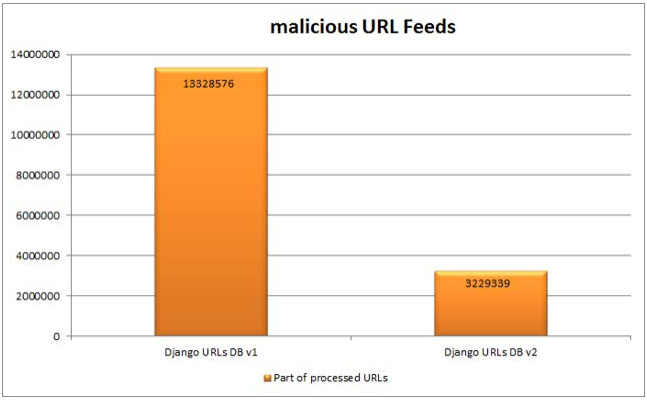


Рис. 6. Статистика кількості URL сигнатур до і після впровадження 1: many

Розроблено методику оцінки збитків від поширення деструктивних вірусних технологій в комп'ютерних мережах підприємств (IT-компаній), яка включає моделі фінансового збитку та інформаційної безпеки, що дозволяє оцінити ефективність інвестицій. Виконана програмна реалізація, де передбачено збереження результатів розрахунків в базу даних для подальшого аналізу роботи підприємства. У сукупності розроблена система дозволяє дати якісну оцінку продуктивності праці і виявити збитки компанії від дії шкідливих програм.

ВИСНОВКИ

Проведене дослідження вирішує науково-практичну задачу – розпізнавання поліморфних мутаторів, що автоматично змінюють синтаксис і логіку роботи при кожній активації.

Для досягнення поставленої мети – істотного зменшення часу і вартості розпізнавання поліморфних мутаторів шляхом розробки і впровадження федеративної архітектури cloud-edge комп'ютингу на основі ML-sandbox і векторно-логічних методів пошуку zero-day шкідливих кодів для захисту інфраструктури кіберфізичного простору – в роботі були вирішені задачі, які дозволили отримати результати, що мають наукову новизну:

1) Нова федеративна ML-архітектура sandbox комп'ютингу, яка характеризується федеративним розподілом в просторі терміналів машинного навчання на основі «пісочниць». Це дозволяє істотно знизити навантаження на канали передачі даних шляхом локальної обробки підозрюваних кодів, підвищити продуктивність сукупного cloud-edge computing, зменшити час навчання і тестування глобальної хмарної «пісочниці», підвищити якість розпізнавання шкідливих кодів, зберегти цілісність і конфіденційність даних на терміналах користувачів.

2) Удосконалена структурна модель ML-комп'ютингу, яка відрізняється від відомих синтезом та аналізом ML-таблиці істинності на основі характеристичного рівняння тестування $T \oplus F \oplus L=0$. Це дає можливість паралельно виконувати логічні операції пошуку шкідливих кодів у локальному кіберпросторі.

3) Удосконалений матрично-логічний метод діагностування шкідливих кодів, що відрізняється від відомих технологій паралельним виконанням логіки алгоритму над рядками і стовпцями попередньо синтезованої таблиці функцій деструктивних компонентів. Це дозволяє підвищити продуктивність edge-комп'ютингу користувача.

4) Удосконалений векторно-матричний метод діагностування шкідливих кодів, який відрізняється від відомих технологій векторним поданням координат матриці деструктивних компонентів. Це дає можливість підвищити продуктивність алгоритму аналізу багатозначних або сигнатурних даних шляхом паралельного виконання трьох логічних операцій.

5) Нові методи:

– детектування модифікованих шкідливих кодів, що заснований на реалізації евристичного аналізатора шкідливого коду за допомогою штучної нейронної мережі;

– детектування досліджуваного зразка заздалегідь встановленими антивірусними рішеннями, які дозволяють в окремому потоці проводити статичне сканування досліджуваного об'єкта без обмежень на кількість запитів

на хвилину, підвищити швидкість обробки об'єктів і обмежити публічний доступ до конфіденційних файлів;

– діагностування поліморфних шкідливих програм за допомогою Yara правил, що дає можливість детектувати нові модифікації, які не виявляються доступними рішеннями;

– створення URL сигнатур нового покоління, що дозволяють успішно детектувати нові шкідливі URL на скомпрометованих легальних серверах, точково блокуючи доступ до шкідливого об'єкта, при цьому не блокуючи весь ресурс. Це дозволяє скоротити розмір бази даних на 75%.

Практичне значення одержаних результатів досліджень визначається:

– тестуванням, верифікацією і впровадженням розроблених програмних засобів перевірки, діагностування шкідливих програм (поліморфних мутаторів), що характеризуються механізмом управління логічною і синтаксичною модифікацією коду шкідливої програми для її маскування від детектування існуючими антивірусними сервісами. Розроблені методи аналізу шкідливих програм успішно виявляли атаки під час їх моделювання, а також підозрілі файли, URL-адреси та електронні листи та зменшували час реакції на інциденти;

– розробкою хмарного сервісу, що об'єднує MAS Sandbox і модифікований розподілений аналізатор шкідливих програм Cuckoo, який дозволяє швидко реагувати на zero-day загрози, зберігати базу знань для кореляції артефактів між поліморфними зразками шкідливих програм, активно шукати нові зразки шкідливих програм та інтегруватися з програмно-апаратними комплексами захисту кіберпростору, що підтримують Cuckoo API. Запропонований хмарний сервіс дозволяє активно застосовувати нейромережні технології виявлення шкідливих об'єктів, збирати артефакти для створення Yara правил і детектувати поліморфні шкідливі програми, а також горизонтально масштабуватися в гібридних хмарах;

– розробкою методики оцінки збитків від поширення деструктивних вірусних технологій в комп'ютерних мережах підприємств (IT-компаній), яка включає моделі фінансового збитку та інформаційної безпеки, що дозволяє оцінити ефективність інвестицій. Виконана програмна реалізація передбачає збереження результатів розрахунків в базу даних для подальшого аналізу роботи підприємства. Розроблена система підрахунку фінансових втрат від простою комп'ютерної мережі дозволяє дати якісну оцінку продуктивності праці і виявити збитки компанії від дії шкідливих програм;

– впровадженням розроблених моделей, методів та інфраструктури у навчальний процес Харківського національного університету радіоелектроніки (акт про впровадження від 27.06.2021); у виробничу діяльність IT компанії EPAM, USA (довідка від 10.06.2021), а саме: критичні компоненти "Системи автоматичного аналізу шкідливих програм" в рамках

проекту внутрішньої безпеки EPAM інтегровано в системи безпеки EPAM (Cofence Антифішинг, TrapX Honeypot, Cisco Umbrella DNS-фільтрація).

Переваги системи MAS перед існуючими аналогами: 1) Відсутність обмеження для аналізу шкідливого файлу великого розміру 1GB і більше. Sandbox інстанси легко вертикально масштабуються і додають потрібну кількість ресурсів. За замовчуванням розмір файлу обмежений 200 MB, але при необхідності ці обмеження знімаються. Ліміти на розмір файлів у існуючих рішеннях, навіть в платних підписках, обмежені 100 MB. 2) Час аналізу шкідливого файлу в системі MAS становить 4 хвилини на одному інстансі. Для збільшення швидкості обробки вхідного потоку додаються нові Sandbox інстанси. При цьому навіть платні аналоги мають обмеження на 20 реквестів на хвилину. 3) Наявність локального мультисканера, при цьому існуючі аналоги використовують тільки Virustotal результати з наявними обмеженнями 4 запити на хвилину. Основна особливість мультисканера полягає у спеціальних налаштуваннях системи, які не дозволяють відправляти чутливі дані в громадські організації.

Запропоновано гібридну архітектуру системи, яка дозволяє проводити ретроспективний пошук за сімействами, відстежуючи зміни в деструктивних компонентах, збирати базу шкідливих URL адрес для блокування трафіка до керуючих серверів, збирати витягнуті завантажені файли, аналізувати вкладення в фішингові листи, інтегруватися з SIEM, IDS, IPS, антифішинг і Honeypot системи, поліпшити якість роботи SOC аналітика і час реакції на інциденти, упереджувати атаки зловмисників і блокувати нові загрози, що не детектуються доступними антивірусними рішеннями.

СПИСОК ОПУБЛІКОВАНИХ РОБІТ ЗА ТЕМОЮ ДИСЕРТАЦІЇ

Список публікацій здобувача, в яких опубліковані основні наукові результати дисертації:

1. *Сапрыкин А.С.* Нейросетевые методы обнаружения вредоносного кода в программных объектах [Текст] / *А.С. Сапрыкин* // Восточно-Европейский журнал передовых технологий. – 2009. – №2/3 (38). – С. 51-55. (Фахове видання. Журнал реферується або індексується міжнародними базами Google Scholar, URAN, Національною бібліотекою України ім. В. І. Вернадського).

2. *Сапрыкин А.С.* Методика оценки убытков предприятия от вредоносных программ [Текст] / *А.С. Сапрыкин*, М.В. Бочарникова, А.С. Адамов // Вестник национального технического университета "ХПИ" (Новое решение в современных технологиях). – 2009. – №8. – С. 58-64. (Фахове видання. Журнал реферується або індексується міжнародними базами Google Scholar, Національною бібліотекою України ім. В. І. Вернадського).

3. Adamov A. Analysis and Detection of Polymorphic Spyware [Text] / A. Adamov, *A. Saprykin* // Hakin9 Magazine. – 2013. – Vol. 8, № 01. – Issue 01/2013 (61). Warsow: Software Press, 2013. – P. 6-11. (Закордонне видання).

4. *Saprykin A.S.* Models and methods for diagnosing Zero-Day threats in cyberspace. Herald of Advanced Information Technology. 2021; Vol. 4, No.2: 155–167. (Фахове видання. Журнал включений в міжнародні бібліотеки та наукометричні бази Academia.edu, ROAD, Національна бібліотека України ім. В. І. Вернадського, Djerelo, Україніка наукова, Index Copernicus).

5. Хаханов В.І. Моделі і методи пошуку шкідливих кодів на архітектурі федеративного машинного навчання [Текст] / В.И. Хаханов, *А.С. Саприкін* // Міжнародний науковий журнал "Modern Scientific Researches." – Мінськ, Білорусь. – 2021. – Вип. 16. – С. 25-38. (Закордонне видання. Журнал реферується або індексується міжнародними базами Index Copernicus, Google Scholar).

6. *Саприкин О.С.* Сучасні технології пошуку вірусів, федеративні алгоритми та архітектури (аналітичний огляд) [Текст] / *О.С. Саприкін* // Радіоелектроніка та інформатика. – 2020. – №4. – С. 11-20. (Журнал реферується або індексується міжнародними базами Index Copernicus, Google Scholar, Scholar Steer, Cyberleninka, TIU Hannover, I2OR, Національною бібліотекою України ім. В.І.Вернадського).

Результати, які засвідчують апробацію матеріалів дисертації:

7. Adamov A. The Problem of Trojan Inclusions in Software and Hardware [Text] / A. Adamov and *A. Saprykin* // Proc. of IEEE 2010 East-West Design & Test Symposium (EWDTS 2010) – Petersburg, RF. – 17-20 Sept. 2010. – P. 449-451. (Входить до міжнародних наукометричних баз Scopus, IEEE Xplore).

8. Adamov A. The problem of Hardware Trojans detection in System-on-Chip [Text] / A. Adamov, *A. Saprykin,* D. Melnik and O. Lukashenko // IEEE 2009 10th International Conference – The Experience of Designing and Application of CAD Systems in Microelectronics (CADSM 2009). – Lviv-Polyana, Ukraine. – 24-28 Feb 2009. – P. 178-179. (Входить до міжнародних наукометричних баз Scopus, IEEE Xplore).

9. *Saprykin A.* Diagnosis method of malicious code in executable files / *A. Saprykin,* V. Kiktenko, S. Galagan, A. Kunitsky // Proceedings of the 5th East-West Design and Test Workshop. – Yerevan, Armenia. – 7-10 Sept. 2007. – P. 664-667.

10. *Сапрыкин А.С.* Тенденции развития вредоносного программного обеспечения для мобильных устройств [Текст] / *А.С. Сапрыкин*, В.А. Федосеев // Матеріали XI Міжнародного молодіжного форуму "Радіоелектроніка та молодь в XXI сторіччі". – Ч. 2. – Харків, Україна. – 10-12 квітня 2007.– С. 240.

11. Федосеев В.А. Методы борьбы с вредоносным программным обеспечением для мобильных устройств [Текст] / В.А. Федосеев, *А.С. Сапрыкин* // Матеріали XI Міжнародного молодіжного форуму

“Радіоелектроніка та молодь в XXI сторіччі”. – Ч. 2. – Харків, Україна. – 10-12 квітня 2007. – С. 257.

12. Бочарникова М.В. Разработка методики оценки ущерба от распространения вирусных технологий на действующих предприятиях [Текст] / М.В. Бочарникова, *А.С. Сапрыкин* // Студенческая конференция IT Security For New Generation. – Москва, РФ – 28-29 августа 2008. – С. 41.

13. Bocharnikova M.V. Calculation of an enterprise's financial losses caused by malware activities [Text] / M.V. Bocharnikova, *A.S. Saprykin*, V.A. Kiktenko, A.S. Adamov // II International Student Conference on Computer Security Issues “It Security For The New Generation” – Moscow, RF. – 28-29 April 2009. – P. 188-190.

14. Adamov A.S. Training course "Computer threats: detection and analysis methods" [Text] / A.S. Adamov, M.V. Kudina, O.A. Chuvilo, *A.S. Saprykin* // II International Student Conference on Computer Security Issues “It Security For The New Generation” – Moscow, RF. – 28-29 April 2009. – P. 158-160.

15. *Саприкін О.* Евристичний аналізатор шкідливого коду на основі штучної нейронної мережі [Текст] // Science, theory and practice. Abstracts of XXIX International Scientific and Practical Conference. – Tokyo, Japan. – 2021. – Pp. 560-563.

16. *Саприкін О.* Хмарний сервіс аналізу поліморфних шкідливих програм [Текст] / *О. Саприкін* // LXVIII Міжнародна інтернет-конференція «Літні наукові дискусії». – Україна, Чернівці. – 10 червня 2021. – С. 157-161.

АНОТАЦІЯ

Саприкін О.С. Моделі автоматизованого аналізу та діагностування поліморфних вірусів у комп’ютерних системах та мережах. – На правах рукопису.

Дисертація на здобуття наукового ступеня кандидата технічних наук (доктора філософії) за спеціальністю 05.13.05 – Комп'ютерні системи та компоненти. – Харківський національний університет радіоелектроніки, Міністерство освіти і науки України, Харків, 2021.

Мета дослідження – істотне зменшення часу і вартості розпізнавання поліморфних мутаторів шляхом розробки і впровадження федеративної архітектури cloud-edge комп'ютингу на основі ML-sandbox і векторно-логічних методів пошуку zero-day шкідливих кодів для захисту інфраструктури кіберфізичного простору. Поліморфний мутатор – механізм управління логічною і синтаксичною модифікацією коду шкідливої програми для її маскування від детектування існуючими антивірусними сервісами.

Наукова новизна результатів досліджень: 1) Вперше запропоновано федеративну ML-архітектуру sandbox комп'ютингу. 2) Удосконалено структурну модель ML-комп'ютингу. 3) Удосконалено матрично-логічний метод діагностування шкідливих кодів. 4) Удосконалено векторно-матричний метод діагностування шкідливих кодів. 5) Вперше запропоновано методи: детектування модифікованих шкідливих кодів; детектування досліджуваного зразка заздалегідь встановленими антивірусними рішеннями; діагностування поліморфних шкідливих програм за допомогою Yara правил; створення URL сигнатур нового покоління, що дає можливість скоротити розмір бази даних на 75%.

Ключові слова: федеративне машинне навчання, «пісочниця» шкідливого програмного забезпечення, хмарна «пісочниця», хмарні обчислення, зловмисне програмне забезпечення, кіберфізична система, логічно-векторний аналіз, ML-обчислення, виявлення шкідливого програмного забезпечення.

АННОТАЦИЯ

Сапрыкин А.С. Модели автоматизированного анализа и диагностирования полиморфных вирусов в компьютерных системах и сетях. – На правах рукописи.

Диссертация на соискание ученой степени кандидата технических наук (доктора философии) по специальности 05.13.05 «Компьютерные системы и компоненты». – Харьковский национальный университет радиоэлектроники, Министерство образования и науки Украины, Харьков, 2021.

Цель исследования – существенное уменьшение времени и стоимости распознавания полиморфных мутаторов путем разработки и внедрения федеративной архитектуры cloud-edge компьютинга на основе ML-sandbox и векторно-логических методов поиска zero-day вредоносных кодов для защиты инфраструктуры киберфизичного пространства. Полиморфный мутатор – механизм управления логической и синтаксической модификацией кода вредоносной программы для ее маскировки от детектирования существующими антивирусными сервисами. Научная новизна исследований: 1) Впервые предложена федеративная ML-архитектура sandbox компьютинга. 2) Усовершенствована структурная модель ML-компьютинга. 3) Усовершенствован матрично-логический метод диагностирования вредоносных кодов. 4) Усовершенствован векторно-матричный метод диагностирования вредоносных кодов. 5) Впервые предложены методы: детектирования модифицированных вредоносных кодов на основе реализации эвристического анализатора; детектирования исследуемого образца заранее установленными антивирусными решениями; диагностирования полиморфных вредоносных программ с помощью Yara правил; создания URL сигнатур нового поколения, что дает возможность сократить размер базы данных на 75%.

Ключевые слова: федеративное машинное обучение, «песочница» вредоносного программного обеспечения, облачная «песочница», облачные вычисления, вредоносное ПО, киберфизическая система, подпись, логически-векторный анализ, ML-вычисления, выявление вредоносного ПО.

ABSTRACT

Saprykin O.S. Models for automated analysis and diagnosis of polymorphic viruses in computer systems and networks. – On the rights of the manuscript.

Thesis for the degree of candidate of technical sciences (PhD) in specialty 05.13.05 – Computer systems and components. – Kharkov National University of Radio Electronics, Ministry of Education and Science of Ukraine, Kharkov, 2021.

The aim of the study is to significantly reduce the time and cost of recognizing polymorphic mutators by developing and implementing a federated cloud-edge computing architecture based on ML-sandbox and vector-logical methods for finding zero-day malicious codes to protect cyberspace infrastructure.

Scientific novelty of research results: 1) Proposed a federal ML-architecture sandbox computing. 2) Improved structural model of ML-computing. 3) Improved matrix-logical method of diagnosing malicious code. 4) Improved vector-matrix method of diagnosing malicious codes. 5) Proposed methods: detection modified malicious codes; detection of the test sample by pre-installed anti-virus solutions; diagnosing polymorphic malware using Yara rules; creation of URLs of signatures of new generation, it allows to reduce the size of a database by 75%.

Key words: federated machine learning, malware sandbox, cloud sandbox, cloud-edge computing, malware, cyber physical system, signature, logic-vector analysis, ML-computing, malware detection.

Підписано до друку 11.08.2021. Формат 60x84 1/16. Папір друк.; Умов. друк. арк. 0,9
Облік. вид. арк. 1,0. Зам. № б/н. Тираж 100 прим.
Надруковано у видавництві ЧП «Степанов В.В.»
61168, Харків, вул. Акад. Павлова, 311