УДК 004.491.2

О.С. САВЕНКО, С.М. ЛИСЕНКО, К.Ю. БОБРОВНІКОВА

Хмельницький національний університет, м. Хмельницький

DNS-МЕТОД ВИЯВЛЕННЯ БОТ-МЕРЕЖ

Анотація. Представлено розроблений DNS-метод виявлення бот-мереж, заснований на властивості синхронної скоординованої активності інфікованих хостів в DNS-трафіку, який ґрунтується на аналізі TTL-періодів, отриманих в DNS-відповідях, та враховує нетипові для звичайних користувачів особливості поведінки, властиві багатьом видам ботнетів: ігнорування TTL-періоду, здійснення DNS-запитів поза локальними DNS-серверами та підвищену кількість порожніх DNS-відповідей з кодом помилки NXDOMAIN (доменне ім’я не існує). Наведено результати експериментів, проведених з метою перевірки ефективності запропонованого методу. Метод дозволяє здійснювати виявлення на початковій стадії поширення інфекції в мережі та виявляти ще невідомі боти.

Ключові слова: синхронна скоординована активність в DNS-трафіку, бот-мережа, бот.

Аннотация. Представлен разработанный DNS-метод обнаружения бот-сетей, основанный на свойстве синхронной скоординированной активности инфицированных хостов в DNS-трафике, который базируется на анализе TTL-периодов, полученных в DNS-ответах, и учитывает нетипичные для обыкновенных пользователей особенности поведения, свойственные многим видам ботнетов: игнорирование TTL-периода, осуществление DNS-запросов вне локальных DNS-серверов и повышенное количество пустых DNS-ответов с кодом ошибки NXDOMAIN (доменное имя не существует). Приведены результаты экспериментов, проведенных с целью проверки эффективности предложенного метода. Метод позволяет осуществлять обнаружение на начальной стадии распространения инфекции в сети и обнаруживать еще неизвестные боты.

Ключевые слова: синхронная скоординированная активность в DNS-трафике, бот-сеть, бот.

Abstract. The developed DNS-based method for botnets detection that is based on the property of the synchronous coordinated activity of infected hosts in DNS-traffic, which is based on the analysis of TTL-periods that were obtained in DNS-responses, and considers atypical for a normal user behaviors that are inherent of many types of botnets: ignoring of the TTL-period, the implementation of DNS-requests outside the local DNS servers and the increased number of empty DNS-responses with error code NXDOMAIN (domain name does not exist) was presented. The results of experiments that were conducted to verify the effectiveness of the proposed method are presented. The method allows to perform the detection at the initial stage of infection in the network and to detect unknown bots.

Key words: synchronous coordinated activity in DNS traffic, botnet, bot.

Вступ

Сьогодні ботнети є не лише інструментом для злочинного або нелегального заробітку в мережі Інтернет, а й засобом здійснення кіберзлочинів. Динамічна географічно розподілена структура та можливість анонімного керування інфікованими хостами незалежно від їх географічного розташування перетворює бот-мережі на глобальну загрозу Інтернет-безпеці. Традиційні методи виявлення ботнетів, базовані на фільтрації пакетів, аналізі трафіка на основі портів, що використовуються, а також відомих сигнатур легко обходяться зловмисником шляхом динамічної зміни шкідливого коду, системи керування та портів, або використанням стандартних портів HTTP/S. При цьому повний аналіз вмісту пакетів є ресурсоємним та погано масштабується в мережах з високим навантаженням. Системи виявлення на базі сигнатур не здатні вирішувати проблем «нульового дня», тому схильні до високої ймовірності неспрацювань, що є основною причиною порушення безпеки. Метод систем-«приманок» – замкнених захищених контролюємих областей, що імітують вразливі мережі, ресурси або служби, застосовується передусім в якості систем спостереження для виявлення спроб вторгнення, дослідження мотивації та технік, що використовуються злочинцем, збору сигнатур та прогнозування розвитку загроз. Проте такий метод виявлення неефективний в перші години інфекції і не піддається масштабуванню.

Актуальність

Переважна більшість ботнетів для керування та контролю над інфікованими хостами використовує DNS [1]. Характерною особливістю поведінки таких видів ботів є скоординованість в DNS-трафіку. Відомі методи, описані в [2-4], які ґрунтуються на цій особливості, мають наступні недоліки: спирання на групові запити лише однакових доменних імен або недостатня гнучкість механізму виявлення міграцій C&C-серверів та пов’язаних запитів; потреба у великих обсягах обчислювальних ресурсів та значний час обробки при застосуванні до великих мереж; коротка тривалість періоду моніторингу; довільний поділ періоду моніторингу на інтервали, в межах яких здійснюється пошук інфікованих груп. Тому актуальною науково-технічною задачею є розробка методу виявлення бот-мереж на основі аналізу DNS-трафіка, який би усував наведені недоліки.

Мета

Метою роботи є розробка методу виявлення бот-мереж в корпоративних мережах. В роботі досліджується можливість виявлення ботів, які входять до складу централізованих ботнетів.

Задачі

4. Розробка методу виявлення бот-мереж на основі аналізу DNS-трафіка, який би усував недоліки існуючих підходів та надавав можливість виявляти ще невідомі боти, а також здійснювати раннє виявлення – на початковій стадії поширення інфекції в мережі.
5. Проведення експериментів з метою перевірки ефективності розробленого методу.

Розв'язання задач

Запропонований метод ґрунтується на властивості синхронної скоординованої активності ботів в DNS-трафіку, що проявляється в одночасних або зосереджених в невеликому проміжку часу DNS-запитах груп хостів під час спроб доступу до C&C-серверів, їх міграціях, виконанні команд або скачуванні оновлень шкідливого програмного забезпечення. Метод враховує нетипові для звичайних користувачів особливості поведінки, властиві багатьом видам ботнетів: ігнорування TTL-періоду [1,5], здійснення DNS-запитів поза локальними DNS-серверами [1] та підвищену кількість порожніх DNS-відповідей з кодом помилки RCODE=3 (NXDOMAIN, доменне ім'я не існує). Принцип функціонування методу подано на рис. 1.

Збір вхідного DNS-трафіка. Вхідний DNS-трафік збирається за допомогою множини мережних давачів $S = \{s_i\}_{i=1}^{N_S}$, де N_s – кількість давачів, підключених до дзеркалюючих портів комутаторів.

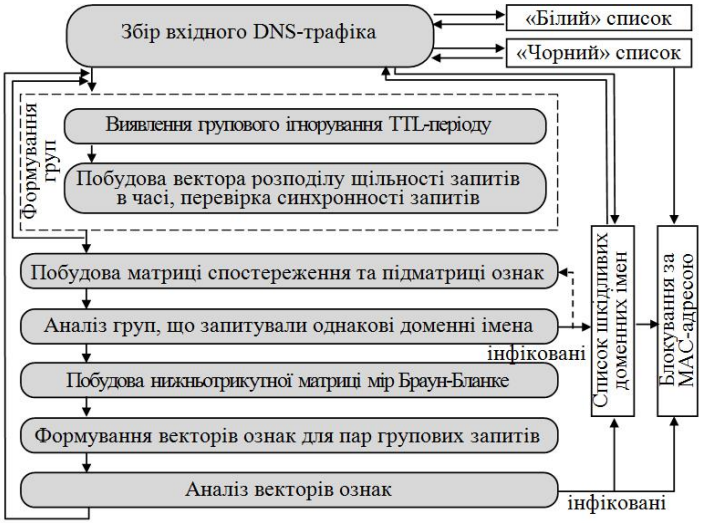

Рисунок 1 – Принцип функціонування методу

Співставлення запитів з «білим» та «чорним» списками. З метою відкидання легітимних запитів та виявлення запитів до відомих шкідливих доменних імен здійснюється співставлення зібраних даних з «білим» та «чорним» списками.

Побудова матриці спостереження та підматриці ознак. Матриця спостереження M_k будується для кожної години часу моніторингу t_k для множини часових інтервалів $T = \{t_k\}_{k=1}^{T_m}$, де T_m – загальний час моніторингу (щонайменше 6 годин). Таким чином, отримуємо множину матриць спостереження $M = \{M_k\}_{k=1}^{T_m}, M_k = (m_{ij})_{i=1,\,j=1}^{N_q,h+6}$, де N_q – кількість групових DNS-запитів, h – кількість різних MAC-адрес, представлених в групах, 6 – кількість ознак в підматриці ознак. Якщо група хостів в межах інтервалу часу Δt_i надсилала запити щодо доменного імені d_i як до локального, так і до інших DNS-серверів, у комірці підматриці ознак $M_k(d_i,S)$ проставляється «0», лише до локального DNS-сервера – «0.5», лише до нелокальних DNS-серверів – «1». Якщо DNS-відгуки для групи містили код помилки NXDOMAIN, у комірці $M_k(d_i,R)$ підматриці ознак проставляється «1», інакше «0». Комірки підматриці ознак $M_k(d_i,M)$ та $M_k(d_i,N)$ обнуляються, в комірках $M_k(d_i,Q)$ проставляється кількість MAC-адрес, представлених у відповідній групі, комірки $M_k(d_i,F)$ заповнюються на наступному етапі. З метою зниження рівня хибних спрацювань прийнято пороговий розмір інфікованих груп $n_t = 4$, які можуть бути виявлені запропонованим методом. DNS-запити груп меншого розміру відкидаються.

Виявлення групового ігнорування TTL-періоду. З метою виявлення групового очищення локальних DNS-кешів на хостах для DNS-відгуків щодо кожного доменного імені від першого зафіксованого DNS-відгуку до спливання TTL-періоду, отриманого в ньому[1], будується матриця відображення MAC-адрес хостів, що запитували доменне ім'я. Якщо $t_{d_i} \geq t_0 + t_s$, де t_{d_i} – час отримання першого DNS-відгуку щодо доменного імені d_i відносно часу початку моніторингу t_0, t_s – часове вікно, в межах якого здійснюється пошук синхронних скоординованих запитів, і MAC-адреса хоста представлена в групі, у відповідній комірці матриці відмічається «1», інакше – «0». Якщо хост повторно надсилав запит щодо d_i, MAC-адреса хоста помічається «1» в рядку матриці, створеному для повторного запиту. Якщо $TTL^r_{d_i m'} < TTL_{d_i m'+1}$, де $TTL^r_{d_i m'} = TTL_{d_i^{m'}} - t_{d_i m'+1}$ – залишок часу до видалення DNS-записів з локальних DNS-кешів на хостах, що запитували d_i, згідно попереднім відгукам, $TTL_{d_i^{m'}}$ – TTL-період, в

[1] При використанні хостами різних DNS-серверів в DNS-відгуках можуть міститись різні значення TTL-періодів в залежності від того, чи були і як давно кешовані DNS-записи для доменного імені на цих серверах. Для типу ботнетів, що розглядається в роботі, така можливість виключена, оскільки боти зберігатимуть синхронність в діях, і, будучи викликаною діями користувачів, не впливатиме на точність виявлення.

межах якого здійснюється пошук, $t_{d_i^{m'+1}}$ – час надходження DNS-відгуку щодо повторного запиту відносно початку зворотнього відліку $TTL_{d_i^{m'}}$, $TTL_{d_i^{m'+1}}$ – TTL-період, отриманий в DNS-відгуку щодо повторного запиту, тоді формування груп триває до сплівання $TTL_{d_i^{m'+1}}$. Якщо $t_{d_i} < t_0 + t_s$, можливо, що частина запитів групи відбулась до початку моніторингу. Тому, якщо зафіксовано кількість повторів MAC-адрес, що більша або дорівнює n_t, для всіх DNS-відгуків перевіряється відповідність значення отриманого TTL-періоду щодо залишку TTL-періоду для попередніх DNS-відгуків, і за приналежністю за цим параметром до різних інтервалів, або, якщо розбіжностей між цими значеннями виявлено не було, відносно інтервалу в часі між повторними запитами груп, який перевищує t_s, вирівнюється поділ повторних запитів на групи. Якщо $\delta \cdot n\left(G_{d_i^{m'}}\right) > n\left(G_{d_i^{m'+x}}\right), m \in Z, x \in Z$, де $n\left(G_{d_i^{m'}}\right)$ та $n\left(G_{d_i^{m'+x}}\right)$ – розміри груп для попереднього та повторного групових запитів, δ – порогове значення подібності між двома групами, рядок для повторного запиту відкидається. Для групових запитів, які не було відкинуто на цьому етапі, перевіряється їх синхронність, як описано нижче. Якщо запити були синхронними, множини MAC-адрес груп хостів об’єднуються та переносяться до матриці спостереження M_k, у комірці підматриці ознак M_k (d_i,F) проставляється «1», якщо групового очищення локальних DNS-кешів не виявлено – «0»:

$$M_k(d_i, F) = \begin{cases} 0, if\ \left(\forall n\left(MAC_j^{\Delta t_i}\right) = 1\right) \vee \left(\delta \cdot n\left(G_{d_i^{m'}}\right) > n\left(G_{d_i^{m'+x}}\right)\right), \\ 1, if\ \delta \cdot n\left(G_{d_i^{m'}}\right) \le n\left(G_{d_i^{m'+x}}\right), \end{cases} \tag{1}$$

де $n\left(MAC_j^{\Delta t_i}\right)$ – кількість поява MAC-адреси хоста в Δt_i, $MAC_j^{\Delta t_i} \in G_{d_i^{m'}}, j = \overline{1, n\left(G_{d_i^{m'}}\right)}, m \in Z, x \in Z$.

Побудова вектора розподілу щільності запитів в часі, перевірка синхронності запитів. Якщо інтервал часу Δt_q між першим та останнім DNS-відгуками для групового запиту щодо d_i більший за тривалість часового вікна t_s, Δt_q розбивається на z інтервалів:

$$z = \left(t_{d_i}^{last} - t_{d_i}^{first}\right) \Big/ \tfrac{1}{3} t_s, \tag{2}$$

де $t_{d_i}^{last}$ та $t_{d_i}^{first}$ – час надходження останнього та першого DNS-відгуків щодо d_i в межах TTL-періоду TTL_{d_i}, де здійснюється пошук, відповідно, або останнього та першого DNS-відгуків щодо d_i, які містили код помилки NXDOMAIN в межах t_k, або останнього та першого DNS-відгуків в межах групи, що повторюється, якщо зафіксовано групове очищення локальних кешів DNS.

Для групового запиту будується z-елементний вектор розподілу щільності запитів в часі $\overline{W_{d_i}} = \left(Query\_Count_j\right)_{j=1}^{z}$, де $Query\_Count_j$ – кількість запитів в межах z-го інтервалу. Для елемента вектора з максимальним значенням $Query\_Count_{max}$ в межах j = max±2, де max – номер елемента з максимальним значенням, відшукуються два суміжні елементи з найбільшими значеннями таким чином, щоб всі три описували розподіл запитів неперервного інтервалу часу, та обчислюється їх сума (Sum_s). Якщо $(1-\delta) \cdot Sum_s > Sum_r$, де Sum_r – сума значень решти елементів вектора, група запитів підлягає подальшому аналізу, інакше – відкидається. Таким чином, інтервал часу Δt_i, в межах якого здійснюється пошук інфікованих груп хостів, визначається за формулою:

$$\Delta t_i = \begin{cases} t_{d_i}^{last} - t_{d_i}^{first}, if\ t_{d_i}^{last} - t_{d_i}^{first} \le t_s, \\ t_s, if\ \left(\left(t_{d_i}^{last} - t_{d_i}^{first} > t_s \vee RCODE_{d_i} = 3\right) \wedge (1-\delta)\cdot Sum_s > Sum_r\right), \end{cases} \quad (3)$$

де $RCODE_{d_i}$ – поле коду DNS-відгуку щодо d_i.

Якщо $TTL_{d_i} \le t_k - t_{d_i}^{first}$, рядок для доменного імені будується в матриці спостереження M_k. Якщо $TTL_{d_i} > t_k - t_{d_i}^{first}$, і в межах TTL_{d_i} не спостерігалось групове очищення локальних кешів DNS на хостах, рядок для доменного імені переноситься до матриці спостереження M_{k+1}.

Оцінка подібності груп хостів. Для порівняння двох груп хостів G_{d_i} та $G_{d_{i+x}}$, x∈Z, які надсилали DNS-запити щодо двох доменних імен d_i та d_{i+x} в інтервали часу Δt_i та Δt_{i+x} відповідно, використаємо коефіцієнт Браун-Бланке [6], який дозволить зменшити рівень хибних спрацювань:

$$K_B\left(G_{d_i}, G_{d_{i+x}}\right) = \frac{n\left(G_{d_i} \cap G_{d_{i+x}}\right)}{\max\left[n\left(G_{d_i}\right), n\left(G_{d_i+x}\right)\right]}, \quad (4)$$

де $n\left(G_{d_i} \cap G_{d_{i+x}}\right)$ – кількість спільних елементів в групах G_{d_i} та $G_{d_{i+x}}$; $n\left(G_{d_i}\right)$ та $n\left(G_{d_i+x}\right)$ – кількість хостів в групах G_{d_i} та $G_{d_{i+x}}$ відповідно; $K_B\left(G_{d_i}, G_{d_{i+x}}\right) \in [0,1]$.

Якщо кількість порівнюваних груп більша двох, їх подібність обчислюється як індекс дисперсності Коха [7]:

$$K_K\left(G_{d'}, ..., G_{d^{n(G)'}}\right) = \frac{C - A}{(n-1)\cdot A}, \quad (5)$$

де $G_{d'}, ..., G_{d^{n(G)'}}$ – порівнювані групи хостів; n(G) – кількість порівнюваних груп; $C = \sum_{i=1}^{n(G)} n\left(G_{d^{i'}}\right)$ – загальна кількість MAC-адрес в усіх групах; $A = n\left(G_{d'} \cup ... \cup G_{d^{n(G)'}}\right)$ – кількість різних MAC-адрес, представлених в групах; $K_K\left(G_{d'}, ..., G_{d^{n(G)'}}\right) \in [0,1]$.

Групи хостів вважатимуться інфікованими, якщо коефіцієнт подібності для них перевищуватиме порогове значення: $K_S \ge \delta$, де $K_S = K_B$ або $K_S = K_K$, δ – порогове значення подібності. Також введемо додатково порогове значення подібності δ`, за якого групи хостів вважатимуться підозрілими.

Оскільки при динамічному розподілі IP-адреса не може бути надійним ідентифікатором для комп'ютерів в мережі, в якості ідентифікаторів хостів використаємо MAC-адреси за умови забезпечення запобігання підміни MAC-адрес в мережі.

Аналіз груп, що запитували однакові доменні імена. На даному етапі здійснюється аналіз груп хостів, що запитували однакові доменні імена в межах t_k. Для матриці M_k, k=1, якщо коефіцієнт подібності $K_B\left(G_{d'}, G_{d''}\right) < \delta'$, та якщо $n\left(G_{d'}\right) < \delta' \cdot n\left(G_{d''}\right)$, де G_d` та G_d`` – перший в часі і наступний групові запити відповідно, можливо, що частина запитів першої групи відбулась раніше початку моніторингу, тому рядок для першого групового запиту видаляється (часові характеристики запитів не зберігаються в матриці спостереження, оскільки такі дані потрібні лише в даному окремому випадку). Якщо для груп, що запитували однакові доменні імена, $K_S \ge \delta$, вони вважаються інфікованими. Якщо $\delta' \le K_S < \delta$, тоді: (1) якщо для будь-якого з групових запитів спостерігалось очищення локальних DNS-кешів або для всіх запитів звертання до нелокальних DNS-серверів, групи хостів вважаються інфікованими, а доменне ім'я шкідливим; доменне ім'я d заноситься в список шкідливих доменних імен; (2) інакше групи вважаються підозрілими. Якщо жодна з умов не задовольняється, доменне ім'я вважається легітимним, групові запити для нього видаляються з матриці. Якщо групи, які запитували одне й те саме доменне ім'я, визначені інфікованими або підозрілими, множини їх MAC-адрес об'єднуються, в матриці залишається один рядок для доменного імені d для подальшого пошуку

пов’язаних з групою запитів: $G_d = G_{d'} \cup ... \cup G_{d^{n(G)'}}$. Якщо група хостів була визначена як інфікована, в комірці підматриці ознак $M_k(d,M)$ проставляється «1», підозріла – «0.5»:

$$M_k(d,M) = \begin{cases} 1, if\ K_S \geq \delta \vee (\delta' \leq K_S < \delta \wedge (\forall M_k(d,S) = 1 \vee \exists M_k(d,F) = 1)), \\ 0.5, if\ \delta' \leq K_S < \delta \wedge (\exists M_k(d,S) \neq 1 \wedge \forall M_k(d,F) \neq 1), \end{cases} \quad (6)$$

де $M_k(d,S) \in \{M_k(d_j,S)\}_{j=1}^{n(G)}, M_k(d,F) \in \{M_k(d_j,F)\}_{j=1}^{n(G)}, d_j = d.$

Значення комірки підматриці ознак $M_k(d,F)$ обчислюється наступним чином:

$$M_k(d,F) = \begin{cases} 1, if\ \exists M_k(d,F) = 1, \\ 0\ else, \end{cases} \quad (7)$$

де $M_k(d,F) \in \{M_k(d_j,F)\}_{j=1}^{n(G)}, d_j = d.$

Значення комірки $M_k(d,S)$ обчислюється аналогічно попередньому етапу. В комірку підматриці ознак $M_k(d,R)$ заноситься значення для останньої групи, що запитувала доменне ім’я.

Побудова нижньотрикутної матриці мір Браун-Бланке. Множина матриць мір Браун-Бланке $B = \{B_k\}_{k=1}^{T_m}, B_k = (b_{ij})_{i=1,j=1}^{n(d_i),n(d_i)+5}$, де $n(d_i)$ – кількість групових запитів щодо різних доменних імен, 5 – кількість ознак в підматриці ознак, будується на основі відповідних матриць спостереження за зростанням кількості MAC-адрес в групах, по стовпчикам. В діагональних комірках матриці відмічається загальна кількість MAC-адрес груп: $B_k(d_i,d_i) = n(G_{d_i})$. Для кожної пари груп хостів обчислюється коефіцієнт Браун-Бланке: $B_k(d_{i+x},d_i) = n(G_{d_{i+x}} \cap G_{d_i}) / n(G_{d_{i+x}})$, де $n(G_{d_{i+x}})$ – діагональний елемент, що відповідає рядку. Обчислення значень комірок для стовпчика припиняється, якщо $n(G_{d_i}) / n(G_{d_{i+x}}) < \delta'$, що дозволить зменшити час та обчислювальні ресурси, необхідні для аналізу.

Формування векторів ознак для пар групових запитів. Для кожної пари групових запитів, для якої виконується умова $K_B \geq \delta`$, згідно матриці мір Браун-Бланке формується вектор ознак, який складається з п’яти елементів: коефіцієнт Браун-Бланке як міра подібності груп хостів за складом, та зведені поведінкові ознаки для двох порівнюваних груп, отримані на основі підматриці ознак, які можуть приймати наступні значеня: "Unusual" (непритаманна ботам), "Neutral" (властива як звичайним користувачам, так і ботам), "Suspicious" (підозріла), "Dangerous" (небезпечна, притаманна ботам). Вектор ознак може бути визначений наступним чином:

$$\overline{W_{d_i,d_{i+x}}} = (K_B(G_{d_i},G_{d_{i+x}}), S_{d_i,d_{i+x}}, F_{d_i,d_{i+x}}, R_{d_i,d_{i+x}}, M_{d_i,d_{i+x}}), \quad (8)$$

де $S_{d_i,d_{i+x}}, F_{d_i,d_{i+x}}, R_{d_i,d_{i+x}}, M_{d_i,d_{i+x}}$ – зведені поведінкові ознаки для двох порівнюваних груп.

Зведені поведінкові ознаки $S_{d_i,d_{i+x}}$ та $M_{d_i,d_{i+x}}$ можуть бути визначені наступним чином:

$$S_{d_i,d_{i+x}} = \begin{cases} Unusual, if\ B_k(d_i,S) = B_k(d_{i+x},S) = 0, \\ Neutral, if\ B_k(d_i,S) = B_k(d_{i+x},S) = 0.5, \\ Dangerous, if\ B_k(d_i,S) = B_k(d_{i+x},S) = 1, \\ Suspicious\ \ else. \end{cases} \quad (9)$$

$$M_{d_i,d_{i+x}} = \begin{cases} Neutral, if\ B_k(d_i,M) = B_k(d_{i+x},M) = 0, \\ Suspicious, if\ ((B_k(d_i,M) = 0.5 \vee B_k(d_{i+x},M) = 0.5) \wedge \\ \wedge B_k(d_i,M) \neq 1 \wedge B_k(d_{i+x},M) \neq 1) \wedge B_k(d_i,M) \neq B_k(d_{i+x},M), \\ Dangerous, if\ B_k(d_i,M) = 1 \vee B_k(d_{i+x},M) = 1 \vee \\ \vee (B_k(d_i,M) = B_k(d_{i+x},M) = 0.5 \wedge B_k(d_i,N) \neq B_k(d_{i+x},N) \vee \\ \vee B_k(d_i,N) = B_k(d_{i+x},N) = 0), \end{cases} \quad (10)$$

де $B_k(d_i,N)$ та $B_k(d_{i+x},N)$ – номер ітерації (k), відмічається лише для груп, визначених підозрілими на етапі аналізу векторів ознак в кожній ітерації процесу аналізу.

Зведені ознаки $F_{d_i,d_{i+x}}$ та $R_{d_i,d_{i+x}}$ визначаються аналогічно, наведемо приклад для першої з них:

$$F_{d_i,d_{i+x}} = \begin{cases} Neutral, if\ B_k(d_i,F) = B_k(d_{i+x},F) = 0, \\ Suspicious, if\ B_k(d_i,F) \neq B_k(d_{i+x},F), \\ Dangerous, if\ B_k(d_i,F) = B_k(d_{i+x},F) = 1. \end{cases} \quad (11)$$

Аналіз векторів ознак. Аналіз векторів ознак здійснюється за наступними правилами, де функція виходу f може приймати чотири значення: "Not_Infected" (неінфіковані), "Not_Suspicious" (не підозрілі), "Suspicious" (підозрілі), "Infected" (інфіковані):

$$f(\overline{W_{d_i,d_{i+x}}}) = \begin{cases} Not\_Infected, if\ K_B(G_{d_i},G_{d_{i+x}}) < \delta \wedge S_{d_i,d_{i+x}} = Unusual \wedge \\ \wedge \forall \overline{W_{d_i,d_{i+x}}}(j) \neq Suspicious\ \wedge \forall \overline{W_{d_i,d_{i+x}}}(j) \neq Dangerous, \\ Not\_Suspicious, if\ K_B(G_{d_i},G_{d_{i+x}}) < \delta \wedge S_{d_i,d_{i+x}} \neq Unusual \wedge \\ \wedge \forall \overline{W_{d_i,d_{i+x}}}(j) \neq Suspicious\ \wedge \forall \overline{W_{d_i,d_{i+x}}}(j) \neq Dangerous, \\ Infected, if\ \exists \overline{W_{d_i,d_{i+x}}}(j) = Dangerous \vee K_B(G_{d_i},G_{d_{i+x}}) \geq \delta, \\ Suspicious\ else. \end{cases} \quad (12)$$

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 вища за ступенем небезпечності. Групи хостів, які було визначено як не інфіковані, відкидаються. Щодо груп хостів, визначених як інфіковані, здійснюються заходи з метою ліквідації інфекції (блокування, усунення вразливостей систем, встановлення (оновлення) антивірусного ПЗ тощо). Групи хостів з матриці спостереження M_k, які не потрапили до матриці мір Браун-Бланке B_k, групи, для яких не було виконано умову $K_B \geq \delta'$, а також визначені як не підозрілі та підозрілі, додаються до даних, отриманих на наступній ітерації спостереження (матриця спостереження M_{k+1}) та знову аналізуються з метою виявлення можливих повторних групових запитів в інтервалі часу t_{k+1}. При цьому, якщо група, яка запитувала доменне ім'я d, була визначена підозрілою, в комірці підматриці ознак M_{k+1}(d,M) для групи проставляється «0.5», а в комірці M_{k+1}(d,N) – номер ітерації k. Ітеративний характер процесу аналізу дозволить підвищити оперативність виявлення інфікованих хостів, оскільки синхронна скоординована активність ботів може бути помічена вже на першій ітерації.

Експерименти

Для перевірки запропонованого підходу було проведено ряд експериментів. З цією метою була згенерована бот-мережа централізованої архітектури, що складалася з k ботів, $k = \overline{1,50}$, які було встановлено на наявні в мережі комп'ютерні системи. Оскільки IP-адреси хостів досліджуваної мережі були сталими в ході експерименту, для його спрощення DNS-трафік локальної мережі збирався в єдиній точці. Для перехоплення DNS-трафіка було застосовано утиліту tcpdump, встановлену на хост, використаний в якості давача, що дозволило збирати та зберігати дані в файлі з метою їх подальшої обробки.

Було проведено 9 експериментів, для кожного з яких змінювались основні параметри, що використовує метод: часове вікно, в межах якого здійснювався пошук синхронних запитів, та порогові значення коефіцієнтів подібності, як показано в табл.1.

Таблиця 1 – Вплив основних параметрів, що використовує метод, на процес виявлення бот-мереж

№ експеримента	T_m, год	t_s, с	δ'	δ	% виявлення	% хибних спрацювань
1	8	30	0.5	0.7	100	12
2	8	15	0.5	0.7	88	10
3	8	5	0.5	0.7	86	0
4	8	30	0.6	0.8	96	8
5	8	15	0.6	0.8	98	0
6	8	5	0.6	0.8	90	0
7	8	30	0.7	0.9	82	0
8	8	15	0.7	0.9	80	0
9	8	5	0.7	0.9	76	0

Було встановлено, що зменшення значення t_s призводить до зниження рівня виявлення, в той час як його збільшення підвищує рівень хибних спрацювань. Зменшення порогових значень коефіцієнта подібності також призводитиме до підвищення рівня хибних спрацювань, а його збільшення – до зниження рівня виявлення через можливість відключення інфікованих хостів. Збільшення часу моніторингу підвищуватиме рівень виявлення, оскільки підвищується ймовірність відслідковування повторних запитів інфікованих груп хостів. Таким чином, адекватними пороговими значеннями для показників, що впливають на рівень виявлення, можуть бути: t_s =15 с, T_m = 6 год (враховуючи тривалість TTL-періодів для більшості ботнетів), δ'=0.6, δ=0.8.

Висновки

Запропоновано DNS-метод виявлення бот-мереж, заснований на властивості синхронної скоординованої активності інфікованих хостів в DNS-трафіку, який ґрунтується на аналізі TTL-періодів, отриманих в DNS-відповідях, та враховує характерні для багатьох видів ботнетів особливості поведінки: ігнорування TTL-періоду, здійснення DNS-запитів поза локальними DNS-серверами та підвищену кількість NXDOMAIN-відповідей. Метод дозволяє виявляти ще невідомі боти, а також здійснювати раннє виявлення – на початковій стадії поширення інфекції в мережі. В подальшому необхідним є дослідження параметрів, що використовує метод, для підвищення ефективності та достовірності роботи методу виявлення бот-мереж.

Список літератури

1. DAMBALLA. Botnet Detection for Communications Service Providers [Електронний ресурс]. – Режим доступу: https://www.damballa.com/downloads/r_pubs/ WP_Botnet_Detection_for_CSPs.pdf.

2. Botnet Detection by Monitoring Group Activities in DNS Traffic. Choi, H., Lee, H., Lee, H., Kim, H.: Seventh IEEE International Conference on Computer and Information Technology (CIT 2007), 2007. – pp. 715-720.

3. Detecting Botnet Activities Based on Abnormal DNS traffic. Manasrah, A.M., Hasan, A., Abouabdalla, O. A., Ramadass, S.: International Journal of Computer Science and Information Security (IJCSIS), Vol. 6, No.1, 2009. – pp. 97–104.

4. Identifying botnets by capturing group activities in DNS traffic. Choi, H., Lee, H.: Computer Networks, 56, 2012. – pp. 20-33.

5. Schiller, C. Botnets: The Killer Web Application/ Craig Schiller, James R. Binkley. – Syngress Publishing, 2007. – 464 p.

6. Pflanzensoziologie Grundzüge der Vegetationskunde. Braun-Blanquet J.: Berlin: Verlaq von Julius springer, 1928. – 330 s.

7. Index of biotal dispersity. Koch, L.// Ecology, V. 38, No. 1, 1957. – pp. 145-148.

Відомості про авторів

Савенко Олег Станіславович – канд. техн. наук, доц., декан факультету програмування та комп'ютерних і телекомунікаційних систем, Хмельницький національний університет.

Лисенко Сергій Миколайович – канд. техн. наук, доц, Хмельницький національний університет.

Бобровнікова Кіра Юліївна – аспірант, Хмельницький національний університет.