УДК 004.491.2

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

Хмельницький національний університет, Україна

ІНФОРМАЦІЙНА ТЕХНОЛОГІЯ ВИЯВЛЕННЯ БОТ-МЕРЕЖ НА ОСНОВІ АНАЛІЗУ DNS-ТРАФІКА

Розроблено інформаційну технологію виявлення бот-мереж на основі аналізу DNS-трафіка, яка побудована на базі двох нових методів: методу виявлення бот-мереж на основі їх групової активності в DNS-трафіку та методу виявлення бот-мереж, які застосовують технології ухилення від виявлення на основі DNS. Перевагами інформаційної технології є те, що процес діагностування не потребує побудови баз сигнатур, надає змогу виявлення вже на початковій стадії поширення інфекції в мережі, може бути застосовуваний як до маленьких, так і до великих мереж, не вимагає значних обсягів обчислювальних ресурсів для обробки даних, дозволяє виявляти відомі та невідомі бот-мережі.

Ключові слова: *бот-мережа, DNS-трафік, групова активність, технології ухилення бот-мереж, «швидкозмінні» мережі, «потік доменів», DNS-тунелювання.*

Вступ

На сьогоднішній день бот-мережі є однією з найбільш небезпечних кібер-загроз [1]. Аналіз існуючих підходів виявлення бот-мереж [2, 3] показав їх недоліки та неспроможність виявлення з високою ефективністю. Тому з метою усунення недоліків відомих інформаційних технологій (ІТ) та підвищення достовірності виявлення бот-мереж в корпоративних мережах було розроблено нову ІТ виявлення бот-мереж на основі аналізу DNS-трафіка.

1. Інформаційна технологія виявлення бот-мереж на основі аналізу DNS-трафіка

Інформаційна технологія побудована на базі моделей ботів та бот-мереж з врахуванням використання ними DNS в процесі функціонування, особливостей поведінки ботів в DNS-трафіку та використання бот-мережами технологій ухилення від виявлення на основі DNS. ІТ надає можливість здійснювати виявлення комп'ютерних систем (КС) в корпоративній мережі, інфікованих як відомими, так і невідомими бот-мережами. Розроблена ІТ представлена узагальненою схемою, поданою на рис. 1. ІТ побудована на базі двох методів виявлення бот-мереж: методу виявлення бот-мереж на основі їх групової активності в DNS-трафіку [2] та методу виявлення бот-мереж, які застосовують технології ухилення від виявлення на основі DNS [3].

ІТ використовує базу знань, що містить: «білий» список популярних легітимних доменних імен; «чорний» список відомих доменних імен бот-мереж; «сірий» список доменних імен бот-мереж, виявлених розробленою ІТ; множини правил для формування та аналізу векторів ознак групових DNS-запитів, які враховують поведінку інфікованих груп КС в DNS-трафіку; знання щодо ознак, які можуть бути отримані з вхідних DNS-повідомлень до ботів бот-мереж, які використовують технології ухилення від виявлення на основі DNS, та ознаки доменних імен таких бот-мереж.

1.1. Метод виявлення бот-мереж на основі їх групової активності в DNS-трафіку

Характерною ознакою функціонування бот-мереж є групова активність ботів. В DNS-трафіку така активність полягає в ініціюванні DNS-запитів групою інфікованих КС одночасно або в невеликому інтервалі часу під час спроб доступу до командно-контролюючих серверів бот-мереж (C&C-серверів), при їх міграціях, виконанні команд або скачуванні оновлень шкідливого програмного забезпечення.

З метою усунення недоліків відомих підходів, заснованих на цій властивості, було розроблено новий метод виявлення бот-мереж [2]. Для виявлення групової активності ботів в DNS-трафіку здійснюється поділ КС мережі, які надсилали DNS-запити, на групи наступним чином. Якщо локальний кеш DNS комп'ютерної системи містить ресурсні записи DNS для доменного імені, повторний DNS-запит цієї КС не виходить за межі локального кеша DNS. Проте для багатьох видів бот-мереж характерною властивістю є ігнорування TTL-періодів DNS, тобто очищення локальних кешів DNS та виконання повторних DNS-запитів щодо доменного імені в межах TTL. Тому 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 [2].

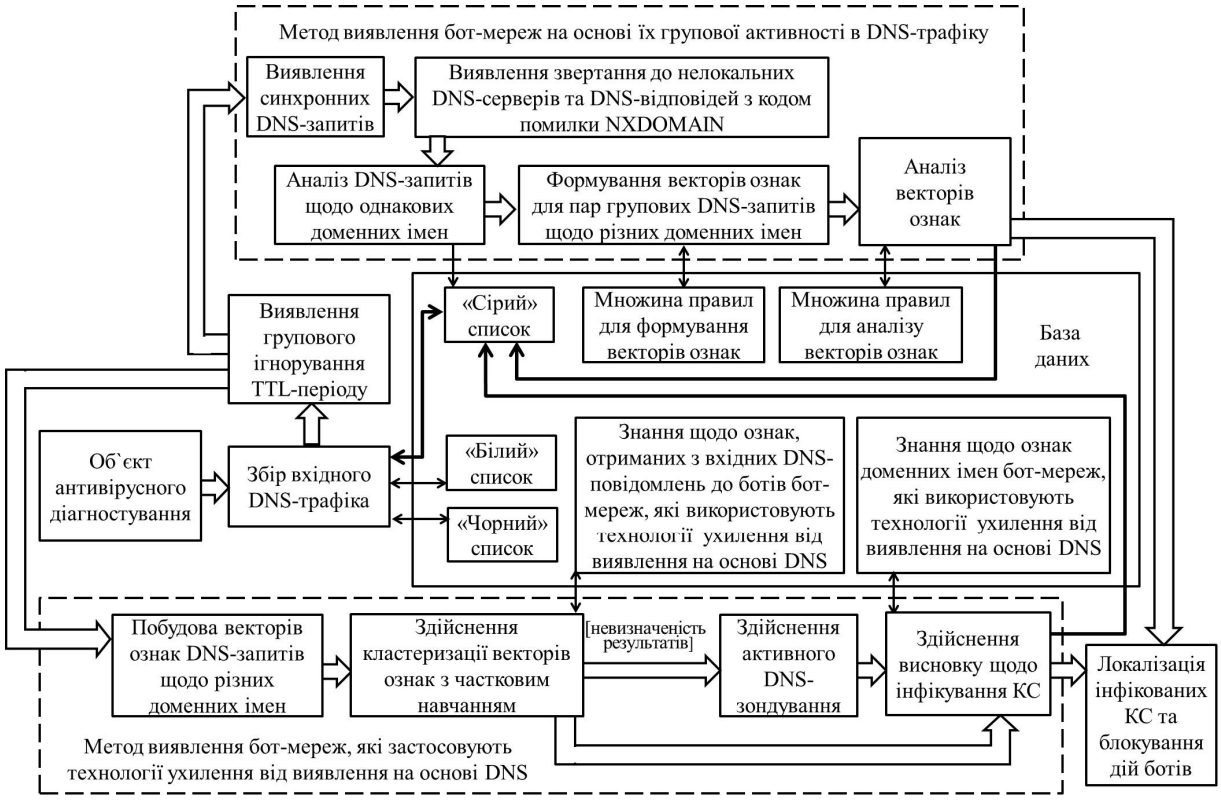

Рис. 1. Схема функціонування ІТ виявлення бот-мереж на основі аналізу DNS-трафіка

Для виокремленого групового DNS-запиту щодо доменного імені перевіряється синхронність запитів. Групові DNS-запити вважатимемо синхронними, якщо спостерігається велика кількість запитів для доменного імені, зосереджених в межах невеликого інтервалу часу, коли боти бот-мережі здійснюють DNS-запити – часу синхронізації ботів t_s. З метою перевірки синхронності запитів здійснюється побудова вектора щільності розподілу DNS-запитів в часі [2]:

$$\overline{W_g} = \left(\Omega_j\right)_{j=1}^{z}, \qquad (1)$$

де Ω_j – кількість DNS-запитів в межах z-го інтервалу; $z = \left(t_{last} - t_{first}\right) / \frac{1}{3} t_s$, де t_{first} та t_{last} – час надходження відповідно першого та останнього DNS-відгуків щодо доменного імені в межах TTL-періоду, протягом якого здійснюється пошук групової активності, або було зафіксовано групове очищення локальних кешів DNS.

Якщо максимальна частка запитів, яка припадає на інтервал t_s, що складається з трьох суміжних елементів вектора $\overline{W_g}$, які описують розподіл DNS-запитів неперервного інтервалу часу, перевищує прийняте порогове значення, то групу запитів вважатимемо синхронною [2]. DNS-запити, визначені як синхронні, підлягають подальшому аналізу, решта – відкидаються.

Визначення приналежності до бот-мереж груп КС, що запитували однакові доменні імена, здійснюється на основі аналізу подібності цих груп за MAC-адресами та аналізу наявності у них особливостей поведінки в DNS-трафіку, притаманних для бот-мереж, а саме: ігнорування TTL-періодів, здійснення DNS-запитів до нелокальних DNS-серверів, а також наявність порожніх DNS-відповідей з кодом помилки RCODE=3 (NXDOMAIN, доменне ім'я не існує). З метою визначення подібності груп КС використовується коефіцієнт Браун-Бланке (для двох груп) або індекс дисперсності Коха (для трьох та більше груп) [2].

Визначення приналежності до бот-мереж груп КС, що запитували різні доменні імена, здійснюється шляхом аналізу векторів ознак для пар групових DNS-запитів, побудованих на основі матриці мір Браун-Бланке [2], для яких коефіцієнт Браун-Бланке перевищує прийняте порогове значення. Вектор ознак складається з п'яти елементів: коефіцієнт Браун-Бланке та зведені поведінкові ознаки для двох порівнюваних груп, які можуть приймати наступні значення: "Unusual" (непритаманна ботам), "Neutral" (властива як користувачам, так і ботам), "Suspicious" (підозріла), "Dangerous" (властива ботам). Вектор ознак може бути визначений наступним чином:

$$\overline{W_{G_1,G_2}} = \left(K_B\left(G_1, G_2\right), S_{G_1,G_2}, F_{G_1,G_2}, R_{G_1,G_2}, M_{G_1,G_2}\right), \qquad (2)$$

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 G_1 та G_2; S_{G_1,G_2} – ознака звертання до локальних / нелокальних DNS-серверів; F_{G_1,G_2} – ознака повторного запиту в межах TTL-періоду; R_{G_1,G_2} – ознака наявності в DNS-відповідях коду помилки NXDOMAIN; M_{G_1,G_2} – ознака "інфікований" чи "підозрілий" щодо групи КС, отримана на проміжних етапах аналізу [2].

Функція аналізу векторів ознак на основі множини правил [2] може приймати чотири значення: "Not_Infected" (неінфіковані), "Not_Suspicious" (не підозрілі), "Suspicious" (підозрілі), "Infected" (інфіковані).

1.2. Метод виявлення бот-мереж, які застосовують технології ухилення від виявлення на основі DNS

З метою ухилення від виявлення бот-мережами використовується ряд технологій на основі DNS: технології «швидкозмінних» мереж (fast flux service networks), «потік доменів» («domain flux»), періодична зміна IP-відображення для шкідливого домена (cycling of IP mapping) та DNS-тунелювання (DNS-tunneling). З метою виявлення бот-мереж, які застосовують технології ухилення на основі DNS, розроблено новий метод [3], який враховує та усуває недоліки відомих методів, спрямованих на виявлення таких бот-мереж. Метод використовує два підходи: пасивний моніторинг вхідного DNS-трафіка та активне DNS-зондування.

На основі даних, зібраних шляхом застосування пасивного моніторингу DNS-трафіка, з врахуванням значень полів TTL, опрацьовуються такі вхідні DNS-повідомлення:

1) кожне перше зафіксоване DNS-повідомлення щодо певного доменного імені в межах TTL-періоду;

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-періоду, в межах якого було отримане це повідомлення.

З метою виявлення застосування технологій ухилення від виявлення на основі DNS використано наступні ознаки:

1) довжина запитаного доменного імені, L_N;

2) кількість унікальних символів в доменному імені, N_U;

3) ентропія доменного імені, E_N;

4) ознака використання рідковживаних типів записів DNS (KEY, NULL тощо), або таких, які зазвичай не використовуються клієнтами (наприклад, TXT), F_{UR};

5) максимальне значення ентропії записів DNS, які містяться в DNS-повідомленнях (CNAME, TXT, NS, MX, KEY, NULL тощо), E_R;

6) максимальна довжина DNS-повідомлення, L_P;

7) кількість A-записів, що відповідають доменному імені, у вхідному DNS-повідомленні, N_A;

8) кількість IP-адрес, пов'язаних з доменним ім'ям, N_{IP};

9) середня дистанція між IP-адресами, пов'язаними з доменним ім'ям, S_{IP};

10) середня дистанція між IP-адресами в A-записах, що відповідають доменному імені, у вхідному DNS-повідомленні, S_A;

11) кількість унікальних IP-адрес в множинах A-записів, що відповідають доменному імені, N_{UA};

12) середня дистанція між унікальними IP-адресами в множинах A-записів, що відповідають доменному імені, S_{UA};

13) кількість доменних імен, які спільно використовують IP-адресу, що відповідає доменному імені, N_D;

14) значення TTL-періоду: мода, T_{mod}; медіана, T_{med}; середнє арифметичне значення, T_{aver};

15) ознака успішності DNS-запиту, F_S.

З перерахованого набору ознак протягом певного інтервалу часу пасивного моніторингу формуються вектори ознак $\overline{W_d}$ щодо кожного запитаного КС мережі доменного імені, які в подальшому подаються на входи нечіткого кластеризатора.

Часткове навчання кластеризатора здійснюється на основі знань щодо ознак, які можуть бути отримані з вхідних DNS-повідомлень до ботів бот-мереж, що використовують такі технології ухилення. Знання можуть бути представлені у вигляді правил, описаних алгоритмом, поданим на рис. 2.

Результатом здійснення нечіткої кластеризації с-середніх з частковим навчанням 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:

$$U = \left[u_{ij}\right], u_{ij} \in [0,1], i = \overline{1,N_z}, j = \overline{1,N_h}, \sum_{j=\overline{1,N_h}} u_{ij} = 1, \quad (3)$$

де N_z – загальна кількість різних доменних імен, запитаних КС мережі; $N_h = 5$ – кількість кластерів, чотири з яких відповідають певній технології ухи-

лення (cycling of IP mapping, «domain flux», «fast flux», DNS-tunneling та кластер, який містить нормальні запити).

```
for all DNS_messages_or_training_data do
    if (T_mod ∈ [0,900] and T_med ∈ [0,900] and T_aver ∈ [0,900]) then
        if ((N_A ∈ (5,∞) and S_A ∈ (65535,∞)) or (N_UA ∈
        (8,∞) and S_UA ∈ (65535,∞)) then
            evasion_technique ← fast_flux
        end
        if (F_S = 0 and N_D ∈ [8,∞]) then
            evasion_technique ← domain_flux
        end
        if (N_IP ∈ (5,∞) and S_IP ∈ (65535,∞)) then
            evasion_technique ← cycling_of_IP_mappings
        end
        if ((L_N ∈ [75,255] and N_U ∈ (27,37]) or E_N ≥ f_Eb32 or (E_R ≥
        f_Eb64 or E_R ≥ f_Eb256) or F_UR = 1 or L_P > 300)) then
            evasion_technique ← DNS_tunneling
        end
    end
end
```

Рис. 2. Правила, на основі яких здійснюється часткове навчання кластеризатора

У випадках, коли об'єкт кластеризації – вектор ознак віднесено до декількох кластерів таким чином, що неможливо дійти однозначного висновку стосовно шкідливості чи легітимності доменного імені, має місце невизначеність результатів.

З метою усунення невизначеності частини результатів доцільним є залучення засобів активного DNS-зондування, а саме: здійснення запитів NS-записів, A-записів, SOA-записів, PTR-записів щодо підозрілих доменних імен. Це надає можливість залучити наступні додаткові ознаки:

1) кількість NS-записів у DNS-відповіді, N_{NS} ;

2) середня дистанція між IP-адресами для множини NS-записів щодо доменного імені, S_{NS} ;

3) кількість різних номерів автономних систем, до яких належать IP-адреси, пов'язані з серверами імен, N_{ASN} ;

4) кількість різних номерів автономних систем (ASN), до яких належать IP-адреси, пов'язані з доменним ім'ям, N_{ASA} ;

5) значення поля retry, отримане у DNS-відповіді на SOA-запит, V_{retry} .

Алгоритм виявлення бот-мереж на базі пасивного моніторингу DNS-трафіка та активного DNS-зондування подано на рис. 3. Локалізація КС, інфікованих ботами, здійснюється за допомогою ведення файлів журналювання, в яких зберігаються MAC-адреси КС мережі, що здійснювали DNS-запити, та запитані ними доменні імена.

Висновки

Розроблено нову інформаційну технологію виявлення бот-мереж на основі аналізу DNS-трафіка, яка дозволяє виконувати наступні задачі: виявлення інфікованих ботами КС мережі на основі їх групової активності в DNS-трафіку; виявлення КС мережі, інфікованих ботами бот-мереж, які застосовують технології ухилення від виявлення на основі DNS; локалізація інфікованих КС мережі. Застосування розробленої інформаційної технології дозволяє підвищити достовірність процесу виявлення бот-мереж в порівнянні з існуючими інформаційними технологіями і виявляти відомі та невідомі бот-мережі.

```
Function passive_analysis
for all gathered_incoming_DNS_messages do
    for all first_DNS_messages_within_TTL or
    all repeated_DNS_messages_within_TTL_from_non_local_DNS_serv
    do
        form feature_vector_W_d
    end
    form data_matrix_of_feature_vectors_V
    data_matrix_of_feature_vectors_V → set_of_clusters_H where
    V(i,) = W_d
    form fuzzy_splitting_matrix_U
    if (u_ij ≥ λ) then
        block
        else
            if (λ' ≤ u_ij < λ) then
                execute active_analysis
            end
    end
end
Function active_analysis
for all W_d do
    if (λ' ≤ u_ij < λ) then
        if ((W_d ∈ H_cycling_of_IP_mapping) and (W_d ∈ H_legitimate)) then
            if ((T_mod ∈ [0,900] and T_med ∈ [0,900] and T_aver ∈
            [0,900]) and (N_IP ∈ (5,∞) and S_IP ∈ (65535,∞) and
            N_ASA > 2)) then
                block
            end
        end
        if ((W_d ∈ H_fast_flux) and (W_d ∈ H_legitimate)) then
            if ((T_mod ∈ [0,900] and T_med ∈ [0,900] and T_aver ∈
            [0,900]) and ((N_A ∈ (5,∞) and S_A ∈ (65535,∞)) or (N_UA ∈
            (8,∞) and (S_UA ∈ (65535,∞)) or N_AS > 2) and (S_NS >
            65535 or N_ASN > 2 and N_NS > 3 and V_retry ∈ [0,900]))) then
                block
            end
        end
        if ((W_d ∈ H_domain_flux) and (W_d ∈ H_legitimate)) then
            if (N_D ∈ [8,∞]) then
                block
            end
        end
        else
            block
    end
end
```

Рис. 3. Алгоритм виявлення бот-мереж, які застосовують технології ухилення від виявлення на основі DNS

Литература

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. – 11.03.2016.

2. Technique for the Botnet Detection Based on DNS-Traffic Analysis [Text] / O. Pomorova, O. Savenko, S. Lysenko, A. Kryshchuk, K. Bobrovnikova. //

Proceedings of the CN 2015. – Brunow, Poland, 16-19 June 2015. – C. 127-138.

3. DNS-based Anti-evasion Technique for Botnets Detection [Text] / S. Lysenko, O. Pomorova, O. Savenko, A. Kryshchuk, K. Bobrovnikova. // Proceedings of the IDAACS 2015. – Warsaw, Poland, 24-26 September 2015. – T. 1. – C. 453-458.

References

1. *DAMBALLA: Botnet Detection for Communications Service Providers.* Available at: https://www.damballa.com/downloads/r_pubs/WP_Botnet_Detection_for_CSPs.pdf (accessed 11.03.2016)

2. Pomorova, O., Savenko, O., Lysenko, S., Kryshchuk, A., Bobrovnikova, K. A Technique for the Botnet Detection Based on DNS-Traffic Analysis. *Proceedings of the CN 2015,* 16-19 June 2015, Brunow, Poland, pp. 127-138.

3. Lysenko, S., Pomorova, O., Savenko, O., Kryshchuk, A., Bobrovnikova, K. DNS-based Anti-evasion Technique for Botnets Detection. *Proceedings of the IDAACS 2015,* 24-26 September 2015, Warsaw, Poland, t.1, pp. 453-458.

Поступила в редакцию 12.03.2016, рассмотрена на редколлегии 14.04.2016

ИНФОРМАЦИОННАЯ ТЕХНОЛОГИЯ ОБНАРУЖЕНИЯ БОТ-СЕТЕЙ НА ОСНОВЕ АНАЛИЗА DNS-ТРАФИКА

О. С. Савенко, С. Н. Лысенко, К. Ю. Бобровникова

Разработана информационная технология обнаружения бот-сетей на основе анализа DNS-трафика, которая построена на базе двух новых методов: метода выявления бот-сетей на основе их групповой активности в DNS-трафике и метода выявления бот-сетей, которые применяют технологии уклонения от обнаружения на основе DNS. Преимуществами информационной технологии является то, что процесс диагностики не требует построения баз сигнатур, предоставляет возможность выявления уже на начальной стадии распространения инфекции в сети, применяем как для маленьких, так и для больших сетей, не требует значительных объемов вычислительных ресурсов для обработки данных и позволяет обнаруживать известные и неизвестные бот-сети.

Ключевые слова: бот-сеть, DNS-трафик, групповая активность, технологии уклонения бот-сетей, «быстросменные» сети, «поток доменов», DNS-туннелирование

THE INFORMATION TECHNOLOGY FOR BOTNET DETECTION BASED ON DNS-TRAFFIC ANALYSIS

O. S. Savenko, S. M. Lysenko, K. Yu. Bobrovnikova

The new information technology for botnet detection based on the analysis of DNS traffic was developed. It is based on two methods: the method of botnets detection based on their group activity in DNS traffic and the method for botnets detection that use evasion techniques based on DNS. The advantages of the developed information technology is that the diagnostic process does not require the construction of signatures databases; it provides the ability to detection at an early stage of infection propagation in the network; it can be used for both small and large networks, does not require significant amounts of computing resources for data processing and allows to detect known and unknown botnets.

Keywords: botnet, DNS-traffic, group activity, botnets` evasion techniques, fast flux service networks, «domain flux», DNS-tunneling

Савенко Олег Станіславович – канд. техн. наук, доц., декан факультету програмування та комп'ютерних і телекомунікаційних систем, Хмельницький національний університет, Хмельницький, Україна, e-mail: savenko_oleg_st@ukr.net.

Лисенко Сергій Миколайович – канд. техн. наук, доц, Хмельницький національний університет, Хмельницький, Україна, e-mail: sirogyk@ukr.net.

Бобровнікова Кіра Юліївна – аспірант, Хмельницький національний університет, Хмельницький, Україна, e-mail: bobrovnikova.kira@gmail.com.

Savenko Oleg Stanislavovych – candidate of science, associate professor, dean of the faculty of programming and computer and telecommunications systems, Khmelnitsky national university, Khmelnitsky, Ukraine, e-mail: savenko_oleg_st@ukr.net.

Lysenko Sergiy Mykolayovych – candidate of science, associate professor, Khmelnitsky national university, Khmelnitsky, Ukraine, e-mail: sirogyk@ukr.net.

Bobrovnikova Kira Yuliivna – post-graduate student, Khmelnitsky national university, Khmelnitsky, Ukraine, e-mail: bobrovnikova.kira@gmail.com.