УКРАЇНА **(19) UA** **(11) 118663** **(13) U**

(51) МПК
G06F 21/55 (2013.01)

МІНІСТЕРСТВО ЕКОНОМІЧНОГО РОЗВИТКУ І ТОРГІВЛІ УКРАЇНИ

(12) ОПИС ДО ПАТЕНТУ НА КОРИСНУ МОДЕЛЬ

(21) Номер заявки: **u 2016 12041**
(22) Дата подання заявки: **28.11.2016**
(24) Дата, з якої є чинними права на корисну модель: **28.08.2017**
(46) Публікація відомостей про видачу патенту: **28.08.2017, Бюл.№ 16**

(72) Винахідник(и):
Савенко Олег Станіславович (UA),
Бобровнікова Кіра Юліївна (UA),
Лисенко Сергій Миколайович (UA),
Савенко Богдан Олегович (UA),
Нічепорук Андрій Олександрович (UA)

(73) Власник(и):
ХМЕЛЬНИЦЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ,
вул. Інститутська, 11, м. Хмельницький, 29016 (UA)

(54) СПОСІБ ІДЕНТИФІКАЦІЇ БОТ-МЕРЕЖ У КОРПОРАТИВНИХ КОМП'ЮТЕРНИХ МЕРЕЖАХ НА ОСНОВІ АНАЛІЗУ DNS-ТРАФІКУ

(57) Реферат:

Спосіб ідентифікації бот-мереж у корпоративних комп'ютерних мережах на основі їх групової активності в DNS-трафіку, що 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 на основі того, що ідентифікація бот-мереж здійснюється шляхом збору вхідного DNS-трафіку та співставлення з «білим» та «чорним» списками доменних імен, що дозволяє виявляти групи КС, які ігнорують TTL-період з подальшою побудовою вектора щільності розподілу запитів в часі для перевірки синхронності запитів і побудовою матриці спостереження для збору та аналізу вхідного DNS-трафіку та виявлення групової активності шляхом аналізу групових запитів щодо одного й того самого доменного імені і для цього побудовою нижньотрикутної матриці мір Браун-Бланке для порівняння груп для формування векторів ознак для пар групових запитів та аналізу векторів ознак для ідентифікації інфікованих комп'ютерних систем.

UA 118663 U

Корисна модель належить до інформаційної безпеки і може використовуватись для ідентифікації бот-мереж у корпоративних комп'ютерних мережах.

Способи ідентифікації бот-мереж на основі їх групової активності в DNS-трафіку дають змогу ідентифікувати інфіковані комп'ютерні системи (КС), що здійснюють таку активність. Перевагами способів на основі DNS є: можливість здійснення ідентифікації бот-мереж на стадії створення їх інфраструктури; порівняно невеликий обсяг трафіку, а тому зменшення потреби в обчислювальних ресурсах, необхідних для аналізу; можливість ідентифікації бот-мереж виключно на основі спостереження за роботою комп'ютерних мереж, залишаючись непоміченим для зловмисника; на відміну від сигнатурних методів, здатність виявляти невідомі боти.

В [1] запропоновано динамічну систему оцінки репутації доменних імен, яка використовує три групи ознак для побудови моделей легітимних та шкідливих доменів - мережні, зональні та доказові ознаки, на основі яких здійснюється оцінка репутації домена. Мережні ознаки доменів (загальна кількість IP-адрес, пов'язаних з доменним ім'ям; їх географічна локація; кількість різних номерів автономних систем (ASN) для них тощо) та зональні ознаки (середня довжина доменних імен; кількість різних доменів верхнього рівня для них; частота, з якою з'являються в доменному імені різні символи тощо) одержуються на базі аналізу DNS-запитів. Доказові ознаки базовані на даних "чорних списків" та систем-«приманок" і дозволяють визначити, в якій мірі домен пов'язаний з відомими шкідливими доменними іменами або IP-адресами.

В [2] запропоновано евристику для виявлення DNSBL-розвідувальної діяльності ботмайстра та складання списків ймовірних ботів. Для визначення, чи знаходяться спам-боти в "чорному списку", ботмайстер виконує DNSBL-запити. DNSBL (DNS blacklist) є списками хостів, що зберігаються з використанням системи DNS та застосовуються для боротьби зі спамом. Недоліком [7] способу є те, що він включає використання одного хоста для вхідного та вихідного поштового серверів. Проте, в великих мережах вони можуть бути розділені, і тоді запити від вхідного поштового сервера можуть бути розцінені, як спроба розвідки.

В [3] запропоновано хост-орієнтований підхід, заснований на аналізі поведінки ботів і не ботів, пов'язаної з реакцією на DNS-відповіді. З метою створення надмірності та підвищення завадостійкості бот-мережі, окрім прямих спроб з'єднання через процедуру перетворення доменного імені, боти виконують зворотні DNS-запити для отримання додаткових доменних імен C&C-сервера. При цьому IP-адреси неуспішних зворотних DNS-запитів ігноруються доброякісним програмним забезпеченням, проте часто використовуються ботами та іншим шкідливим програмним забезпеченням (ШПЗ). Визначено чотири підозрілі процеси, що можуть мати місце в RD-поведінці (reaction-to-DNS response behavior, RD-behavior): (1) успішне пряме перетворення імені, підключення не відбулось - аномальний процес; (2) успішне зворотне перетворення імені, підключення не відбулось - притаманний для ШПЗ, в тому числі ботів; (3) неуспішне зворотне перетворення імені, підключення відбулось - притаманний для ШПЗ, в тому числі ботів; (4) неуспішне зворотне перетворення імені, підключення не відбулось - домінуючий для ботів.

Спосіб [4] передбачає моніторинг, захоплення DNS-трафіку в різних часових інтервалах та вимірювання відношення подібності між будь-якими двома групами КС, що запитують одне й те саме доменне ім'я. Для обчислення значення подібності між групами КС використовується коефіцієнт Жаккара. Недоліком методу є те, що він спирається на групові запити лише однакових доменних імен, не враховуючи міграцій C&C-серверів та інших DNS-запитів, пов'язаних з діяльністю бот-мережі.

Спосіб [5], орієнтований на виявлення ботів в мережах класу C, для обчислення подібності між двома групами КС використовує коефіцієнт Кульчинського. З метою виявлення міграцій C&C-серверів бот-мереж порівнюються списки IP-адрес КС, що запитували різні доменні імена, але які подібні за розмірами в межах 10 %. Недоліками [7] цього способу є значне зростання часу обробки та потреба у великих обсягах обчислювальних ресурсів при застосуванні до великих мереж, недостатня гнучкість механізму виявлення міграцій C&C-серверів та пов'язаних з функціонуванням бот-мережі DNS-запитів.

Для визначення множини доменних імен бот-мережі, і таким чином виявлення міграцій C&C-серверів, спосіб [6] використовує кластеризацію методом x-середніх (x-means) ознак, вилучених з DNS-трафіку. Виокремлено три групи таких ознак: (1) засновані на DNS-лексикології; (2) засновані на інформації, вилученій з DNS-запитів; (3) засновані на інформації, вилученій з DNS-відповідей. Перша група ознак об'єднує наступні ознаки, вилучені з доменного імені: (1) кількість міток в доменному імені; (2) середня довжина мітки домена; (3) найбільша довжина мітки домена; (4) наявність домена другого рівня, занесеного до "чорних списків". Друга група містить ознаки: (1) кількість надісланих запитів щодо доменного імені; (2) кількість різних IP-адрес КС, що надсилали запити; (3) кількість різних номерів автономних систем (ASN), до яких належать

IP-адреси КС, що надсилали запити; (4) тип запиту (A, NS, CNAME, MX, PTR); (5) оцінка подібності груп КС для домена. До третьої групи віднесено ознаки: (1) кількість різних повернутих IP-адрес доменного імені; (2) кількість різних номерів автономних систем, до яких належать IP-адреси доменного імені; (3) кількість різних країн локації IP-адрес доменного імені; (4) значення поля TTL в DNS-відповіді. З метою обчислення значення подібності між двома групами КС використовується косинусний коефіцієнт. Недоліками способу, описаного в [6] є те, що він орієнтований на виявлення ботів у великих розподілених мережах, тому не придатний для невеликих локальних мереж [7]. Коротка тривалість періоду моніторингу (одна година), зумовлена зменшенням ймовірності зміни динамічних IP-адрес як ідентифікаторів КС в мережі, призводить до неспроможності виявлення групових запитів, якщо повторний запит групи КС відбувся поза межами періоду моніторингу.

Спільним недоліком способів [4, 5, 6, 7] є довільний поділ періоду моніторингу на інтервали, в межах яких здійснюється пошук груп інфікованих КС, що призводить до зменшення рівня виявлення. Іншим недоліком описаних методів є використання для порівняння двох груп КС симетричних мір подібності (Жаккара, Кульчинського, косинусного коефіцієнта), що є доцільним для оцінки подібності рівновеликих груп, тому може призводити до хибних спрацювань.

Найближчим аналогом заявленого способу можна вважати мультиагентний спосіб локалізації бот-мереж у корпоративних комп'ютерних мережах [8] та спосіб виявлення комп'ютерних атак нейромережевою штучною імунною системою, описаний в [9].

У відомих аналогах є недостатньо висока достовірність ідентифікації нових бот-мереж.

Задачею корисної моделі є підвищення достовірності ідентифікації бот-мереж у корпоративних комп'ютерних мережах на основі аналізу DNS-трафіку.

Поставлена задача вирішується тим, що ідентифікація бот-мереж в корпоративних мережах ґрунтується на властивості групової активності ботів в DNS-трафіку [7]. Така активність проявляється в зосереджених в невеликому проміжку часу групових DNS-запитах КС під час спроб доступу до командно-контролюючих серверів, їх міграціях, виконанні команд або скачуванні оновлень шкідливого програмного забезпечення. При цьому враховуються особливості поведінки інфікованих груп КС, характерні для багатьох видів бот-мереж: групи КС ігнорують TTL-період DNS, тобто очищують локальні кеші DNS та здійснюють повторні запити щодо доменного імені до завершення TTL-періоду, а також здійснюють DNS-запити, використовуючи нелокальні DNS-сервери, і також відслідковується підвищена кількість порожніх DNS-відповідей з кодом помилки RCODE=3 (NXDOMAIN); враховується міграція C&C-серверів та інших DNS-запитів, пов'язаних з функціонуванням бот-мережі; враховується можлива зміна розмірів груп КС в результаті поширення ботів.

Для оцінки різних за розмірами груп використано несиметричну міру подібності на відміну від [4, 5, 6], де для порівняння груп КС використовуються симетричні міри подібності, які доцільно використовувати для оцінки подібності рівновеликих груп. Спосіб передбачає збір лише DNS-трафіку і поділ періоду моніторингу на інтервали, в межах яких здійснюється пошук інфікованих груп КС, ґрунтується на основі врахування значень TTL, які містяться в DNS-повідомленнях.

Спосіб містить наступні кроки:

1) збір вхідного DNS-трафіку;

2) співставлення з "білим" та "чорним" списками доменних імен;

3) виявлення груп КС, які ігнорують TTL-період;

4) побудова вектора щільності розподілу запитів в часі для перевірки синхронності запитів;

5) побудова матриці спостереження для збору та аналізу вхідного DNS-трафіку;

6) виявлення групової активності шляхом аналізу групових запитів щодо одного й того самого доменного імені;

7) побудова нижньотрикутної матриці мір Браун-Бланке для порівняння груп;

8) формування векторів ознак для пар групових запитів;

9) аналіз векторів ознак для ідентифікації інфікованих КС.

Вхідний DNS-трафік збирається за допомогою множини мережних давачів $E=\{e_i\}_{i=1}^{N_E}$, де N_E - кількість давачів, підключених до дзеркалюючих портів комутаторів.

Для відкидання легітимних DNS-запитів здійснюється співставлення зібраних даних з "білим" списком відомих легітимних доменних імен. Для виявлення DNS-запитів до відомих шкідливих доменних імен здійснюється співставлення зібраних даних з "чорним" списком відомих шкідливих доменних імен.

На наступному етапі здійснюється виявлення груп КС корпоративної мережі, які ігнорують TTL-період. З цією метою КС очищають локальні кеші DNS для уможливлення здійснення

повторних DNS-запитів в межах TTL-періоду DNS. Для виявлення цього факту будується матриця спостереження V_{MAC}, кожен рядок якої містить MAC-адреси КС, які здійснювали запити щодо конкретного доменного імені в межах TTL-періоду.

Таким чином, рядки матриці V_{MAC} містять MAC-адреси КС, які, ймовірно, здійснюють групову активність. Якщо MAC-адреса КС представлена в групі G_d, то у відповідній комірці матриці позначається "1", інакше - "0". Якщо КС повторно надсилала запит щодо доменного імені d, то MAC-адреса КС позначається "1" в рядку матриці V_{MAC}, створеному для повторного запиту:

$$V_{MAC}(h_{j,i}) = \begin{cases} 0, \text{if } h_j \notin G_d \\ 0, \text{if } h_j \in G_d \end{cases}, \text{ where } j = \overline{1, N_G},$$

де h_j - MAC-адреси КС, які здійснювали DNS-запити щодо d в межах TTL, і - номер рядка матриці.

Формування груп триває до спливання найбільшого значення TTL-періоду, отриманого в DNS-відповіді щодо повторного запиту.

Якщо N_G та $N_{G_{rep}}$ розміри груп для попереднього та повторного групових запитів, δ - порогове значення подібності між двома групами і $\delta \cdot N_G > N_{G_{rep}}$ $S\text{-}N_G > N_c$, то рядок матриці V_{MAC} для повторного запиту відкидається. Для групових запитів, які не були відкинуті на цьому етапі, перевіряється їх синхронність.

Побудова вектора щільності розподілу DNS-запитів в часі для перевірки синхронності DNS-запитів. Групи запитів синхронні, якщо спостерігається велика кількість запитів для доменного імені в межах часу, коли боти бот-мережі здійснюють запити - часу синхронізації ботів ts. 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: $z = (t_{last} - t_{first}) / \left(\frac{1}{3} ts\right)$, 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. Такий поділ надає можливість мінімізувати кількість DNS-запитів, які не потрапили в інтервал ts (Фіг.1, а).

Для групового запиту будується z-елементний вектор щільності розподілу запитів в часі $\overline{W_d} = (\Omega_j)_{j=1}^{z}$, де Ω_j - кількість запитів в межах z-го інтервалу. Для елемента вектора $\overline{W_d}$ з максимальним значенням Ω_{max} в межах у $j = max \pm 2$ відшукуються два суміжні елементи з найбільшими значеннями таким чином, щоб всі три елементи описували розподіл запитів неперервного інтервалу часу, та обчислюється їх сума (Sum_s).

Якщо $(1-\delta)(Sum_s + Sum_r) \geq Sum_r$, то множини MAC-адрес груп КС в матриці V_{MAC} об'єднуються, і груповий запит підлягає подальшому аналізу, інакше така група відкидається, де Sum_r - сума значень решти елементів вектора $\overline{W_d}$ (Фіг.1, б).

З метою збору та подальшого аналізу вхідного DNS-трафіку для кожного визначеного інтервалу часу моніторингу t_m будується матриця спостереження M_m, m - номер ітерації спостереження. Вона містить доменні імена d_i запитані групами КС; MAC-адреси груп КС h_i, отримані з матриці V_{MAC}; 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 h_i 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":

$$M_m(d, F) = \begin{cases} 0, \text{if } \left(\forall n\left(h_j^{\Delta t}\right) = 1\right) \vee \gamma < \delta, \\ 1, \text{if } \gamma \geq \delta, \end{cases} \quad (2)$$

де $n\left(h_j^{\Delta t}\right)$ - кількість появ MAC-адреси КС в Δt.

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":

$$M_m(d,F)=\begin{cases}0, \text{if } \left(\exists\chi_{d,k,j,IP}\in\Xi\right)\wedge\left(\exists\chi_{j,k,j,IP}\notin\Xi\right)\\ 0{,}5, \text{if } \forall\chi_{d,k,j,IP}\in\Xi\\ 1, \text{if } \forall\chi_{d,k,j,IP}\notin\Xi\end{cases}, \quad (3)$$

де Ξ - множина IP-адрес локальних DNS-серверів мережі.

Якщо DNS-відгуки для групи містили код помилки NXDOMAIN, то у комірці матриці спостереження $M_m(d_i, R)$ позначається "1", інакше "0":

$$M_m(d,R)=\begin{cases}1, \text{if } \chi_{j,k,HD,RC}=3,\\ 0 \text{ otherwise.}\end{cases} \quad (4)$$

Комірки матриці спостереження $M_m(d_i, M)$ та $M_m(d_i, N)$ заповнюються нулями. В комірку матриці спостереження $M_m(d_i, N_G)$ заноситься кількість MAC-адрес, представлених у відповідній групі.

Для зниження рівня хибних спрацювань прийнято пороговий розмір інфікованих груп $n_t=4$, які можуть бути виявлені запропонованим способом. DNS-запити груп меншого розміру відкидаються.

Для порівняння двох груп КС G_1 та G_2, які надсилали DNS-запити щодо двох доменних імен d_1 та d_2 в інтервалах часу Δt_1, та Δt_2 відповідно, використано коефіцієнт Браун-Бланке [7], який є несиметричною мірою подібності, а тому придатний для оцінки різних за розмірами груп і дозволяє оцінити подібності двох груп КС з високою точністю:

$$K_B(G_1,G_2)=\frac{N_0}{\max\left[N_{G_1}, N_{G_2}\right]}, \quad (5)$$

де N_o - кількість спільних елементів в групах G_1 та G_2;

N_{G1} та N_{G2} - кількість КС в групах G_1 та G_2 відповідно; $K_B(G_1,G_2)\in[0,1]$.

Якщо кількість порівнюваних груп більша двох, то для оцінки подібності груп КС використовується індекс дисперсності Коха [7]:

$$K_K(G_1,...,G_q)=\frac{C-A}{(q-1)\cdot A}, \quad (6)$$

де $G_1,\ldots,G_q$ - порівнювані групи КС;

q - кількість порівнюваних груп;

$C=\sum_{i=1}^{q}N_{G_i}$ - загальна кількість MAC-адрес в усіх групах;

A - кількість різних MAC-адрес, представлених в групах; $K_K(G_1,...,G_q)\in[0,1]$.

Групи КС вважатимуться інфікованими, якщо коефіцієнт подібності для груп перевищує порогове значення $K_B\geq\delta$ або $K_K\geq\delta$, де δ - порогове значення подібності. Крім цього введено додатково порогове значення подібності δ', яке вказує на підозрілість груп КС, якщо $\delta'\leq K_B<\delta$ або $\delta'\leq K_K<\delta$. Як ідентифікатори КС в мережі використано MAC-адреси за умови забезпечення запобігання підміни MAC-адрес.

На наступному етапі здійснюють аналіз матриці спостереження M_m з метою виявлення групових запитів щодо однакових доменних імен. Для цього порівнюють групи за MAC-адресами. В залежності від кількості групових запитів щодо певного доменного імені d обирається коефіцієнт Браун-Бланке для порівняння двох груп або індекс дисперсності Коха для 3 і більше груп (Фіг. 2, а). Якщо результат порівняння перевищує поріг $K_B\geq\delta$ або $K_K\geq\delta$, то групи КС вважаються інфікованими.

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, $M_m(d_1,F)=1$, та використання групами нелокальних DNS-серверів, $M_m(d_1,S) = 1$. Якщо для будь-якого з групових запитів спостерігалось групове ігнорування TTL-періоду або для всіх запитів

спостерігалось звертання до нелокальних DNS-серверів, то групи КС вважаються інфікованими. Інакше, групи КС вважаються підозрілими.

Якщо групи, які запитували одне й те саме доменне ім'я, визначені інфікованими або підозрілими, то множини їх MAC-адрес об'єднуються в один рядок для доменного імені d в матриці спостереження M_m (Фіг. 2, б) з метою подальшого пошуку пов'язаних з групою DNS-запитів. Якщо група КС була визначена як інфікована, то в комірці матриці спостереження $M_m(d, M)$ проставляється "1", якщо група КС була визначена як підозріла - "0.5":

$$M_m(d,M)=\begin{cases} 1, \text{if } \tau(G_1,...,G_n)=\text{infected}, \\ 0{,}5, \text{if } \tau(G_1,...,G_n)=\text{suspicious}. \end{cases} \quad (7)$$

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

У випадку об'єднання множин MAC-адрес КС комірки матриці спостереження M_m заповнюються за наступними правилами. Якщо для будь-якого з групових запитів спостерігалось групове ігнорування TTL-періоду, в комірці матриці спостереження $M_m(d, F)$ проставляється "1", інакше - "0":

$$M_m(d,F)=\begin{cases} 1, \text{if } \exists F=1, \\ 0 \text{ otherwise}, \end{cases} \quad (8)$$

де $\left\{F \middle| F \in \left\{M_m(d_j,F)\right\}_{j=G_1}^{G_n}\right\}$.

Якщо запити груп КС здійснювались як до локального, так і до інших DNS-серверів, то у комірці матриці спостереження $M_m(d, S)$ проставляється "0"; якщо запити груп здійснювались лише до локального DNS-сервера, то у комірці матриці спостереження $M_m(d, S)$ проставляється "0.5", якщо запити груп здійснювались до нелокальних DNS-серверів, то у комірці матриці спостереження $M_m(d, S)$ проставляється "1":

$$M_m(d,S)=\begin{cases} 0{,}5, \text{ if } \forall S=0{,}5, \\ 1, \text{if } \forall S=1, \\ 0 \text{ otherwise}, \end{cases} \quad (9)$$

де $\left\{S \middle| S \in \left\{M_m(d_j,S)\right\}_{j=G_1}^{G_n}\right\}$.

Якщо DNS-відгуки для останньої групи КС містили код помилки NXDOMAIN, то у комірці матриці спостереження $M_m(d, R)$ проставляється "1", інакше "0":

$$M_m(d,R)=\begin{cases} 1, \text{if } M_m(d_{G_n},R)=1, \\ 0 \text{ otherwise}, \end{cases} \quad (10)$$

де $M_m(d_{G_n},R)$ - значення комірки для групи КС, що останньою запитувала доменне ім'я.

В комірках матриці спостереження $M_m(d, N)$ для груп КС, які вважаються підозрілими, проставляється номер ітерації спостереження:

$M_m(d, N) = m$. (11)

Якщо жодна з умов не задовольняється, то групові запити для таких доменних імен видаляються з матриці спостереження M_m.

На наступному етапі 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 (Фіг. 3). Обчислення значень комірок для кожного стовпця припиняється, якщо $N_{Gi}/N_{Gi+1} < \delta`$ (тобто відношення розмірів порівнюваних груп є меншим за порогове значення $\delta`$ - порожні комірки матриці B_m). Це дозволить зменшити час та обчислювальні ресурси, необхідні для аналізу.

Для кожної пари групових запитів, якщо виконується умова $K_B \geq \delta`$, згідно з матрицею мір Браун-Бланке B_m формується вектор ознак $\overline{W_{G_1,G_2}}$ (Фіг. 4).

Вектор 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", "Dangerous" (властива ботам):

$$\overline{W_{G_1,G_2}} = \left(K_B(G_1,G_2), S_{G_1,G_2}, F_{G_1,G_2}, R_{G_1,G_2}, M_{G_1,G_2},\right) \quad (12)$$

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} визначаються наступним чином:

$$S_{G_1,G_2} = \begin{cases} \text{Unusual, if } B_m(d_1,S) = B_m(d_2,S) = 0, \\ \text{Neutral, if } B_m(d_1,S) = B_m(d_2,S) = 0{,}5, \\ \text{Dangerous, if } B_m(d_1,S) = B_m(d_2,S) = 1, \\ \text{Suspicious otherwise.} \end{cases} \quad (13)$$

$$M_{G_1,G_2} = \begin{cases} \text{Neutral, if } B_m(d_1,M) = B_m(d_2,M) = 0, \\ \text{Suspicious, if } \left(B_m(d_1,M) = 0{,}5 \vee B_m(d_2,M) = 0{,}5\right) \wedge \\ \wedge B_m(d_1,M) \neq 1 \wedge B_m(d_2,M) \neq 1 \wedge B_m(d_1,M) \neq B_m(d_2,M), \\ \text{Dangerouz, if } B_m(d_1,M) = 1 \vee B_m(d_2,M) = 1 \vee \\ \vee \left(B_m(d_1,M) = B_m(d_2,M) = 0{,}5 \wedge B_m(d_1,N) \neq B_m(d_2,N) \vee \right. \\ \left. \vee B_m(d_1,N) = B_m(d_2,N) = 0\right). \end{cases} \quad (14)$$

Зведені ознаки $F_{G1,G2}$ та $R_{G1,G2}$ визначаються аналогічно. Нижче наведено приклад для першої з них:

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

На наступному етапі з метою ідентифікації інфікованих КС здійснюється аналіз векторів ознак для пар групових запитів. Аналіз векторів ознак $\overline{W_{G_1,G_2}}$ здійснюється за наступними правилами, де функція виходу $f\left(W_{G_1,G_2}\right)$ може приймати чотири значення: "Not_Infected" (неінфіковані), "Not_Suspicious" (не підозрілі), "Suspicious" (підозрілі), "Infected" (інфіковані):

$$f\left(\overline{W_{G_1,G_2}}\right) = \begin{cases} \text{Not\_\_Infected, if } K_B(G_1,G_2) < \delta \wedge S_{G_1,G_2} \neq \text{Unusual} \wedge \\ \wedge \forall \overline{W_{G_1,G_2}}(j) \neq \text{Suspicious} \wedge \forall \overline{W_{G_1,G_2}}(j) \neq \text{Dangerous}, \\ \text{Not\_\_Suspicious, if } K_B(G_1,G_2) < \delta \wedge S_{G_1,G_2} \neq \text{Unusual} \wedge \\ \wedge \forall \overline{W_{G_1,G_2}}(j) \neq \text{Suspicious} \wedge \forall \overline{W_{G_1,G_2}}(j) \neq \text{Dangerous}, \\ \text{Infected, if } \exists \overline{W_{G_1,G_2}}(j) = \text{Dangerous} \vee K_B(G_1,G_2) \geq \delta, \\ \text{Suspicious otherwise.} \end{cases} \quad (16)$$

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

Одна й та сама група в межах ітерації може отримати декілька різних оцінок. В такому випадку пріоритет має оцінка з вищим ступенем небезпечності. Групи КС, які було визначено як не інфіковані, відкидаються. 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}) для виявлення можливих повторних групових запитів. При цьому, якщо група, яка запитувала доменне ім'я d, була визначена підозрілою, в комірці матриці спостереження $M_{m+1}(d, M)$ для цієї групи проставляється "0.5", а в комірці M_{m+1} (d, N) - номер ітерації m.

Представлений схемою зв'язку кроків на Фіг. 5 розроблений спосіб ідентифікації бот-мереж в корпоративних комп'ютерних мережах на основі аналізу DNS-трафіка ґрунтується на властивості групової активності ботів в DNS-трафіку та враховує анормальну поведінку груп КС, властиву багатьом видам бот-мереж. Він дає змогу ідентифікувати ще невідомі боти вже на початковій стадії поширення інфекції в корпоративній мережі та може бути застосований як до

малих, так і до великих мереж, та не вимагає значних обсягів обчислювальних ресурсів для обробки даних.

Розроблений спосіб ідентифікації бот-мереж у корпоративних комп'ютерних мережах на основі їх групової активності в DNS-трафіку 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.

Джерела інформації:

1. Antonakakis M. Building a Dynamic Reputation System for DNS [Text] / M. Antonakakis, R. Perdisci, D. Dagon, W. Lee, N. Feamster // 19th Usenix Security Symposium, 2010. -PP. 273-290.

2. Ramachandran A. Revealing botnet membership using DNSBL counter-intelligence [Text] / A. Ramachandran, N. Feamster, D. Dagon // 2nd Workshop on Steps to Reducing Unwanted Traffic on the Internet (SRUTI 2006), 2006. - P. 8-13.

3. Morales J.A. Analyzing DNS activities of bot processes [Text] / J.A. Morales, A. Al-Bataineh, S. Xu, R. Sandhu // Proceedings of the 4th International Conference on Malicious and Unwanted Software (MALWARE), 2009. - P. 98-103.

4. Manasrah A.M. Detecting Botnet Activities Based on Abnormal DNS traffic [Text] / A.M. Manasrah, A. Hasan, O.A. Abouabdalla, S. Ramadass // International Journal of Computer Science and Information Security (IJCSIS), Vol. 6, № l, 2009. - P. 97-104.

5. Choi H. Botnet Detection by Monitoring Group Activities in DNS Traffic [Text] / H. Choi, H. Lee, H. Lee, H. Kim // Seventh IEEE International Conference on Computer and Information Technology (CIT 2007), 2007. - P. 715-720.

6. Choi H. Identifying botnets by capturing group activities in DNS traffic [Text] / H. Choi, H. Lee // Computer Networks, 56, 2012. - P. 20-33.

7. 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. - P. 127-138, ISBN: 978-3-319-19418-9.

8. 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, СМ. 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.

9. Спосіб виявлення комп'ютерних атак нейромережевою штучною імунною системою [Текст]: пат. №74822 Україна, МПК(2012) H04W 12/08, G06F 21/00, G06F 12/14. Комар М. П., Саченко А. О., Головко В. А., Безобразов С. В.; заявник і патентовласник Тернопільський національний економічний університет. -№u201205349; заявл. 28.04.12; опубл. 12.11.12, Бюл. №21.

ФОРМУЛА КОРИСНОЇ МОДЕЛІ

Спосіб ідентифікації бот-мереж у корпоративних комп'ютерних мережах на основі їх групової активності в DNS-трафіку, який **відрізняється** тим, що 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 на основі того, що ідентифікація бот-мереж здійснюється шляхом збору вхідного DNS-трафіку та співставлення з «білим» та «чорним» списками доменних імен, що дозволяє виявляти групи КС, які ігнорують TTL-період з подальшою побудовою вектора щільності розподілу запитів в часі для перевірки синхронності запитів і побудовою матриці спостереження для збору та аналізу вхідного DNS-трафіку та виявлення групової активності шляхом аналізу групових запитів щодо одного й того самого доменного імені і для цього побудовою нижньотрикутної матриці мір Браун-Бланке для порівняння груп з метою формування векторів ознак для пар групових запитів та аналізу векторів ознак для ідентифікації інфікованих комп'ютерних систем, де для кожної пари групових запитів, якщо виконується умова $K_B \geq \delta`$, де $K_B$ - коефіцієнт Браун-Бланке, обчислений для пар груп комп'ютерних систем, δ` -

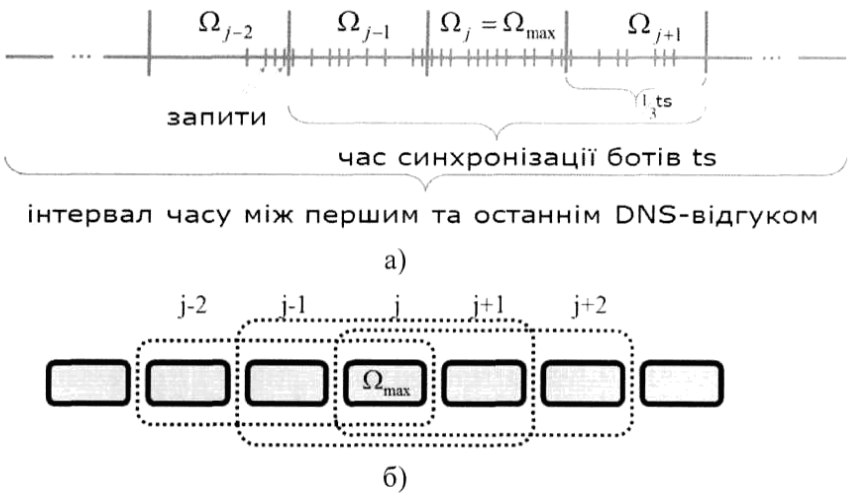
порогове значення подібності, яке вказує напідозрілість груп комп'ютерних систем, згідно матриці мір Браун-Бланке B_m формується вектор ознак $\overline{W_{G_1,G_2}}$ з п'яти елементів: коефіцієнт Браун-Бланке та зведені поведінкові ознаки для двох порівнюваних груп, отримані на основі матриці B_m, які можуть приймати наступні значення: "Unusual" (непритаманна ботам), "Neutral" (властива як користувачам, так і ботам), "Suspicious", "Dangerous" (властива ботам):

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

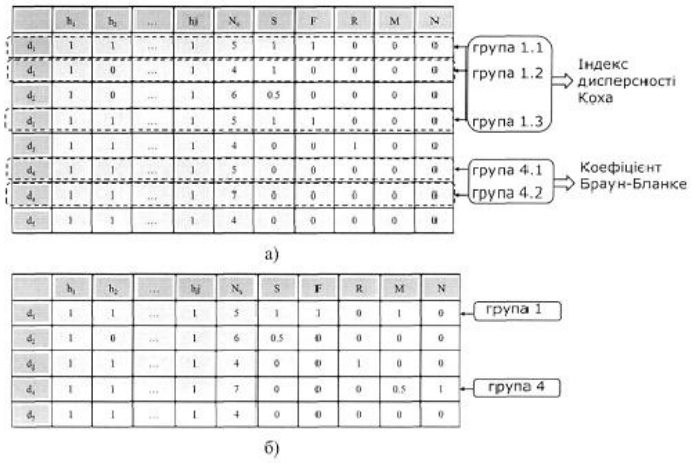
де $S_{G_1,G_2}, F_{G_1,G_2}, R_{G_1,G_2}, M_{G_1,G_2}$ - зведені поведінкові ознаки для двох порівнюваних груп і зведені поведінкові ознаки S_{G_1,G_2} та M_{G_1,G_2}, F_{G_1,G_2} та R_{G_1,G_2}, а далі здійснюється аналіз векторів ознак для пар групових запитів $\overline{W_{G_1,G_2}}$ за наступними правилами, де функція виходу $f\left(W_{G_1,G_2}\right)$ може приймати чотири значення: "Not_Infected" (неінфіковані), "Not_Suspicious" (не підозрілі), "Suspicious" (підозрілі), "Infected" (інфіковані):

$$f\left(\overline{W_{G_1,G_2}}\right) = \begin{cases} \text{Not\_\_Infec ted, if } K_B(G_1,G_2) < \delta \wedge S_{G_1,G_2} \neq \text{Unusual} \wedge \\ \wedge \forall \overline{W_{G_1,G_2}}(j) \neq \text{Suspicious} \wedge \forall \overline{W_{G_1,G_2}}(j) \neq \text{Dangerous}, \\ \text{Not\_\_Suspicio us, if } K_B(G_1,G_2) < \delta \wedge S_{G_1,G_2} \neq \text{Unusual} \wedge \\ \wedge \forall \overline{W_{G_1,G_2}}(j) \neq \text{Suspicious} \wedge \forall \overline{W_{G_1,G_2}}(j) \neq \text{Dangerous}, \\ \text{Infected, if } \exists \overline{W_{G_1,G_2}}(j) = \text{Dangerous} \vee K_B(G_1,G_2) \geq \delta, \\ \text{Suspicious otherw ise} \end{cases},$$

де $j = \overline{2,5}$ - номер елемента в векторі ознак, причому одна й та сама група в межах ітерації може отримати декілька різних оцінок, то в такому випадку пріоритет має оцінка з вищим ступенем небезпечності, причому групи КС, які було визначено як не інфіковані, відкидаються, а щодо груп КС, визначених як інфіковані, здійснюються заходи з метою ліквідації інфекції (блокування, усунення вразливостей системи, встановлення (оновлення) антивірусного програмного забезпечення тощо), а стосовно ж 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}) для виявлення можливих повторних групових запитів і при цьому, якщо група, яка запитувала доменне ім'я d, була визначена підозрілою, в комірці матриці спостереження $M_{m+1}(d,M)$ для цієї групи проставляється "0,5", а в комірці $M_{m+1}(d,N)$ - номер ітерації т .



Фіг.1



Фіг.2

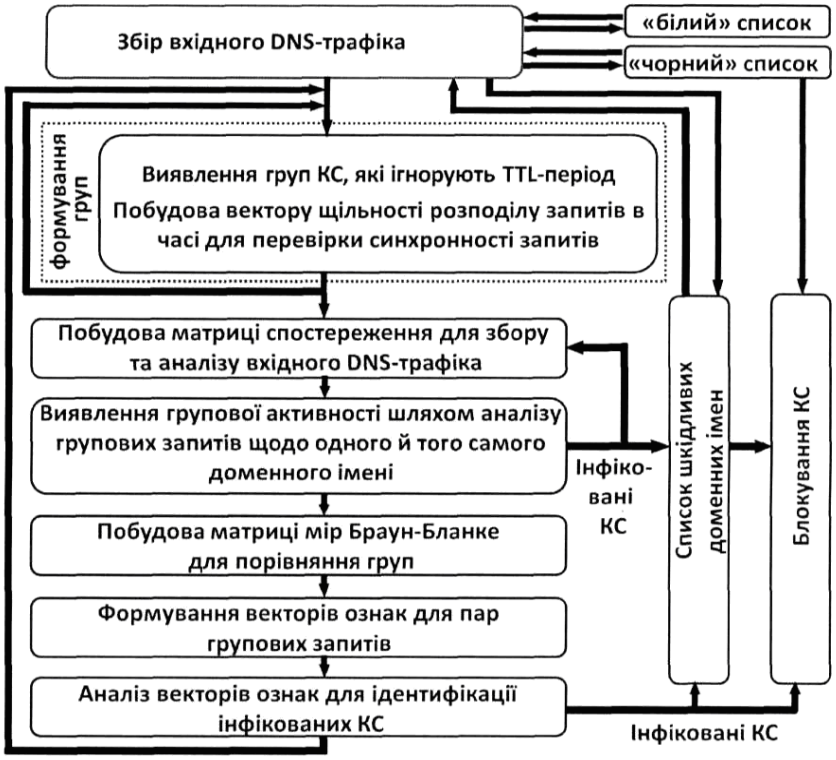
	d_3	d_5	d_1	d_2	d_4	…	N_G	S	F	R	M	N
d_3	1					…	4	0	0	1	0	0
d_5	1	1				…	4	0	0	0	0	0
d_1	0.8	0.8	1			…	5	1	1	0	1	0
d_2			0.5	1		…	6	0.5	0	0	0	0
d_4			0.71	0.43	1	…	7	0	0	0	0.5	1

Фіг.3

	d_3	d_5	d_1	d_2	d_4	…	N_G	S	F	R	M	N
d_3	1					…	4	0	0	1	0	0
d_5	1	1				…	4	0	0	0	0	0
d_1	0.8	0.8	1			…	5	1	1	0	1	0
d_2			0.5	1		…	6	0.5	0	0	0	0
d_4			0.71	0.43	1	…	7	0	0	0	0.5	1

$K_0(G_1,G_4)$	$S_{G1,G4}$	$F_{G1,G4}$	$R_{G1,G4}$	$M_{G1,G4}$
0.71	Suspicious	Suspicious	Neutral	Dangerous

Фіг.4



Фіг.5

Комп'ютерна верстка Л. Бурлак

Міністерство економічного розвитку і торгівлі України, вул. М. Грушевського, 12/2, м. Київ, 01008, Україна

ДП "Український інститут інтелектуальної власності", вул. Глазунова, 1, м. Київ – 42, 01601