МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ТЕРНОПІЛЬСЬКИЙ НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ
ІМЕНІ ІВАНА ПУЛЮЯ

Бобровнікова Кіра Юліївна

УДК 004.491. 2

ІНФОРМАЦІЙНА ТЕХНОЛОГІЯ ВИЯВЛЕННЯ БОТ-МЕРЕЖ У КОРПОРАТИВНИХ МЕРЕЖАХ НА ОСНОВІ АНАЛІЗУ DNS-ТРАФІКА

05.13.06 – інформаційні технології

АВТОРЕФЕРАТ
дисертації на здобуття наукового ступеня
кандидата технічних наук

Тернопіль-2017

Дисертацією є рукопис.

Робота виконана в Хмельницькому національному університеті Міністерства освіти і науки України

Науковий керівник: кандидат технічних наук, доцент
Савенко Олег Станіславович,
Хмельницький національний університет,
декан факультету програмування та комп'ютерних і телекомунікаційних систем

Офіційні опоненти: доктор технічних наук, професор
Крилов Віктор Миколайович,
Одеський національний політехнічний університет,
професор кафедри прикладної математики та інформаційних технологій

кандидат технічних наук, доцент
Козак Руслан Орестович,
Тернопільський національний технічний університет імені Івана Пулюя,
завідувач кафедри кібербезпеки

Захист відбудеться «22» лютого 2017 р. о 14.00 годині на засіданні спеціалізованої вченої ради К58.052.06 у Тернопільському національному технічному університеті імені Івана Пулюя за адресою: 46001, м. Тернопіль, вул. Руська, 56, ауд. 58.

З дисертацією можна ознайомитися у науково-технічній бібліотеці Тернопільського національного технічного університету імені Івана Пулюя за адресою: 46001, м. Тернопіль, вул. Руська, 56.

Автореферат розісланий «18» січня 2017 р.

Вчений секретар
спеціалізованої вченої ради М.Є. Фриз

ЗАГАЛЬНА ХАРАКТЕРИСТИКА РОБОТИ

Актуальність роботи. Одним з найбільш небезпечних видів шкідливого програмного забезпечення (ШПЗ) на сьогоднішній день є бот-мережі. Переважна більшість бот-мереж для керування інфікованими комп'ютерними системами (КС) використовує DNS. Динамічна географічно розподілена структура бот-мереж, застосування різноманітних технологій ухилення від виявлення, в тому числі на основі DNS, та можливість анонімного керування інфікованими КС ускладнює виявлення бот-мереж та вживання контрзаходів з метою припинення їх діяльності.

Перспективним напрямком є підходи виявлення бот-мереж на основі DNS. Перевагами методів на основі DNS є: можливість здійснення виявлення бот-мереж на стадії створення їх інфраструктури; порівняно невеликий обсяг трафіка, а тому зменшення потреби в обчислювальних ресурсах, необхідних для аналізу; можливість виявлення бот-мереж виключно на основі спостереження за роботою мереж, залишаючись непоміченим для зловмисника; на відміну від сигнатурних методів, здатність виявляти невідомі боти.

Дослідження в цій сфері проводять Хольц Т., Девід Дж. Малан, Ентонакакіс М., Дейгон Д., Хюнсанг Чоі, Дітріх Д., Ішайз Х., Погребенник В.Д., Котенко І.В., Рувінська В.М., Сиротинський О.І., Касперський Є.В., Касперські К., Собейкіс В.Г. Проте, оскільки відомі підходи виявлення бот-мереж на основі аналізу DNS-трафіка не враховують певні важливі фактори, які вказують на функціонування бот-мереж, це призводить до великої кількості хибних спрацювань.

При організації процесу виявлення бот-мереж на основі аналізу DNS-трафіка важливим питанням є можливість однозначної ідентифікації КС в мережі. Оскільки IP-адреса не може бути надійним ідентифікатором для КС в мережі, то в якості ідентифікаторів КС доцільно використовувати їх MAC-адреси за умови забезпечення запобігання підміни MAC-адрес. Для цього виникає необхідність доступу до мережного обладнання та контролю над ним, а саме до керованих комутаторів мережі. Тому в роботі розглядається саме корпоративна комп'ютерна мережа.

При здійсненні виявлення невідомих ботів в корпоративній мережі (КМ) виникає протиріччя між двома важливими показниками: інформаційною безпекою КМ та оперативністю роботи інформаційних систем (ІС) КМ. Для підвищення оперативності роботи ІС корпоративної мережі потрібно знижувати рівень хибних спрацювань, відкидаючи певну частину підозрілих об'єктів аналізу, проте з метою підвищення рівня інформаційної безпеки КМ потрібно забезпечити відповідну реакцію на такі підозрілі об'єкти, зменшуючи рівень пропусків ШПЗ.

Тому, для усунення недоліків відомих підходів запропоновано нові методи виявлення бот-мереж в корпоративних мережах, засновані на аналізі DNS-трафіка, які враховують нову базову ознаку синхронності DNS-запитів та залучають методи інтелектуального аналізу даних, що дозволяє підвищити рівень достовірності виявлення невідомих ботів та оперативність роботи ІС корпоративної мережі.

Зв'язок роботи з науковими програмами, планами, темами. Дослідження, представлені в дисертації, проводились в рамках держбюджетної НДР Хмельницького національного університету №4Б-2015 «Розвиток наукових та інженерних основ надійності електронної техніки шляхом удосконалення технології її тестування на вібрації та удари», виконавцем якої була автор дисертації.

Мета і задачі дослідження. Метою дисертаційної роботи є розроблення інформаційної технології виявлення бот-мереж в корпоративних мережах на основі аналізу DNS-трафіка для підвищення достовірності виявлення бот-мереж.

Для досягнення поставленої мети необхідно вирішити наступні задачі:

1) дослідити особливості функціонування бот-мереж з врахуванням системи доменних імен та розробити відповідні модель бот-мереж з врахуванням системи доменних імен, модель DNS-трафіка та модель процесу виявлення бот-мереж на основі аналізу DNS-трафіка;

2) проаналізувати сучасні інформаційні технології виявлення бот-мереж на основі аналізу DNS-трафіка з метою визначення шляхів підвищення достовірності виявлення бот-мереж;

3) розробити метод ідентифікації бот-мереж на основі їх групової активності в DNS-трафіку, який дозволить ідентифікувати інфіковані комп'ютерні системи, що здійснюють таку активність;

4) розробити метод виявлення бот-мереж, які застосовують технології ухилення від виявлення на основі DNS, який дозволить виявляти інфіковані комп'ютерні системи в мережі, що належать до бот-мереж, які використовують технології ухилення на основі DNS;

5) розробити інформаційну технологію виявлення бот-мереж на основі аналізу DNS-трафіка та дослідити достовірність виявлення бот-мереж;

6) розробити алгоритми та реалізувати інформаційну технологію виявлення бот-мереж на основі аналізу DNS-трафіка у вигляді програмного забезпечення та впровадити її у виробництво з метою підвищення достовірності виявлення бот-мереж у корпоративних мережах.

Об'єкт дослідження – процес виявлення бот-мереж у корпоративних мережах на основі аналізу DNS-трафіка.

Предмет дослідження – моделі, методи та програмні засоби інформаційної технології виявлення бот-мереж в корпоративних мережах на основі аналізу DNS-трафіка.

Методи дослідження. Для розв'язання поставлених задач використано основні положення системного аналізу, методів аналізу даних, теорії мір близькості та теорії комп'ютерних мереж.

Наукова новизна одержаних результатів.

1. Набула подальшого розвитку модель процесу виявлення бот-мереж в корпоративних мережах на основі аналізу DNS-трафіка. Розроблена модель відрізняється від відомих моделей тим, що задіює розроблені модель DNS-трафіка та модель бот-мереж з врахуванням системи доменних імен, що дозволило використання цієї моделі для виявлення вже відомих та нових бот-мереж.

2. Вперше розроблено метод ідентифікації бот-мереж у корпоративних мережах на основі їх групової активності в DNS-трафіку, який, на відміну від відомих, уможливлює уточнений поділ періоду моніторингу на інтервали, в межах яких здійснюється пошук груп інфікованих комп'ютерних систем, що ґрунтується на основі аналізу значень TTL, які містяться в DNS-повідомленнях, використовує нову ознаку синхронності DNS-запитів, а також враховує особливості поведінки груп інфікованих комп'ютерних систем, характерні для багатьох видів бот-мереж, що дозволило підвищити достовірність виявлення бот-мереж в порівнянні з відомими антивірусними програмними засобами.

3. Вперше розроблено метод виявлення бот-мереж, які застосовують технології ухилення від виявлення на основі DNS, у корпоративних мережах, який грунтується на залученні кластерного аналізу множини ознак, одержаних з корисного навантаження DNS-повідомлень, які вказують на використання таких технологій ухилення, що дозволило підвищити достовірність виявлення бот-мереж в порівнянні з відомими антивірусними програмними засобами.

4. Набула подальшого розвитку інформаційна технологія виявлення бот-мереж в корпоративних мережах на основі аналізу DNS-трафіка, яка дозволяє ідентифікувати боти, що здійснюють групову активність в DNS-трафіку, а також виявляти боти бот-мереж, які застосовують технології ухилення від виявлення на основі DNS. Застосування розробленої інформаційної технології дозволяє підвищити достовірність процесу виявлення бот-мереж в порівнянні з відомими інформаційними технологіями та виявляти нові бот-мережі.

Практичне значення одержаних результатів полягає в розробленні програмного забезпечення (ПЗ) інформаційної технології (ІТ) виявлення бот-мереж в корпоративних мережах на основі аналізу DNS-трафіка. Результати експериментальних досліджень з використанням розробленого ПЗ підтверджують вірність наукових положень запропонованої ІТ, оскільки впровадження інформаційної технології підвищує достовірність діагностування на 8-22% у порівнянні з відомими програмними засобами виявлення бот-мереж.

Теоретичні та практичні результати роботи впроваджено:

– державне підприємство «Новатор», відділ автоматизованих систем управління;

– товариство з обмеженою відповідальністю «ITT - telecommunication company»;

– у навчальному процесі Хмельницького національного університету при викладанні дисциплін «Програмування комп'ютерних мереж», «Технічна діагностика і надійність комп'ютерних пристроїв та систем» та «Інженерія програмного забезпечення».

Особистий внесок здобувача. Всі основні результати дослідження, які виносяться на захист, отримані автором особисто. В роботах, опублікованих у співавторстві, автору належить: проведено аналіз відомих методів виявлення бот-мереж на основі DNS [11-13]; розроблено моделі бот-мережі, DNS-трафіка та модель процесу виявлення бот-мереж в мережах на основі аналізу ознак, які можуть бути вилучені з DNS-трафіка [14]; розроблено новий метод ідентифікації бот-мереж на основі аналізу DNS-трафіка, який грунтується на властивості групової активності ботів в DNS-трафіку та враховує анормальну поведінку груп КС, властиву для бот-мереж [3, 8]; розроблено новий метод виявлення бот-мереж в корпоративних мережах на основі пасивного моніторингу DNS-трафіка та активного DNS-зондування з використанням кластерного аналізу векторів ознак, вилучених з корисного навантаження DNS-повідомлень, що дозволило здійснювати виявлення бот-мереж, які використовують технології ухилення від виявлення на основі DNS [4, 5, 9, 10, 15, 16]; розроблено ІТ виявлення бот-мереж на основі аналізу DNS-трафіка, яка побудована на базі двох нових методів: методу ідентифікації бот-мереж на основі їх групової активності в DNS-трафіку та методу виявлення бот-мереж, які застосовують технології ухилення від виявлення на основі DNS [6]; розроблено концептуальну схему локалізації бот-мереж в частині застосування інформації з різних КС мережі для аналізу трафіку з метою виявлення аномальної поведінки КС, властивої для бот-мереж [7].

Апробація результатів дисертації. Основні положення та результати проведених у дисертаційній роботі досліджень доповідалися та обговорювалися на міжнародних та всеукраїнських конференціях, а саме: Міжнародна конференція «Контроль і управління в складних системах» (м. Вінниця, 2014 р.); Міжнародна науково-практична конференція молодих вчених та студентів «Інформаційне, програмне та технічне забезпечення систем управління організаційно-технологічними комплексами» (м. Луцьк, 2015 р.); Computer Networks International Conference (Брунов, Польща, 2015, 2016 рр.); International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (Варшава, Польща, 2015 р.); Міжнародна конференція з автоматичного управління та інформаційних технологій (м. Київ, 2015 р.); Міжнародний науково-практичний семінар молодих вчених та студентів «Програмовані логічні інтегральні схеми та мікропроцесорна техніка в освіті і виробництві» (м. Луцьк, 2016 р.); Міжнародна наукова конференція ім. Т. А. Таран «Інтелектуальний аналіз інформації» (м. Київ, 2016 р.); Cyber Forum DESSERT B2S-S2B (м. Чернівці, 2016 р.); науково-практичні конференції професорсько-викладацького складу Хмельницького національного університету 2014-2016 рр.

Публікації. Основні матеріали дисертації викладено в 16 наукових публікаціях, з них 6 статей, опублікованих у наукових фахових виданнях України (2 з яких є одноосібними), 3 – у виданнях, зареєстрованих в наукометричній базі Index Copernicus, і 2 – у періодичних закордонних виданнях, зареєстрованих у наукометричній базі Scopus, 1 з яких – у виданні, зареєстрованому в наукометричній базі Web of Science, 8 – у матеріалах конференцій, з них 1 – індексоване у наукометричних базах Scopus та Web of Science, та 1 патент на корисну модель.

Обсяг і структура дисертації. Дисертація складається зі вступу, чотирьох розділів та висновків, викладених на 150 сторінках основного тексту, списку використаних джерел (138 найменувань). Робота містить 47 рисунків, 7 таблиць та 3 додатки.

ОСНОВНИЙ ЗМІСТ РОБОТИ

У вступі обґрунтовано актуальність теми дисертації, визначено об’єкт та предмет, мету і задачі дослідження, визначено наукову новизну та практичну цінність одержаних результатів. Вказано зв’язок роботи з науковими програмами та НДР за місцем виконання роботи. Надано інформацію щодо кількості публікацій та апробації результатів дисертації.

У першому розділі досліджено принципи функціонування бот-мереж з врахуванням системи доменних імен (DNS) та методи, на яких базуються ІТ виявлення бот-мереж на основі DNS.

Зважаючи на надані ресурсами Virus Bulletin і AV-TEST результати рейтингів тестування відомих програмних засобів виявлення бот-мереж в КМ, які визначають ефективність цих засобів на рівні 76–92%, а також беручи до уваги результати проведеного порівняльного аналізу цих програмних засобів, які підтвердили їх неспроможність здійснювати виявлення бот-мереж в мережах з високою достовірністю, виникає необхідність створення нової ІТ виявлення бот-мереж в КМ. Результати проведених досліджень показали, що з метою підвищення рівня ефективності та достовірності виявлення бот-мереж в корпоративних мережах

перспективним підходом є виявлення бот-мереж в КМ на основі аналізу DNS-трафіка.

У другому розділі визначено область дослідження та розроблено модель бот-мереж з врахуванням системи доменних імен, модель DNS-трафіка та модель процесу виявлення бот-мереж в корпоративних мережах на основі аналізу DNS-трафіка.

З врахуванням використання DNS на різних фазах життєвого циклу бот-мережі розроблено модель бот-мережі (рис. 1):

$$M_{BN} = \langle C, A, B, \Psi, Z, L, F \rangle, \quad (1)$$

де $C = \{c_j\}_{j=1}^{N_C}$ – множина контролюючих елементів бот-мережі, N_C – кількість контролюючих елементів бот-мережі, $C = \{c_j\}_{j=1}^{N_C} = \{\langle\langle D, I\rangle, \langle N, E\rangle\rangle_j\}_{j=1}^{N_C}$, де $D = \{d_j\}_{j=1}^{N_D}$, $I = \{i_j\}_{j=1}^{N_I}$ – множини доменних імен та IP-адрес контролюючих елементів бот-мережі відповідно, $N = \{n_j\}_{j=1}^{N_N}$, $E = \{e_j\}_{j=1}^{N_E}$ –множини доменних імен та IP-адрес авторитетних серверів імен для d відповідно, N_D, N_I, N_N, N_E – кількість доменних імен, які відповідають контролюючим елементам бот-мережі, IP-адрес, які співставляються з цими доменними іменами, доменних імен авторитетних серверів імен та їх IP-адрес відповідно; $A = \{a_j\}_{j=1}^{3}$ – тип архітектури бот-мережі, a_1 – централізована, a_2 – розподілена, a_3 – гібридна; $B = \{b_j^p\}_{j=1}^{N_B}$ – множина мережних протоколів, що використовуються для керування бот-мережею, N_B – кількість мережних протоколів, $p \in P$, $P = \{1..65535\}$ – множина портів, що використовуються для керування бот-мережею; $\Psi = \{\psi_j\}_{j=1}^{4}$ – множина технологій ухилення від виявлення бот-мереж на основі DNS, ψ_1 – періодична зміна IP-відображення (cycling of IP mapping), ψ_2 – «потік доменів» (domain flux), ψ_3 – «швидкозмінні мережі» (fast flux), ψ_4 – DNS-тунелювання (DNS-tunneling); $Z = \{z_j\}_{j=1}^{N_Z}$ – множина ботів, що входять до складу бот-мережі, N_Z – кількість ботів бот-мережі; $L = \{l_j\}_{j=1}^{5}$ – множина стадій життєвого циклу бот-мережі; $F = \{f_j\}_{j=1}^{N_F}$ – множина функцій ботів, що визначається відповідною фазою життєвого циклу бот-мережі, N_F –

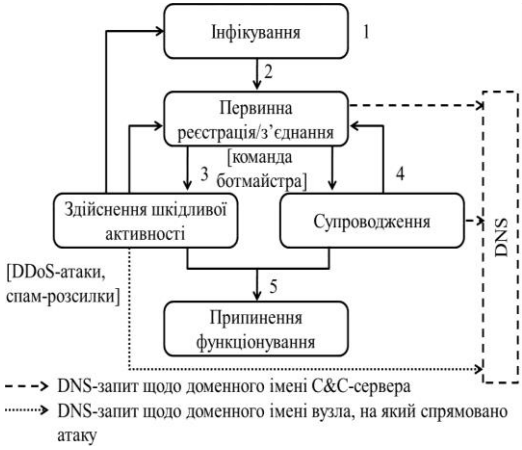


Рисунок 1 – Схема життєвого циклу бот-мережі з врахуванням системи доменних імен

кількість функцій ботів бот-мережі; функція інфікування вузла $l_1 \Rightarrow \mathrm{Y} \xrightarrow{f_1} \{h_{\inf} | h_{\inf} \in H\}$, де Y – множина шкідливих дій, закладених в функціонал бот-мережі, H – множина КС в глобальній мережі, $h_{\inf}$ – інфікована ботом КС; функція приєднання інфікованої КС до бот-мережі $l_2 \Rightarrow Z \cup \{h_{\inf} | h_{\inf} \in H\} \xrightarrow{f_2} Z'$; функція оновлення версії ШПЗ бота бот-мережі $l_3 \Rightarrow z \times z' \xrightarrow{f_3} z'$; функція виконання команди на здійснення шкідливої активності $l_4 \Rightarrow Z \times \{p | p \in P\} \xrightarrow{f_4} \mathrm{Y}$, де P – множина команд, які можуть бути виконані ботами бот-мережі; функція припинення функціонування бота бот-мережі $l_5 \Rightarrow Z \setminus \{z | z \in Z\} \xrightarrow{f_5} Z'$.

Побудована модель бот-мереж враховує застосування бот-мережами технологій ухилення від виявлення на основі DNS. При періодичній зміні IP-відображення буде $d \to \{i_1,...,i_n\}$, $\psi_1 \Rightarrow a_1$. Для технології ухилення «потік доменів» має місце відповідність $\{i\} \to \{d_1,...,d_n\}$, або $\{i_1,...,i_n\} \to \{d_1,...,d_m\}$ в межах TTL-періоду для записів DNS щодо доменного імені C&C-сервера, $\psi_2 \Rightarrow a_1$. Для «однопоточної» «швидкозмінної» мережі буде $\{c_1,...,c_n\} \coloneqq \{\mathrm{x} \mid \mathrm{x} \in Z \wedge x \in C\}$, $d \to \{i_1,...,i_n\}$ в межах інтервалу часу, визначеного TTL-періодом DNS. Для «двопоточної» «швидкозмінної» мережі для кожного авторитетного сервера імен n має місце відповідність $d \to \{i_1,...,i_n\}$, $\{n_1,...,n_m\} \to \{e_1,...,e_n\}$, $\{n_1,...,n_m\} \coloneqq \{\mathrm{x} \mid \mathrm{x} \in Z \wedge x \in N\}$, $\psi_3 \Rightarrow a_2$. При використанні технології DNS-тунелювання множина доменних імен D зони DNS фактично виступає в ролі аналога доменних імен C&C-сервера бот-мережі, IP-адреса e DNS-сервера зловмисника зазвичай залишається сталою, $\{d_1,...,d_n\} \to e$, $\psi_4 \Rightarrow a_1 \vee a_3$.

Розроблена модель DNS-трафіка корпоративної мережі:

$$M_T = \langle \chi, H, S, D \rangle, \tag{2}$$

де χ – множина DNS-повідомлень, надісланих від та до множини H КС корпоративної мережі, $\chi = \chi^O \cup \chi^I$, де χ^O – множина вихідних DNS-повідомлень мережі, χ^I – множина вхідних DNS-повідомлень мережі; S – множина DNS-серверів, до яких було надіслано DNS-запити та від яких було одержано DNS-відповіді КС мережі, $S = S^L \cup S^N$, де S^L – множина локальних DNS-серверів корпоративної мережі, S^N – множина нелокальних DNS-серверів; D – множина запитаних КС мережі доменних імен, $D = \{d_i\}_{i=1}^{N_D}$, де N_D – кількість різних доменних імен.

Множина захоплених вхідних DNS-повідомлень може бути представлена як $\chi^T = \bigcup_{j=d_1}^{d_{N_D}} \bigcup_{k=1}^{N_{TTL}} \chi_{j,k}$, де χ_j – підмножини вхідних DNS-повідомлень щодо певного доменного імені, захоплені протягом часу спостереження; $\chi_{j,k}$ – підмножини вхідних DNS-повідомлень щодо певного доменного імені, захоплені в межах певного TTL-періоду; $\chi_{j,k} = \{\chi_{j,k,i}\}_{i=1}^{N_{\chi,j,k}}$, де $\chi_{j,k,i}$ – DNS-повідомлення, захоплене в

межах певного TTL-періоду, $N_{\chi,j,k}$ – кількість DNS-повідомлень, захоплених в межах певного TTL-періоду.

На основі моделі бот-мереж з врахуванням системи доменних імен розроблено модель процесу виявлення бот-мереж на основі аналізу DNS-трафіка, формалізована схема та часова діаграма якої подані на рис. 2:

$$M_D = \left\langle \chi^T, f_1^T, C_{GA}{}^T, C_{ET}{}^T, f_2{}^T, T \right\rangle, \tag{3}$$

де χ^T – множина захоплених вхідних DNS-повідомлень до множини H КС корпоративної мережі; $f_1{}^T$ – функція співставлення доменних імен з «білим» та «чорним» списками; $C_{GA}{}^T$ – множина алгоритмів ідентифікації бот-мереж на основі їх групової активності в DNS-трафіку; $C_{ET}{}^T$ – множина алгоритмів виявлення бот-мереж, які застосовують технології ухилення від виявлення на основі DNS; $f_2{}^T$ – функція локалізації КС, інфікованих ботами, та блокування дій ботів; $T = \{t_m\}_{m=0}^{N_T}$ – інтервал часу

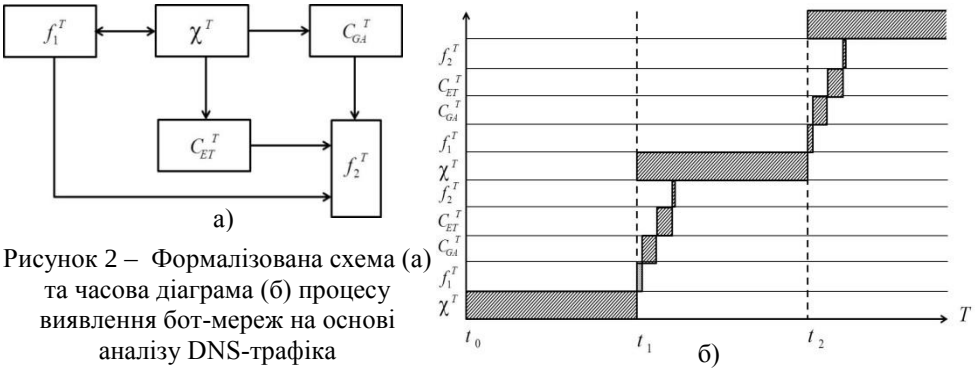


Рисунок 2 – Формалізована схема (а) та часова діаграма (б) процесу виявлення бот-мереж на основі аналізу DNS-трафіка

спостереження, де N_T – кількість ітерацій спостереження.

Таким чином, розроблено модель бот-мереж з врахуванням системи доменних імен, модель DNS-трафіка та модель процесу виявлення бот-мереж в корпоративних мережах на основі аналізу DNS-трафіка, які є основою методів виявлення бот-мереж на основі аналізу DNS-трафіка.

У третьому розділі автором розроблено метод ідентифікації бот-мереж у корпоративних мережах на основі їх групової активності в DNS-трафіку та метод виявлення бот-мереж, які використовують технології ухилення від виявлення на основі DNS, а також схему процесу виявлення бот-мереж в корпоративних мережах на основі аналізу DNS-трафіка (рис. 3).

Розроблений метод ідентифікації бот-мереж заснований на властивості групової активності бот-мереж в DNS-трафіку, яка проявляється в зосереджених в невеликому проміжку часу групових DNS-запитах КС під час спроб доступу до C&C-серверів бот-мереж, їх міграціях, виконанні команд або скачуванні оновлень ШПЗ. Метод

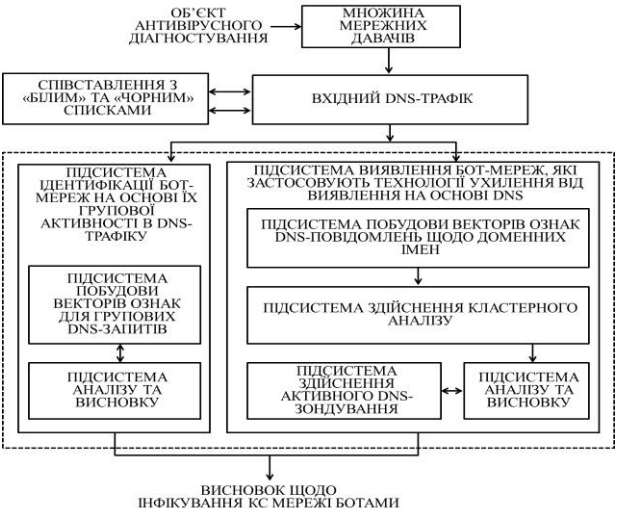


Рисунок 3 – Укрупнена схема процесу виявлення бот-мереж в корпоративних мережах на основі аналізу DNS-трафіка

складається з наступних кроків (рис. 4):

1. Збір вхідного DNS-трафіка за допомогою множини мережних давачів, підключених до дзеркалюючих портів комутаторів (рис. 5).

2. Співставлення з «білим» та «чорним» списками доменних імен з метою відкидання легітимних DNS-запитів та виявлення DNS-запитів щодо відомих шкідливих доменних імен.

3. Виявлення груп КС, які ігнорують TTL-період. Поділ КС на групи здійснюється в межах TTL-періодів, отриманих КС у вхідних DNS-повідомленнях щодо кожного доменного імені, та з врахуванням можливості ігнорування TTL. З цією метою будується матриця спостереження V_{MAC} (рис. 6), кожен рядок якої містить MAC-адреси КС h_j, які здійснювали запити щодо конкретного доменного імені в межах TTL-періоду. Якщо КС повторно надсилали запити щодо доменного імені d в межах TTL, то для таких запитів в матриці спостереження будується новий рядок. Позначимо N_G та $N_{G_{rep}}$ розміри груп для попереднього та повторного групових запитів, δ – порогове значення подібності між двома групами, обчислене експериментально. Якщо $\delta \cdot N_G > N_{G_{rep}}$, то рядок

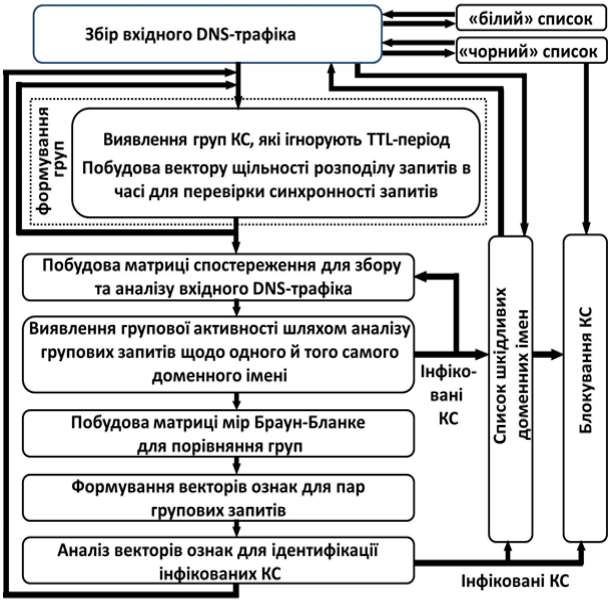


Рисунок 4 – Принцип функціонування методу ідентифікації бот-мереж на основі їх групової активності в DNS-трафіку

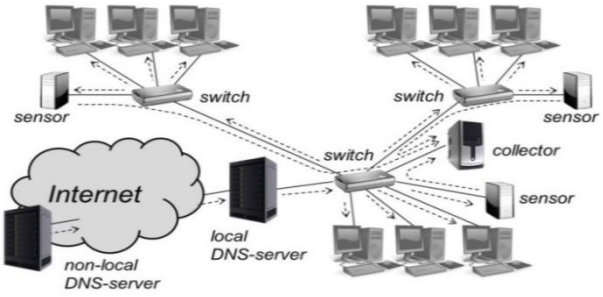


Рисунок 5 – Узагальнена схема корпоративної мережі як об'єкту діагностування на наявність бот-мереж

матриці V_{MAC} для повторного запиту відкидається.

Для групових запитів, які не були відкинуті на цьому етапі, перевіряється їх синхронність, як описано нижче.

4. Побудова вектору щільності розподілу запитів в часі для перевірки синхронності запитів. Групи запитів можна розглядати як синхронні, якщо спостерігається велика кількість запитів для доменного імені в межах часу, коли боти бот-мережі здійснюють запити – часу синхронізації ботів ts, обчисленого експериментально. Якщо інтервал часу між першим та останнім DNS-відгуками Δt для групового запиту щодо доменного імені d більший, ніж тривалість часового вікна ts, то інтервал часу Δt розбивається на z підінтервалів: $z = \left(t_{last} - t_{first}\right) / \left(\frac{1}{3} ts\right)$, де t_{last} та t_{first} – час надходження останнього та першого DNS-відгуків щодо доменного імені d в межах TTL-періоду, протягом якого здійснюється пошук групової активності або зафіксовано групове очищення локальних кешів DNS (рис. 7). Для групового запиту будується вектор щільності розподілу запитів в часі $\overline{W_d} = \left(\Omega_j\right)_{j=1}^{z}$, де Ω_j – кількість запитів в межах z-го інтервалу. Для елемента вектору $\overline{W_d}$ з максимальним значенням $\Omega_{\max}$ в межах $j = \max \pm 2$ відшукуються два суміжні елементи з найбільшими значеннями таким чином, щоб всі три елементи описували розподіл DNS-запитів неперервного інтервалу часу (рис. 7, б), та обчислюється їх сума $\left(Sum_s\right)$. Якщо $(1-\delta)\left(Sum_s + Sum_r\right) \geq Sum_r$, то множини MAC-адрес груп КС в матриці V_{MAC} об'єднуються, і груповий запит підлягає подальшому аналізу, інакше група відкидається, де Sum_r – сума значень решти елементів вектору $\overline{W_d}$.

h_1	h_2	h_3	h_4	...	h_j
1	1	1	1	...	1
1	0	1	1	...	0

Рисунок 6 – Матриця спостереження V_{MAC}

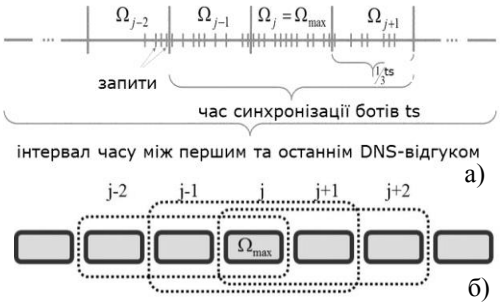


Рисунок 7 – Перевірка синхронності DNS-запитів: поділ інтервалу Δt на z підінтервалів (а) та пошук найбільшої кількості DNS-запитів в межах інтервалу часу ts (б)

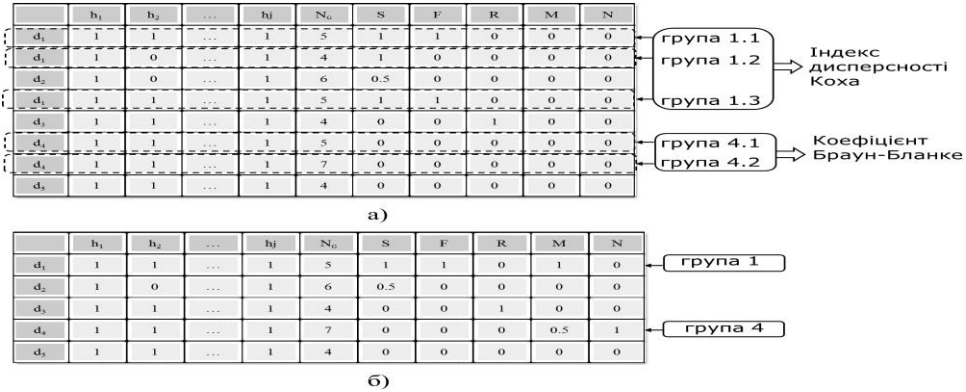
5. Побудова матриці спостереження для збору та аналізу вхідного DNS-трафіка. Для кожного визначеного інтервалу часу моніторингу t_m будується матриця спостереження M_m, де m – номер ітерації спостереження. Вона містить доменні імена d_i, запитані групами КС; MAC-адреси груп КС h_j, отримані з матриці V_{MAC}; а також ознаки наявності у груп КС особливостей поведінки в DNS-трафіку, притаманних для бот-мереж, а саме: ознаку звертання до локальних / нелокальних DNS-серверів – S; ознаку повторного запиту в межах TTL-періоду – F; ознаку наявності у DNS-відповідях коду помилки NXDOMAIN – R; ознаку "інфікований" чи "підозрілий" щодо групи КС, отриману на проміжних етапах аналізу – M; номер ітерації спостереження, на якій зафіксовано ознаку "підозрілий" – N; кількість КС у групі – N_G. Якщо було виявлено синхронність запитів, то множини MAC-адрес груп КС переносяться з матриці V_{MAC} до матриці спостереження M_m. Якщо було виявлено групове очищення локальних кешів DNS, то у комірці матриці спостереження $M_m(d_i, F)$ позначається «1», інакше – «0». Якщо група КС надсилала запити щодо доменного імені d_i як до локального, так і до інших DNS-серверів, то у комірці матриці спостереження $M_m(d_i, S)$ позначається «0», якщо лише до локального DNS-сервера – «0.5», якщо лише до нелокальних DNS-серверів – «1». Якщо DNS-відгуки для групи містили код помилки NXDOMAIN, то у комірці матриці спостереження $M_m(d_i, R)$ позначається «1», інакше «0».

6. Виявлення групової активності шляхом аналізу групових запитів щодо одного й того самого доменного імені. Визначення приналежності до бот-мереж груп КС, що запитували однакові доменні імена, здійснюється на основі аналізу подібності цих груп КС за MAC-адресами та аналізу наявності у них особливостей поведінки в DNS-трафіку, притаманних для бот-мереж. В залежності від кількості групових запитів щодо певного доменного імені d обирається коефіцієнт Браун-Бланке K_B для порівняння двох груп або індекс дисперсності Коха K_K для 3 і більше груп (рис. 8, а).

Групи КС вважаються інфікованими, якщо коефіцієнт подібності для груп перевищує порогове значення $K_B \geq \delta$ або $K_K \geq \delta$, де δ – порогове значення подібності. Також, введено порогове значення подібності δ`, яке вказує на підозрілість груп КС, якщо $\delta` \leq K_B < \delta$ або $\delta` \leq K_K < \delta$. В якості ідентифікаторів КС в мережі використовуються MAC-адреси за умови забезпечення запобігання їх підміни.

Якщо результат порівняння становить $\delta` \leq K_B < \delta$ або $\delta` \leq K_K < \delta$, то здійснюється додатковий аналіз матриці спостереження M_m щодо наявності факту ігнорування TTL-періоду групами та використання групами нелокальних DNS-серверів за визначеними правилами. Якщо групи, які запитували одне й те саме доменне ім'я, визначені інфікованими або підозрілими, то з метою подальшого пошуку пов'язаних з групою DNS-запитів множини їх MAC-адрес об'єднуються в один рядок для доменного імені d в матриці спостереження M_m (рис. 8, б), і комірки матриці спостереження M_m для об'єднаних рядків заповнюються за визначеними правилами. Якщо жодна з умов не задовольняється, то групові запити для таких доменних імен видаляються з матриці

спостереження M_m. Якщо група КС була визначена інфікованою, доменне ім'я d заноситься до списку шкідливих доменних імен.



а)

б)

Рисунок 8 – Матриця спостереження M_m: порівняння групових запитів для одного й того самого доменного імені d (а) та об'єднання множин MAC-адрес в один рядок для доменного імені d (б)

7. Побудова нижньотрикутної матриці мір Браун-Бланке для порівняння груп. На основі матриці спостереження M_m будується нижньотрикутна матриця мір Браун-Бланке B_m. Ознаки N_G, S, F, R, M, N з матриці M_m переносяться до матриці B_m. Рядки матриці B_m формуються за зростанням кількості MAC-адрес в групах N_G, по стовпцях. Також, в матрицю B_m заносяться коефіцієнти Браун-Бланке, обчислені для пар груп КС (рис. 9). Обчислення значень комірок для кожного стовпця припиняється, якщо відношення розмірів порівнюваних груп є меншим за порогове значення δ'.

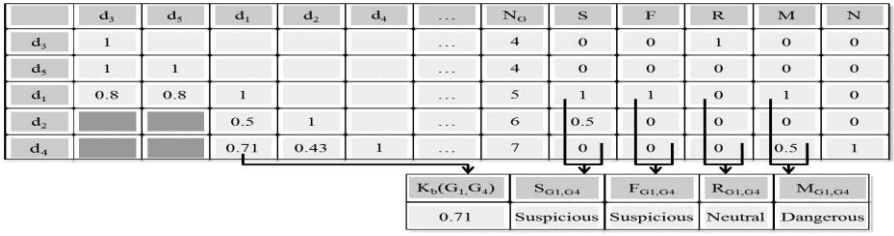


Рисунок 9 – Побудова матриці мір Браун-Бланке B_m та формування векторів ознак для пар групових запитів

8. Формування векторів ознак для пар групових запитів. Для кожної пари групових запитів, якщо виконується умова $K_B \geq \delta`$, згідно матриці мір Браун-Бланке B_m формується вектор ознак $\overline{W_{G_1,G_2}}$ (рис. 9), який складається з п'яти елементів: коефіцієнт Браун-Бланке та зведені поведінкові ознаки для двох

порівнюваних груп, отримані на основі матриці B_m, які можуть приймати наступні значення: «Unusual» (непритаманна ботам), «Neutral» (властива як користувачам, так і ботам), «Suspicious» (підозріла), «Dangerous» (властива ботам):

$$\overline{W_{G_1,G_2}} = \left(K_B(G_1,G_2), S_{G_1,G_2}, F_{G_1,G_2}, R_{G_1,G_2}, M_{G_1,G_2}\right), \tag{4}$$

де $S_{G_1,G_2}, F_{G_1,G_2}, R_{G_1,G_2}, M_{G_1,G_2}$ – зведені поведінкові ознаки для двох порівнюваних груп.

Зведені поведінкові ознаки S_{G_1,G_2} та M_{G_1,G_2} можуть бути визначені наступним чином:

$$S_{G_1,G_2} = \begin{cases} Unusual, if\ B_m(d_1,S) = B_m(d_2,S) = 0, \\ Neutral, if\ B_m(d_1,S) = B_m(d_2,S) = 0.5, \\ Dangerous, if\ B_m(d_1,S) = B_m(d_2,S) = 1, \\ Suspicious\ otherwise. \end{cases} \tag{5}$$

$$M_{G_1,G_2} = \begin{cases} Neutral, if\ B_m(d_1,M) = B_m(d_2,M) = 0, \\ Suspicious, if\ ((B_m(d_1,M) = 0.5 \vee B_m(d_2,M) = 0.5) \wedge \\ \wedge B_m(d_1,M) \neq 1 \wedge B_m(d_2,M) \neq 1) \wedge B_m(d_1,M) \neq B_m(d_2,M), \\ Dangerous, if\ B_m(d_1,M) = 1 \vee B_m(d_2,M) = 1 \vee \\ \vee (B_m(d_1,M) = B_m(d_2,M) = 0.5 \wedge B_m(d_1,N) \neq B_m(d_2,N) \vee \\ \vee B_m(d_1,N) = B_m(d_2,N) = 0). \end{cases} \tag{6}$$

Зведені ознаки F_{G_1,G_2} та R_{G_1,G_2} визначаються аналогічно. Нижче наведено приклад для F_{G_1,G_2}:

$$F_{G_1,G_2} = \begin{cases} Neutral, if\ B_m(d_1,F) = B_m(d_2,F) = 0, \\ Suspicious, if\ B_m(d_1,F) \neq B_m(d_2,F), \\ Dangerous, if\ B_m(d_1,F) = B_m(d_2,F) = 1. \end{cases} \tag{7}$$

9. Аналіз векторів ознак для виявлення інфікованих КС. Визначення приналежності до бот-мереж груп КС, що запитували різні доменні імена, здійснюється шляхом аналізу векторів ознак для пар групових DNS-запитів, для яких коефіцієнт Браун-Бланке перевищує прийняте порогове значення. Аналіз векторів ознак $\overline{W_{G_1,G_2}}$ здійснюється за наступними правилами, де функція виходу $f\left(\overline{W_{G_1,G_2}}\right)$ може приймати чотири значення: «Not_Infected» (неінфіковані), «Not_Suspicious» (не підозрілі), «Suspicious» (підозрілі), «Infected» (інфіковані):

$$f\left(\overline{W_{G_1,G_2}}\right) = \begin{cases} Not\_Infected, if\ K_B(G_1,G_2) < \delta \wedge S_{G_1,G_2} = Unusual \wedge \\ \wedge \forall \overline{W_{G_1,G_2}}(j) \neq Suspicious \wedge \forall \overline{W_{G_1,G_2}}(j) \neq Dangerous, \\ Not\_Suspicious, if\ K_B(G_1,G_2) < \delta \wedge S_{G_1,G_2} \neq Unusual \wedge \\ \wedge \forall \overline{W_{G_1,G_2}}(j) \neq Suspicious \wedge \forall \overline{W_{G_1,G_2}}(j) \neq Dangerous, \\ Infected, if\ \exists \overline{W_{G_1,G_2}}(j) = Dangerous \vee K_B(G_1,G_2) \geq \delta, \\ Suspicious\ otherwise. \end{cases} \tag{8}$$

де $j = \overline{2,5}$ – номер елемента в векторі ознак.

Щодо груп КС, визначених як інфіковані, здійснюються заходи з метою ліквідації інфекції (блокування, усунення вразливостей системи, встановлення (оновлення) антивірусного ПЗ тощо). Групи КС з матриці спостереження M_m, які не потрапили до матриці мір Браун-Бланке B_m, та групи, для яких не було виконано умову $K_B \geq \delta'$, а також групи, визначені як не підозрілі та підозрілі, аналізуються разом з даними, що будуть отримані на наступній ітерації спостереження (матриця спостереження M_{m+1}) з метою виявлення повторних групових запитів.

Розроблено новий метод виявлення бот-мереж, які застосовують технології ухилення від виявлення на основі DNS. Метод складається з наступних кроків (рис. 10):

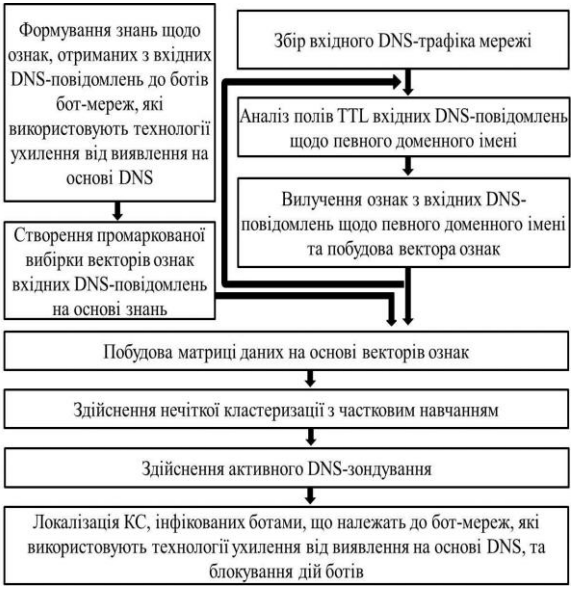


Рисунок 10 – Принцип функціонування методу виявлення бот-мереж, які застосовують технології ухилення від виявлення на основі DNS

1. Збір вхідного DNS-трафіку мережі за допомогою мережних давачів, підключених до дзеркалюючих портів комутаторів.

2. Аналіз полів TTL вхідних DNS-повідомлень щодо певного доменного імені. На основі значень полів TTL опрацьовуються такі вхідні DNS-повідомлення: (1) кожне перше зафіксоване DNS-повідомлення щодо певного доменного імені в межах TTL; (2) кожне DNS-повідомлення, отримане КС повторно в межах TTL-періоду, якщо джерелом повідомлення є нелокальний DNS-сервер, і TTL-період, зазначений в цьому повідомленні, відрізняється від залишку TTL, в межах якого було отримане це повідомлення.

3. Вилучення ознак з вхідних DNS-повідомлень щодо певного доменного імені та побудова вектору ознак. Подамо вектор ознак, отриманих з вхідних DNS-повідомлень щодо певного доменного імені d, наступним чином:

$$\overline{W_e} = \left(l_N, n_U, e_N, t_{\text{mod}}, t_{med}, t_{aver}, n_A, n_{IP}, s_{IP}, s_A, n_{UA}, s_{UA}, n_D, f_{UR}, e_R, l_P, f_S\right), \qquad (9)$$

де l_N – довжина доменного імені; n_U – кількість унікальних символів в доменному імені; e_N – ентропія доменного імені; t_{mod} – TTL-період, мода (мінімальне значення у випадку, якщо множина є мультимодальною); t_{med} – TTL-період,

медіана; t_{aver} – TTL-період, середнє арифметичне значення; n_A – кількість A-записів, що відповідають доменному імені, у вхідному DNS-повідомленні (ознака використовується, якщо $n_A > 1$); n_{IP} – кількість IP-адрес, пов'язаних з доменним ім'ям (ознака використовується, якщо $n_A = 1$); s_{IP} – середня дистанція між IP-адресами, пов'язаними з доменним ім'ям (ознака використовується, якщо $n_A = 1$); s_A – середня дистанція між IP-адресами в множині A-записів для доменного імені у вхідному DNS-повідомленні (ознака використовується, якщо $n_A > 1$); n_{UA} – кількість унікальних IP-адрес в множинах A-записів, що відповідають доменному імені, у вхідних DNS-повідомленнях (ознака використовується, якщо $n_A > 1$); s_{UA} – середня дистанція між унікальними IP-адресами в множинах A-записів, що відповідають доменному імені, у вхідних DNS-повідомленнях (ознака використовується, якщо $n_A > 1$); n_D – кількість доменних імен, які спільно використовують IP-адресу; f_{UR} – бінарна ознака використання рідковживаних типів записів DNS (KEY, NULL тощо), або таких, які зазвичай не використовуються клієнтами (наприклад, TXT, які найбільш часто використовуються для тунелювання); e_R – максимальне значення ентропії записів DNS, які містяться в DNS-повідомленнях (CNAME, TXT, NS, MX, KEY, NULL тощо); l_P – середній розмір DNS-повідомлень щодо доменного імені; f_S – бінарна ознака успішності DNS-запиту ($f_S = 0$, якщо DNS-запит невдалий, $f_S = 1$, якщо DNS-запит успішний). Також, метод використовує функцію залежності $f_{E_{Bn}}$ ентропії поля DNS-повідомлення від його довжини, де n – основа кодування.

4. Побудова матриці даних на основі векторів ознак. З векторів ознак вхідних DNS-повідомлень формується матриця даних $V = (v_{ij})_{i=1,j=1}^{N_D,N_S}$, $V(i,) = \overline{W_e}$, де N_D – загальна кількість різних доменних імен, запитаних КС мережі, N_S – загальна кількість ознак вхідних DNS-повідомлень, які вказують на використання технологій ухилення від виявлення на основі DNS.

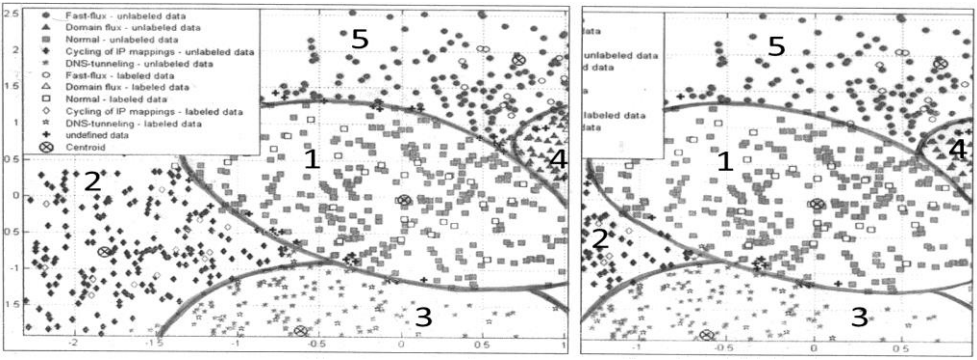
5. Здійснення нечіткої кластеризації з частковим навчанням з метою виявлення запитів, які можуть свідчити про функціонування ботів, що належать до бот-мереж, які використовують технології ухилення від виявлення на основі DNS. На основі ознак, властивих вхідним DNS-повідомленням до ботів, формуються знання, які можуть бути представлені у вигляді наступних правил:

$$
\begin{gathered}
if\ (t_{\mathrm{mod}} \in [0{,}900]\ and\ \ t_{med} \in [0{,}900]\ and\ t_{aver} \in [0{,}900]\ )\ and \\
and\ \ ((n_A \in (5,\infty)\ and\ \ s_A \in (65535,\infty))or(n_{UA} \in (8,\infty)\ and\ \ s_{UA} \in (65535,\infty))) \Rightarrow fast\_flux \\
if\ t_{\mathrm{mod}} \in [0{,}900]\ and\ \ t_{med} \in [0{,}900]\ and\ t_{aver} \in [0{,}900]\ and \\
and\ \ f_S = 0\ and\ n_D \in [8;\infty] \Rightarrow domain\_flux \\
if\ t_{\mathrm{mod}} \in [0{,}900]\ and\ \ t_{med} \in [0{,}900]\ and\ t_{aver} \in [0{,}900]\ and \\
and\ n_{IP} \in (5,\infty)\ and\ s_{IP} \in (65535,\infty) \Rightarrow cycling\ \ of\ IP\ mappings \\
if\ ((l_N \in [75{,}255]\ and\ \ n_U \in (27{,}37])\ or\ (e_N \geq f_{E_{B32}}\ or \\
or\ \ (e_R \geq f_{E_{B64}}\ or\ \ e_R \geq f_{E_{B256}})\ or\ \ f_{UR} = 1\ ))\ and\ l_P > 300 \Rightarrow DNS\_tunneling\,.
\end{gathered}
\tag{10}
$$

На основі знань (10) формується промаркована навчальна вибірка для часткового навчання кластеризатора з метою визначення початкових центрів кластерів. Кожен вектор ознак з промаркованої вибірки даних належить одному з множини наперед визначених кластерів $X = \{x_i\}_{i=1}^{5}$; приналежність вектора ознак кластеру x_1 свідчить про застосування технології ухилення «cycling of IP mapping», x_2 – «domain flux», x_3 – «fast flux», x_4 – «DNS-tunneling», x_5 – кластер, який містить нормальні DNS-запити.

Об'єктами кластеризації є вектори ознак $\overline{W_e}$. Застосування нечіткої кластеризації c-means дозволяє підвищити точність та інформативність результатів у випадках, якщо об'єкти кластеризації розташовані на межах кластерів. В якості відстані між об'єктом кластеризації та центром кластера застосовано норму Махаланобіса, оскільки вона надає можливість виокремлювати кластери у формі гіпереліпсоїдів з орієнтованими в довільних напрямках осями, що дозволяє врахувати можливу наявність у досліджуваних даних викидів. В якості способу нормування було обрано обчислення стандартизованого вкладу (Z-вкладу).

Результатом кластеризації є матриця нечіткого розбиття U, де кожен елемент матриці u_{ij} визначає ступінь приналежності і-го елемента множини об'єктів кластеризації до j-го кластера: $U = [u_{ij}], u_{ij} \in [0,1], i = \overline{1, N_D}, j = \overline{1,5}$, $\sum_{j=\overline{1,5}} u_{ij} = 1$. Приймемо λ та $\lambda`$ як порогові значення приналежності об'єкта кластеризації до кластера, за яких доменне ім'я вважається шкідливим або підозрілим відповідно. Якщо $u_{ij} \geq \lambda$, $j = \overline{1,4}$, то об'єкт відноситься до кластера, який відповідає одній з технологій ухилення. Належність вектору ознак до п'ятого кластера свідчить про виконання запитів до легітимних ресурсів. У випадку, якщо $\lambda` \leq u_{ij} < \lambda$, $j = \overline{1,4}$, то об'єкт може належати кільком кластерам, тому має місце певна невизначеність результатів (рис. 11).



а) б)

Рисунок 11 – Графічне представлення результатів кластеризації: невизначеність частини результатів (а) та усунення невизначеності частини результатів кластеризації (б). Цифрами позначено: 1 – кластер, що містить легітимні DNS-запити, 2 – «cycling of IP mapping», 3 – «DNS-tunneling», 4 – «domain flux», 5 – «fast flux»

6. Здійснення активного DNS-зондування. В ситуації, коли засоби пасивного моніторингу DNS-трафіка не можуть дати результату, з метою одержання додаткових ознак, які можуть вказувати на шкідливість доменного імені, доцільним є залучення засобів активного DNS-зондування шляхом здійснення запитів NS-записів, A-записів, SOA-записів, PTR-записів. З метою уточнення частини результатів кластеризації використовуються наступні ознаки, отримані засобами активного DNS-зондування (рис. 11): n_{NS} – кількість NS-записів у DNS-відповіді; s_{NS} – середня дистанція між IP-адресами для множини NS-записів щодо доменного імені; v_{retry} – значення поля retry, отримане у DNS-відповіді на SOA-запит; n_{ASN} – кількість різних номерів автономних систем (ASN), до яких належать IP-адреси, пов'язані з серверами імен; n_{ASA} – кількість різних номерів автономних систем, до яких належать IP-адреси, пов'язані з доменним іменем.

Висновок щодо наявності застосування технологій ухилення на основі DNS здійснюється із застосуванням знань, які можуть бути представлені у вигляді правил:

$$\begin{gathered} if\ t_{\mathrm{mod}} \in [0{,}900] and\ \ t_{med} \in [0{,}900] and\ t_{aver} \in [0{,}900] and \\ and\ \ n_{IP} \in (5,\infty) and\ s_{IP} \in (65535,\infty) and\ \ n_{ASA} > 2 \Rightarrow cycling\ \ of\ IP\ mappings \\ if\ (t_{\mathrm{mod}} \in [0{,}900] and\ \ t_{med} \in [0{,}900] and\ t_{aver} \in [0{,}900]\ ) and \\ and\ \ ((n_A \in (5,\infty)\ and\ \ s_A \in (65535,\infty)) or (n_{UA} \in (8,\infty)\ and\ \ s_{UA} \in (65535,\infty)) or\ n_{AS} > 2)\ and \\ and\ \ (s_{NS} \in (65535,\infty) or\ n_{ASN} > 2\ and\ n_{NS} > 3\ and\ v_{retry} \in [0{,}900])) \Rightarrow fast\_flux \\ if\ \ n_D \in [8;\infty] \Rightarrow domain\_flux \end{gathered} \tag{11}$$

7. Локалізація КС, інфікованих ботами, що належать до бот-мереж, які використовують технології ухилення від виявлення на основі DNS, та блокування дій ботів. На основі приналежності векторів $\overline{W_e}$ до кластерів здійснюється визначення доменних імен, до яких зверталися боти бот-мереж. З метою блокування дій ботів, які здійснювали шкідливі запити, локалізація КС мережі здійснюється за допомогою ведення файлів журналювання MAC-адрес КС, що здійснювали DNS-запити, та запитаних ними доменних імен.

Таким чином, розроблено метод ідентифікації бот-мереж у корпоративних мережах на основі їх групової активності в DNS-трафіку та метод виявлення бот-мереж, які застосовують технології ухилення від виявлення на основі DNS, що надало можливість виявляти як відомі, так і ще невідомі бот-мережі, а також здійснювати виявлення бот-мереж на початковій стадії поширення інфекції в мережі. Запропоновано методику визначення ефективності та достовірності інформаційної технології виявлення бот-мереж на основі аналізу DNS-трафіка.

У четвертому розділі з метою усунення недоліків відомих ІТ та підвищення достовірності виявлення бот-мереж в корпоративних мережах автором було розроблено інформаційну технологію виявлення бот-мереж на основі аналізу DNS-трафіка (рис. 12). Розроблено алгоритми ПЗ виявлення бот-мереж в корпоративних мережах на основі аналізу DNS-трафіка та проведено дослідження їх складності, що підтверджує можливість програмної реалізації розробленої ІТ.

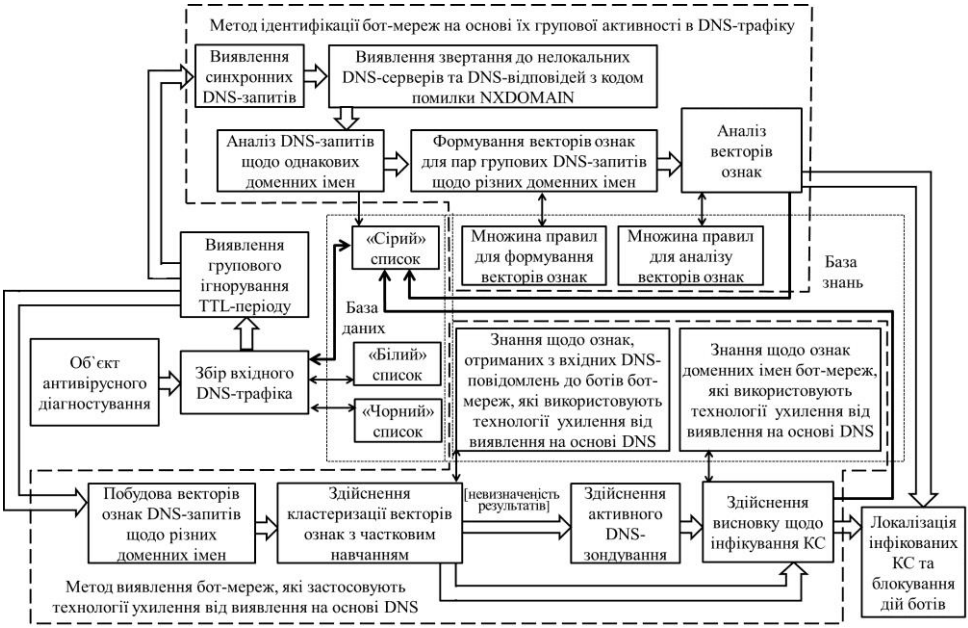


Рисунок 12 – Схема функціонування ІТ виявлення бот-мереж на основі аналізу DNS-трафіка

Розроблена ІТ виявлення бот-мереж на основі аналізу DNS-трафіка дозволяє здійснювати виявлення КС в корпоративній мережі, інфікованих як відомими, так і невідомими бот-мережами (табл. 1, рис. 13).

Таблиця 1

Достовірність виявлення бот-мереж в корпоративних мережах

№ зп	Засіб антивірусного діагностування	Середня достовірність виявлення, %	Середнє значення помилки 1-го роду, %	Середнє значення помилки 2-го роду, %	Середня тривалість часу, затраченого на виявлення, хв.
1.	Розроблений антивірусний засіб (АЗ)	96,22	3,78	3,44	30
2.	Avast Endpoint Protection Suite	84,80	15,20	11,66	25
3.	Avira Small Business Security Suite	86,38	13,62	9,52	42
4.	Dr.Web CureNet!	86,18	13,82	10,28	38
5.	ESET Endpoint Security	86,92	13,08	7,48	24
6.	Kaspersky Endpoint Security	88,86	11,14	8,30	31
7.	McAfee Endpoint Protection Suite	85,92	14,08	9,24	33
8.	Microsoft System Center Endpoint Protection	78,58	21,42	3,94	21
9.	Panda Endpoint Protection	83,56	16,44	5,84	29

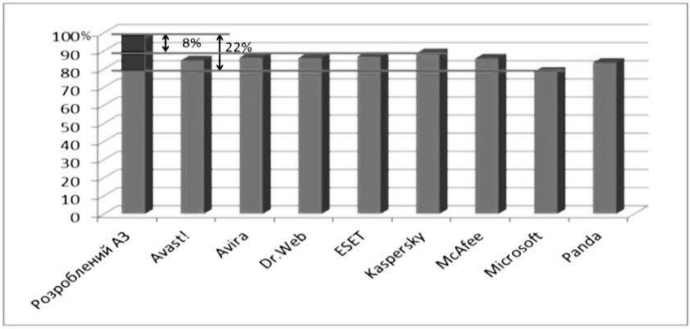


Рисунок 13 – Порівняльний аналіз розробленої ІТ з відомими (достовірність виявлення, %)

Для розрахунку достовірності виявлення бот-мереж шляхом застосування програмного забезпечення розробленої ІТ в порівнянні з відомими антивірусними засобами було проведено ряд експериментів (в загальній кількості 90, кожен тривалістю 24 години). З цією метою було створено спеціальне ПЗ, яке дозволяло здійснювати DNS-запити і мало властивості ботів бот-мереж з функціоналом, невідомим для антивірусних засобів. Для оцінки достовірності діагностування кожного антивірусного засобу мережа з 100 КС інфікувалась 60 створеними зразками вірусних програм.

Отже, результати дослідження показують, що використання розробленого програмного забезпечення ІТ виявлення бот-мереж на основі аналізу DNS-трафіка дозволяє підвищити рівень достовірності виявлення бот-мереж на 8-22% в порівнянні з відомими антивірусними програмними засобами та досягти зниження рівня помилок другого роду (хибних спрацювань) до 4%, що на 13-70% нижче в порівнянні з відомими антивірусними програмними засобами. Застосування ПЗ розробленої інформаційної технології виявлення бот-мереж на основі аналізу DNS-трафіка дозволяє підвищити оперативність роботи ІС корпоративної мережі та рівень безпеки корпоративної мережі, що підтверджується відповідними актами впровадження.

ВИСНОВКИ

В дисертаційній роботі вирішена актуальна науково-технічна задача виявлення бот-мереж в корпоративних мережах на основі аналізу DNS-трафіка. Обмеження областю корпоративних мереж та володіння апріорними знаннями стосовно системи доменних імен надало можливість підвищити достовірність виявлення нових та вже відомих бот-мереж. Враховуючи широке поширення КС в переважній більшості галузей, вирішення цієї задачі має важливе значення.

Основні наукові і практичні результати роботи полягають у наступному:

1.Проведено аналіз сучасних інформаційних технологій виявлення бот-мереж, який показав не високу достовірність виявлення ними невідомих нових ботів бот-мереж.

2. Досліджено особливості функціонування бот-мереж з врахуванням системи доменних імен та застосування технологій ухилення від виявлення бот-мереж на основі DNS.

3. Розроблено модель бот-мереж з врахуванням системи доменних імен, а також використання бот-мережами технологій ухилення від виявлення на основі DNS, що надає можливість підвищити достовірність виявлення бот-мереж.

4. Розроблено модель DNS-трафіка та модель процесу виявлення бот-мереж в мережах на основі аналізу DNS-трафіка, яка відрізняється від відомих моделей тим, що дозволяє здійснювати виявлення вже відомих та невідомих бот-мереж.

5. Розроблено та досліджено метод ідентифікації бот-мереж на основі їх групової активності в DNS-трафіку, який, на відміну від відомих, уможливлює уточнений поділ періоду моніторингу на інтервали, в межах яких здійснюється пошук груп інфікованих комп'ютерних систем, що ґрунтується на основі аналізу значень TTL, які містяться в DNS-повідомленнях, використовує нову ознаку синхронності DNS-запитів, а також враховує особливості поведінки груп інфікованих комп'ютерних систем, характерні для багатьох видів бот-мереж, що дозволило підвищити достовірність виявлення нових бот-мереж.

6. Розроблено та досліджено метод виявлення бот-мереж, які застосовують технології ухилення від виявлення на основі DNS, який ґрунтується на залученні кластерного аналізу множини ознак, одержаних з корисного навантаження DNS-повідомлень, які вказують на використання таких технологій ухилення, що дозволило підвищити достовірність виявлення нових бот-мереж.

7. Розроблені методи інтегровано в інформаційну технологію виявлення бот-мереж в корпоративних мережах на основі аналізу DNS-трафіка. Інформаційну технологію реалізовано у вигляді програмного забезпечення, яке дозволяє ідентифікувати боти, що здійснюють групову активність в DNS-трафіку, а також виявляти боти бот-мереж, які застосовують технології ухилення від виявлення на основі DNS. Застосування розробленої інформаційної технології дозволяє підвищити достовірність процесу виявлення бот-мереж в порівнянні з відомими інформаційними технологіями на 8-22% та виявляти як відомі, так і нові бот-мережі.

СПИСОК ОПУБЛІКОВАНИХ ПРАЦЬ ЗА ТЕМОЮ ДИСЕРТАЦІЇ

1. Бобровнікова К.Ю. Модель інформаційної технології виявлення бот-мереж на основі аналізу DNS-трафіка [Текст] / К.Ю. Бобровнікова // Вісник Хмельницького національного університету. – 2015. – № 6 (231). – С.164-172 (індексується в наукометричній базі Index Copernicus).

2. Бобровнікова К.Ю. Методи та програмне забезпечення інформаційної технології виявлення бот-мереж на основі аналізу DNS– трафіка [Текст] / К.Ю. Бобровнікова // Вісник Хмельницького національного університету. – 2016. – № 2. – С.53-57 (індексується в наукометричній базі Index Copernicus).

3. Савенко О.С. DNS-метод виявлення бот-мереж [Текст] / О. С. Савенко, С. М. Лисенко, К. Ю. Бобровнікова // Інформаційні технології та комп'ютерна інженерія, 2014. – № 3. – С. 39-45.

4. Савенко О.С. Метод виявлення бот-мереж, що використовують технології ухилення на основі DNS [Текст] / О. С. Савенко, С. М. Лисенко, К. Ю. Бобровнікова // Комп'ютерно-інтегровані технології: освіта, наука, виробництво. – 2015. – № 19. – С. 71-78.

5. Савенко О.С. Метод виявлення бот-мереж на основі пасивного моніторингу DNS-трафіка та активного DNS-зондування [Текст] / О. С. Савенко, С. М. Лисенко, К. Ю. Бобровнікова // Комп'ютерно-інтегровані технології: освіта, наука, виробництво. – 2016. – № 22. – С. 136-143.

6. Савенко О. С. Інформаційна технологія виявлення бот-мереж на основі аналізу DNS-трафіка [Текст] / О. С. Савенко, С. М. Лисенко, К. Ю. Бобровнікова // Радіоелектронні і комп'ютерні системи. – 2016. – № 5. – С. 38-42 (індексується в наукометричній базі Index Copernicus).

7. Мультиагентний спосіб локалізації бот-мереж у корпоративних комп'ютерних мережах [Текст] : пат. 108238 Україна : МПК G06F 21/55 (2013.01) / О.В. Поморова, О.С. Савенко, А.Ф. Крищук, С.М. Лисенко, К.Ю. Бобровнікова, А.О. Нічепорук; заявник та власник Хмельницький національний університет. – № u2016 00127 ; заявл. 04.01.16 ; опубл. 11.07.16, Бюл. № 13.

8. Pomorova O. A Technique for the Botnet Detection Based on DNS-Traffic Analysis [Text] / O. Pomorova, O. Savenko, S. Lysenko, A. Kryshchuk, K. Bobrovnikova // Communications in Computer and Information Science. – 2015. – Vol. 522. – pp. 127-138, ISSN: 1865-0929 (part scientometric Web of Science and SCOPUS).

9. Lysenko S. DNS-based Anti-evasion Technique for Botnets Detection [Text] / S. Lysenko, O. Pomorova, O. Savenko, A. Kryshchuk, K. Bobrovnikova // Proceedings of the 2015 IEEEE 8th International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS), IDAACS'2015, Warsaw, Poland, September 24-26, 2015, Vol.1. – PP. 453-458, ISBN: 978-1-4673-8359-2 (Print) (part scientometric Web of Science and SCOPUS).

10. Lysenko S. Anti-evasion Technique for the Botnets Detection Based on the Passive DNS Monitoring and Active DNS Probing / S. Lysenko, O. Pomorova, O. Savenko, A. Kryshchuk, K. Bobrovnikova [Text] // Communications in Computer and Information Science. – 2016. – Vol. 608. – PP. 83-95, ISSN: 1865-0929 (part scientometric SCOPUS).

11. Бобровнікова К.Ю. Методи виявлення бот-мереж, що використовують технології ухилення на основі DNS [Текст] / К.Ю.Бобровнікова, Д.В.Муравський, О.О.Павлова // Збірник наукових праць IV Всеукраїнської науково-практичної конференції молодих учених та студентів «Інтелектуальні технології в системному програмуванні», Хмельницький, 2015. – С. 25-29.

12. Бобровнікова К.Ю. Аналіз DNS-методів виявлення бот-мереж [Текст] / К.Ю.Бобровнікова, А.І.Наконечний, М.А.Репушанська // Збірник наукових праць IV Всеукраїнської науково-практичної конференції молодих учених та студентів «Інтелектуальні технології в системному програмуванні», Хмельницький, 2015. – С. 29-33.

13. Савенко О.С. Дослідження DNS-методів виявлення бот-мереж [Текст] / О. С. Савенко, С. М. Лисенко, К. Ю. Бобровнікова // Контроль і управління в

складних системах (КУСС-2014). XII Міжнародна конференція. Тези доповідей. Вінниця, 14-16 жовтня 2014 року. – Вінниця: ВНТУ. – 2014. – С. 87.

14. Савенко О.С. Модель процесу виявлення бот-мереж на основі аналізу DNS-трафіка [Текст] / О. С. Савенко, К. Ю. Бобровнікова // Матеріали 3-ї Міжнародної конференції з автоматичного управління та інформаційних технологій, ICACIT-2015 . – К.: КПІ. – 2015. – С. 60-67.

15. Бобровнікова К.Ю. Метод виявлення бот-мереж на основі пасивного моніторингу DNS-трафіка та активного DNS-зондування [Текст] / К. Ю. Бобровнікова, О. С. Савенко, С. М. Лисенко // Збірник тез доповідей міжнародного науково-практичного семінару молодих вчених та студентів «Програмовані логічні інтегральні схеми та мікропроцесорна техніка в освіті і виробництві». – м. Луцьк, ЛНТУ. – 2016. – С. 20-23.

16. Савенко О.С. Применение кластерного анализа для решения задачи обнаружения бот-сетей на основе анализа DNS-трафика [Текст] / О.С. Савенко, С. Н. Лысенко, К.Ю. Бобровникова // Сборник трудов XVI Международной научной конференции «Интеллектуальный анализ информации (ИАИ-2016)» им. Т.А.Таран. Сборник трудов. – К.: Просвіта. – 2016. – С.186-192.

АНОТАЦІЯ

Бобровнікова К. Ю. Інформаційна технологія виявлення бот-мереж у корпоративних мережах на основі аналізу DNS-трафіка. – *Рукопис.*

Дисертація на здобуття наукового ступеня кандидата технічних наук за спеціальністю 05.13.06 – Інформаційні технології. – Тернопільський національний технічний університет імені Івана Пулюя, Тернопіль, 2016.

Дисертація присвячена вирішенню актуальної наукової задачі – створенню інформаційної технології з метою підвищення достовірності та ефективності виявлення бот-мереж в корпоративних мережах на основі аналізу DNS-трафіка.

Проведено огляд та виявлено недоліки роботи інформаційних технологій і програмних засобів виявлення бот-мереж в мережах. Розроблено модель бот-мереж з врахуванням DNS, модель DNS-трафіка та модель процесу виявлення бот-мереж в корпоративних мережах на основі аналізу DNS-трафіка. На базі розроблених моделей побудовано інформаційну технологію виявлення бот-мереж у корпоративних мережах на основі аналізу DNS-трафіка, яка ґрунтується на двох розроблених нових методах виявлення бот-мереж: методу ідентифікації бот-мереж на основі їх групової активності в DNS-трафіку та методу виявлення бот-мереж, які застосовують технології ухилення від виявлення на основі DNS.

Здійснено програмну реалізацію інформаційної технології виявлення бот-мереж у корпоративних мережах на основі аналізу DNS-трафіка, що дало змогу виявляти відомі та невідомі боти бот-мереж з високою достовірністю.

Ключові слова: бот, бот-мережа, DNS-трафік, групова активність, технології ухилення від виявлення бот-мереж, виявлення бот-мереж, корпоративна мережа.

АННОТАЦИЯ

Бобровникова К. Ю. Информационная технология обнаружения бот-сетей в корпоративных сетях на основе анализа DNS-трафика. – *Рукопись.*

Диссертация на соискание ученой степени кандидата технических наук по специальности 05.13.06 – Информационные технологии. – Тернопольский национальный технический университет имени Ивана Пулюя, Тернополь, 2016.

Диссертация посвящена решению актуальной научной задачи – созданию информационной технологии с целью повышения достоверности и эффективности обнаружения бот-сетей в корпоративных сетях на основе анализа DNS-трафика.

Проведен обзор и выявлены недостатки работы информационных технологий и программных средств обнаружения бот-сетей в сетях. Разработана модель бот-сетей с учетом DNS, модель DNS-трафика и модель процесса обнаружения бот-сетей в корпоративных сетях на основе анализа DNS-трафика. Разработан новый метод идентификации бот-сетей в корпоративных сетях на основе их групповой активности в DNS-трафике, который учитывает также анормальное поведение групп КС в DNS-трафике, свойственное многим видам бот-сетей. Разработанный метод позволяет идентифицировать неизвестные боты уже на начальной стадии распространения инфекции в корпоративной сети, может быть применен как для маленьких, так и для больших сетей, и не требует значительных объемов вычислительных ресурсов для обработки данных.

Разработан новый метод обнаружения бот-сетей, которые применяют технологии уклонения от обнаружения на основе DNS, такие как технология «быстросменных» сетей (fast flux service networks), «поток доменов» («domain flux»), периодическая смена IP-отображения для вредоносного домена (cycling of IP mapping) и DNS-туннелирование (DNS-tunneling). Метод использует два подхода: пассивный мониторинг входящего DNS-трафика и активное DNS-зондирование. С целью обнаружения бот-сетей, которые применяют технологии уклонения от обнаружения на основе DNS, используется нечеткая кластеризация c-means с частичным обучением. Объектами кластеризации являются векторы признаков, которые получены из полезной нагрузки входных DNS-сообщений относительно запрошенных компьютерными системами сети доменных имен, на основе пассивного мониторинга DNS-трафика. С целью устранения неопределенности части результатов кластеризации используются дополнительные признаки, которые указывают на применение технологий уклонения бот-сетей на основе DNS, которые могут быть получены на основе активного DNS-зондирования.

На базе моделей бот-сетей с учетом DNS и модели процесса обнаружения бот-сетей в корпоративных сетях на основе анализа DNS-трафика разработана информационная технология обнаружения бот-сетей в корпоративных сетях на основе анализа DNS-трафика, которая основывается на разработанных методе идентификации бот-сетей на основе их групповой активности в DNS-трафике и методе обнаружения бот-сетей, которые применяют технологии уклонения от обнаружения на основе DNS.

Осуществлена программная реализация информационной технологии обнаружения бот-сетей в корпоративных сетях на основе анализа DNS-трафика, что дало возможность обнаруживать известные и неизвестные боты бот-сетей с

высокой достоверностью. Программное обеспечение, реализующее информационную технологию обнаружения бот-сетей, позволяет выполнять следующие задачи: выявление известных и неизвестных ботов бот-сетей на основе пассивного мониторинга DNS-трафика и активного DNS-зондирования; локализация инфицированных ботами компьютерных систем сети. Использование разработанного программного обеспечения позволяет повысить уровень достоверности обнаружения бот-сетей на 8-22% по сравнению с известными антивирусными программными средствами и достичь снижения уровня погрешностей первого рода до 4%, что на 13-70% ниже по сравнению с известными антивирусными программными средствами.

Ключевые слова: бот, бот-сеть, DNS-трафик, групповая активность, технологии уклонения от обнаружения бот-сетей, обнаружение бот-сетей, корпоративная сеть.

ANNOTATION

Bobrovnikova K. Y. The Information technology for botnets detection in corporate area networks based on DNS-traffic analysis. – *Manuscript.*

Thesis for a Candidate of Technical Sciences degree in specialty 05.13.06 – Information technology. Ternopil Ivan Pul'uj National Technical University, 2016.

The dissertation is devoted to solving of the important scientific problem - the creation of information technology to increase reliability and efficiency of DNS-based botnet detection in the corporate area networks.

The weaknesses of the information technologies and software tools for botnets detection in the networks were outlined. The model of botnets which takes into account DNS, the model of DNS-traffic and the model of process of botnets detection in corporate area networks which is based on an analysis of the DNS-traffic were developed. On the basis of the developed models the information technology for botnet detection based on the analysis of DNS-traffic was developed. It is based on two new developed methods: the method of botnets identification based on their group activity in DNS-traffic and the method for botnets detection that use DNS-based evasion techniques.

The software of the information technology for botnets detection in the corporate area networks that based on an analysis of the DNS-traffic was developed. Usage of the developed software makes it possible to detect known and unknown bots of the botnets with high reliability.

Key words: bot, botnet, DNS-traffic, group activity, botnet`s evasion techniques, botnets detection, corporate area network.