команди $e_\phi ... e_\eta u_\vartheta ... u_\sigma$ можуть бути різними для кожного нового запуску вірусу.

На наступному етапі проводиться збір векторів ознак зразків підозрілих кодів з мережі.

Висновок

Запропоновано модель інформаційної технології виявлення метаморфних вірусів на основі порівняння векторів ознак підозрілого файлу з його копією після емуляції виконання, та формування нечіткого логічного висновку на основі моделей поведінки метаморфних вірусів. Запропонована модель є основою реалізації методу виявлення нових метаморфних віусів або копій вже існуючих.

Література

1. Ször P. Striking Similarities: Win32/Simile and Metamorphic Virus Code, white paper. – Symantec Security Response, 2003.

2. Rad B.B. Metamorphic Virus Variants Classification Using Opcode Frequency Histogram / Rad B. B., Masrom M. – Proc. ICCOMP 10 the 14th WSEAS international conference of computers, 2011. – pp. 147-155.

3. Zhang Q. MetaAware: Identifying Metamorphic Malware / Q. Zhang, Douglas S. Reeve – Proc. Twenty Third Annual IEEE Conference on Computer Security Applications, 2007. – pp. 411-420

4. Bilar D. Statistical structure: fingerprinting malware for classification and analysis / D. Bilar – Proceding of black hat, 2008.

5. Kaspersky Lab What a metamorphic virus – Definition [Електронний ресурс] режим доступу: http://usa.kaspersky.com/internet-security-center/definitions/metamorphic-virus#.VkO9VdLhDIU

6. E. Al Daoud Computer virus strategies and detection methods / E. Al Daoud, I. H. Jebril, B. Qaibeh Open Problems compt., math., vol. 1 No.2 September 2008

7. Лисенко С. М. Метод виявлення поліморфного коду ботів ботнет-мереж / С. М. Лисенко, О. С. Савенко, А. О. Нічепорук // Радіоелектрон. і комп'ют. системи. - 2014. - № 5. - С. 129-134.

8. E. Al Daoud et al Detecting Metamorphic viruses by using Arbitrary Length of Control Flow Graphs and Nodes Alignment / Daoud E. Al. – UbiCC Journal, Vol. 4, No 3, pp.628–633, 2009.

9. A.H. Sung / Static analyzer of vicious executables (SAVE) A.H. Sung Proc. 20th Annual Computer Security Applications Conference, 2004.

Рецензія/Peer review : 4.11.2015 р. Надрукована/Printed :5.12.2015 р.
Рецензент: д.т.н., проф. Мартинюк В.В.

УДК 004.491.2

К.Ю. БОБРОВНІКОВА
Хмельницький національний університет

МОДЕЛЬ ІНФОРМАЦІЙНОЇ ТЕХНОЛОГІЇ ВИЯВЛЕННЯ БОТ-МЕРЕЖ НА ОСНОВІ АНАЛІЗУ DNS-ТРАФІКА

В роботі запропоновано модель інформаційної технології виявлення бот-мереж, яка ґрунтується на аналізі ознак, вилучених з корисного навантаження DNS-повідомлень. Модель заснована на властивості групової активності ботів в DNS-трафіку, враховує особливості поведінки груп хостів, властиві бот-мережам, та застосовує кластерний аналіз векторів ознак, які вказують на використання бот-мережами технологій ухилення від виявлення на основі DNS.

Ключові слова: бот-мережа, DNS-трафік, групова активність в DNS-трафіку, технології ухилення бот-мереж

K.Y.BOBROVNIKOVA
Khmelnitsky National University, Khmelnitsky, Ukraine

THE MODEL OF INFORMATION TECHNOLOGY FOR BOTNETS DETECTION BASED ON DNS-TRAFFIC ANALYSIS

Abstract - The model of information technology for botnets detection that based on an analysis of the features obtained from the payload of DNS-messages was proposed. The model is based on the property of bots group activity in the DNS-traffic, takes into account abnormal behaviors of the hosts' group, which are similar to botnets, and uses a cluster analysis of the feature vectors, which indicate the botnet's usage the DNS-based evasion techniques.

Keywords: botnet, DNS-traffic, group activity in DNS-traffic, botnet`s evasion techniques

Вступ

Бот-мережі є одним з найбільш небезпечних видів шкідливого програмного забезпечення. Щороку по всьому світу бот-мережами інфікується близько 500 млн. персональних комп'ютерів, кожну секунду – близько 18 [1]. Бот-мережі використовуються для здійснення DDoS-атак, поширення шкідливого

програмного забезпечення, викрадення конфіденційних даних, організації анонімних проксі-серверів, з метою здійснення корпоративного шпіонажу, фішингу, поширення спаму, надання сервісу віддалених машин тощо.

Методи виявлення бот-мереж, засновані на фільтрації пакетів, аналізі трафіка на основі портів та відомих сигнатур обходяться зловмисником шляхом динамічної зміни шкідливого коду, системи керування та портів або використанням портів HTTP/S.

Переважна більшість бот-мереж для керування інфікованими хостами використовує DNS. На відміну від традиційних методів виявлення, методи на основі DNS не вимагають значних обсягів обчислювальних ресурсів та здатні виявляти ще невідомі боти.

Постановка задачі

DNS-запити ботів, що входять до складу мережі, вирізняються характерною ознакою – груповою активністю, тобто скоординованістю в DNS-трафіку, що полягає у здійсненні ними одночасних або зосереджених в невеликому проміжку часу DNS-запитах. Відомі методи, засновані на властивості групової активності в DNS-трафіку, описані в [2-4], мають наступні недоліки: потреба у значному часі обробки та великих обсягах обчислювальних ресурсів при застосуванні до великих мереж; спирання на групові запити лише однакових доменних імен або недостатня гнучкість механізму виявлення міграцій C&C-серверів та активності груп ботів; довільний поділ періоду моніторингу на інтервали, в межах яких здійснюється пошук інфікованих груп; коротка тривалість періоду моніторингу; необхідність додаткового збору мережного трафіка.

Існує ряд технологій ухилення від виявлення бот-мереж, які засновані на використанні DNS: періодична зміна IP-відображення для шкідливого домена (cycling of IP mapping), технологія «потік доменів» (domain flux), технологія «швидкозмінних» мереж (fast-flux service network) та DNS-тунелювання (DNS-tunneling). Застосування DNS-тунелювання дозволяє зловмиснику використовувати протокол DNS в якості носія для трафіку командування та контролю бот-мережею. Технологія «швидкозмінних» мереж полягає в поєднанні використання дуже коротких TTL-періодів та циклічного методу розподілення навантаження і забезпечення відмовостійкості round-robin DNS. Технології «потік доменів» та періодична зміна IP-відображення застосовують поєднання коротких TTL-періодів та частих змін доменного імені C&C-сервера та IP-адрес для доменного імені C&C-сервера відповідно. Використання цих технологій ускладнює виявлення та блокування C&C-серверів бот-мереж, проте може бути виявлене шляхом аналізу ознак, вилучених з корисного навантаження DNS-повідомлень. Основними недоліками методів, описаних в [5-7], які спрямовані на виявлення застосування технологій ухилення на основі DNS, є необхідність активного DNS-зондування, і тому неможливість реалізації на основі пасивного аналізу DNS-трафіка; необхідність залучення інформації, отриманої від інших сервісів (WHOIS тощо); зосередження на виявленні вузького кола шкідливого ПЗ.

Крім того, для багатьох видів бот-мереж властиві певні особливості поведінки ботів, які простежуються в DNS-трафіку та є нетиповими для DNS-запитів звичайних користувачів. Зазвичай неінфіковані хости в локальній мережі використовують локальні DNS-сервери для здійснення DNS-запитів; боти в локальній мережі можуть використовувати або локальні, або власні DNS-сервери чи безкоштовні сервіси DNS (OpenDNS, FreeDNS). Для багатьох видів бот-мереж характерним є ігнорування ботами TTL-періоду, який містився у відповіді від авторитативного DNS-сервера на DNS-запит. Це означає, що бот виконує очищення локального кеша DNS та здійснює повторний DNS-запит щодо доменного імені до завершення TTL-періоду, що надає можливість підвищити гнучкість та надійність керування бот-мережею. Для бот-мереж характерною є підвищена кількість DNS-відповідей з кодом помилки RCODE = 3 (NXDOMAIN, доменне ім'я не існує), пов'язаних з частою зміною доменного імені C&C-сервера бот-мережі або проблемами з його доступністю.

Тому метою роботи є розробка моделі інформаційної технології виявлення бот-мереж, яка грунтується на аналізі DNS-трафіка та усуває недоліки існуючих підходів.

Основна частина

Подамо модель інформаційної технології виявлення бот-мереж на основі аналізу DNS-трафіка наступним чином (рис.1):

$$M_D = \langle C_{TG}, C_{WB}, C_{GA}, C_{ET}, C_L \rangle, \quad (1)$$

де C_{TG} – процес збору вхідного DNS-трафіка; C_{WB} – процес співставлення запитаних доменних імен з «білим» та «чорним» списками; C_{GA} – процес виявлення групової активності в DNS-трафіку; C_{ET} – процес виявлення застосування технологій ухилення від виявлення бот-мереж на основі використання DNS; C_L – процес локалізації хостів, інфікованих ботами, та блокування дій ботів.

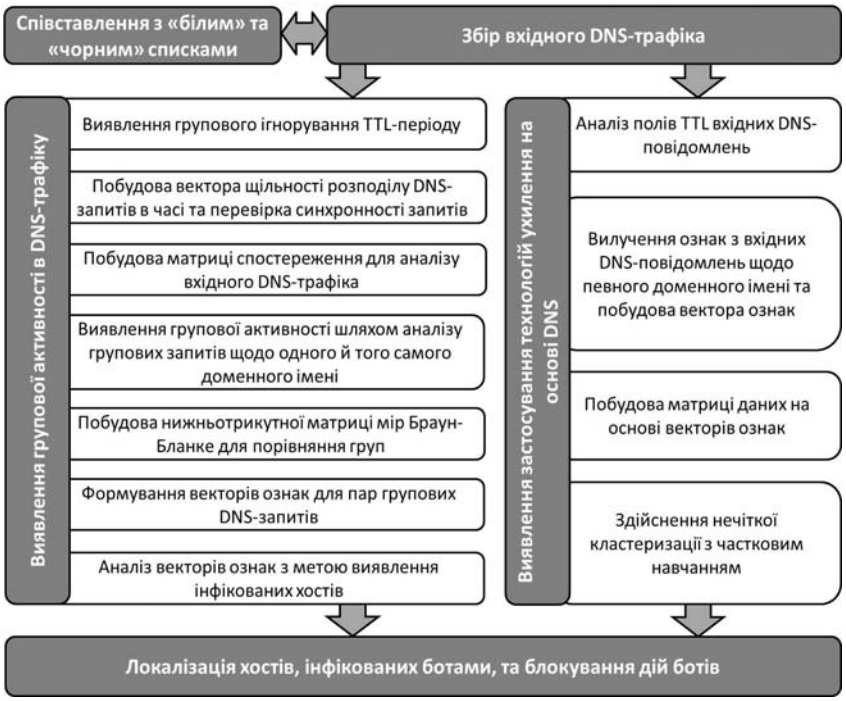

Рис.1. Узагальнена схема моделі інформаційної технології виявлення бот-мереж

Процес збору вхідного DNS-трафіка визначимо кортежем:

$$C_{TG} = \langle H, E, \chi \rangle, \tag{2}$$

де H – множина об'єктів діагностування – хостів мережі, які здійснювали DNS-запити протягом часу моніторингу; E – множина мережних давачів, налаштованих для збору вхідного DNS-трафіка; χ – множина захоплених вхідних DNS-повідомлень (DNS-відгуків).

Позначимо множину запитаних хостами мережі доменних імен як $D=\{d_i\}_{i=1}^{N_D}$, де N_D – кількість різних доменних імен. Представимо множину хостів мережі, які здійснювали DNS-запити протягом часу моніторингу, наступним чином: $H = \bigcup_{j=d_1}^{d_{N_D}} \bigcup_{k=1}^{N_T} H_{j,k}$, де H_j – підмножини MAC-адрес хостів, які надсилали DNS-запити щодо певного доменного імені протягом часу моніторингу; $H_{j,k}$ – підмножини MAC-адрес хостів, які надсилали DNS-запити щодо певного доменного імені в межах певного TTL-періоду; N_T – загальна кількість таких підмножин; $H_{j,k} = \{h_{j,k,i}\}_{i=1}^{N_{H,j,k}}$, де $h_{j,k,i}$ – MAC-адреса певного хоста мережі; $N_{H,j,k}$ – кількість хостів мережі, які надсилали DNS-запити в межах певного TTL-періоду. Аналогічно множину захоплених вхідних DNS-повідомлень представимо як $\chi = \bigcup_{j=d_1}^{d_{N_D}} \bigcup_{k=1}^{N_T} \chi_{j,k}$, де χ_j – підмножини вхідних DNS-повідомлень щодо певного доменного імені, захоплені протягом часу моніторингу; $\chi_{j,k}$ – 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.

З врахуванням полів вхідного DNS-повідомлення, дані з яких можуть бути використані для виявлення DNS-запитів бот-мереж, згідно стандарту RFC 1035 [8] опишемо захоплений DNS-відгук щодо певного доменного імені кортежем:

$$\chi_{j,k,i} = \left\langle \chi_{j,k,i,\text{H}}, \chi_{j,k,i,TS}, \chi_{j,k,i,IP}, \left\langle \chi_{j,k,i,HD}, \chi_{j,k,i,ANS}, \chi_{j,k,i,ATH}, \chi_{j,k,i,ADD} \right\rangle \right\rangle, \tag{3}$$
$$j = d_1, \ldots, d_{N_D}, k = \overline{1, N_T}, i = \overline{1, N_{\chi,j,k}},$$

де $\chi_{j,k,i,\text{H}}$ – MAC-адреса хоста, який здійснював DNS-запит; $\chi_{j,k,i,TS}$ – часовий штамп (час

надходження DNS-пакета); $\chi_{j,k,i,IP}$ – IP-адреса джерела DNS-пакета; $\chi_{j,k,i,HD}, \chi_{j,k,i,ANS}, \chi_{j,k,i,ATH}, \chi_{j,k,i,ADD}$ – секції DNS-повідомлення: заголовок (Header), секція відгуків (Answer), секція серверів імен (Authority) та секція додаткової інформації (Additional) відповідно.

Заголовок DNS-повідомлення може бути описаний наступним чином:

$$\chi_{j,k,i,HD} = \left\langle \chi_{j,k,i,HD,ID}, \chi_{j,k,i,HD,OPC}, \chi_{j,k,i,HD,RC}, \chi_{j,k,i,HD,QDC}, \chi_{j,k,i,HD,ANC}, \chi_{j,k,i,HD,NSC}, \chi_{j,k,i,HD,ARC} \right\rangle, \quad (4)$$
$$j = d_1,...,d_{N_D}, k = \overline{1,N_T}, i = \overline{1,N_{\chi,j,k}},$$

де $\chi_{j,k,i,HD,ID}$ – ідентифікатор, що дозволяє пов'язати запит та відгук (поле ID); $\chi_{j,k,i,HD,OPC}$ – тип запиту (поле OPCODE), $\chi_{j,k,i,HD,OPC} \in \{0,..,2\}$, 0 – стандартний, 1 – інверсний, 2 – запит стану сервера; $\chi_{j,k,i,HD,RC}$ – код відгуку (поле RCODE), $\chi_{j,k,i,HD,RC} \in \{0,..,5\}$, 0 – немає помилки, 1 – помилка в форматі запиту, 2 – збій сервера, 3 – доменне ім'я не існує тощо; $\chi_{j,k,i,HD,QDC}$ – кількість записів в секції запитів (поле QDCOUNT); $\chi_{j,k,i,HD,ANC}, \chi_{j,k,i,HD,NSC}, \chi_{j,k,i,HD,ARC}$ – кількість ресурсних записів в секціях відгуків, серверів імен та додаткової інформації (поля ANCOUNT, NSCOUNT, ARCOUNT) відповідно.

Секції відгуків, серверів імен та додаткової інформації мають однаковий формат та можуть бути описані як множини ресурсних записів наступним чином:

$$\chi_{j,k,i,S} = \left\{ \left(\chi_{j,k,i,S,NM}, \chi_{j,k,i,S,TP}, \chi_{j,k,i,S,TTL}, \chi_{j,k,i,S,RDL}, \chi_{j,k,i,S,RDT} \right)_n \right\}_{n=1}^{N_{RR,S}}, \quad (5)$$
$$j = d_1,...,d_{N_D}, k = \overline{1,N_T}, i = \overline{1,N_{\chi,j,k}},$$

де $S \in \{"ANS","ATH","ADD"\}$, $\chi_{j,k,i,S,NM}$ – ім'я домена, до якого відноситься ресурсний запис (поле NAME); $\chi_{j,k,i,S,TP}$ – тип коду ресурсного запису (поле TYPE), визначає значення та формат даних в полі RDATA [8]; $\chi_{j,k,i,S,TTL}$ – час життя записів DNS (поле TTL); $\chi_{j,k,i,S,RDL}$ – довжина поля RDATA (поле RDLENGTH); $\chi_{j,k,i,S,RDT}$ – рядок, що описує ресурс (поле RDATA); $N_{RR,S}$ – кількість ресурсних записів в секції (дорівнює значенню $\chi_{j,k,i,HD,ANC}, \chi_{j,k,i,HD,NSC}, \chi_{j,k,i,HD,ARC}$ для відповідної секції).

Процес співставлення запитаних доменних імен з «білим» та «чорним» списками визначимо кортежем:

$$C_{WB} = \langle D, W, B \rangle, \quad (6)$$

де W, B – множини доменних імен, занесених до «білого» списку популярних легітимних доменних імен та «чорного» списку відомих шкідливих доменних імен відповідно.

Процес виявлення групової активності в DNS-трафіку опишемо кортежем:

$$C_{GA} = \langle P_F, P_S, P_M, P_E, P_B, P_P, P_A \rangle, \quad (7)$$

де P_F – підпроцес виявлення групового ігнорування TTL-періоду; P_S – підпроцес побудови вектора щільності розподілу DNS-запитів в часі та перевірки синхронності запитів; P_M – підпроцес побудови матриці спостереження для аналізу вхідного DNS-трафіка; P_E – підпроцес виявлення групової активності шляхом аналізу групових запитів щодо одного й того самого доменного імені; P_B – підпроцес побудови нижньотрикутної матриці мір Браун-Бланке для порівняння груп; P_P – підпроцес формування векторів ознак для пар групових DNS-запитів; P_A – підпроцес аналізу векторів ознак з метою виявлення інфікованих хостів.

Підпроцес виявлення групового ігнорування TTL-періоду визначимо кортежем:

$$P_F = \langle \eta, \gamma \rangle, \quad (8)$$

де η – процедура побудови матриці спостереження V_{MAC}; γ – функція порівняння множин хостів для попереднього та повторного групових запитів в межах TTL-періоду.

Кожен рядок матриці спостереження V_{MAC} (рис.2) містить MAC-адреси хостів, які здійснювали DNS-запити щодо певного доменного імені в межах TTL-періоду. Якщо MAC-адреса хоста представлена в групі, то у відповідній комірці матриці відмічається «1», інакше – «0». Якщо хост повторно надсилав запит щодо доменного імені, то MAC-адреса хоста помічається «1» в рядку матриці, створеному для повторного запиту.

h_1	h_2	h_3	h_4	…	h_i
1	1	1	1	…	1
1	0	1	1	…	0

Рис.2. Матриця спостереження V_{MAC}

Процедуру побудови матриці спостереження V_{MAC} опишемо кортежем:

$$\eta = \left\langle D, \chi_{j,k}, \chi_{j,k,i,TS}, \chi_{j,k,i,ANS,TTL}, H_{j,k} \right\rangle, j = d_1,...,d_{N_D}, k = \overline{1, N_T}, i = \overline{1, N_{\chi,j,k}}. \quad (9)$$

Визначимо функцію γ порівняння двох множин хостів для попереднього та повторного групових запитів в межах TTL-періоду наступним чином: $\gamma : \{g_{1,i} \le g_{2,i}, K_B\}$, $K_B : \{K_B \in [0;1]; G_1, G_2 \to K_B\}$, де G_1 та G_2 – множини MAC-адрес хостів (групи), що здійснювали повторні DNS-запити в межах TTL-періоду, $G_1 \subseteq H_{j,k}, G_2 \subseteq H_{j,k}$, $g_{1,i} \in G_1, g_{2,i} \in G_2$, K_B – коефіцієнт подібності Браун-Бланке для двох порівнюваних груп; δ – порогове значення подібності. Якщо $\gamma < \delta$, то рядок матриці V_{MAC} для групового запиту $min(G_1, G_2)$ видаляється.

Підпроцес побудови вектора щільності розподілу DNS-запитів в часі та перевірки синхронності запитів визначимо кортежем:

$$P_S = \langle \theta, \varphi, \vartheta \rangle, \quad (10)$$

де θ – процедура побудови вектора щільності розподілу DNS-запитів в часі; φ – функція визначення синхронності DNS-запитів; ϑ – функція об'єднання множин MAC-адрес груп хостів з матриці V_{MAC}.

Групи запитів можна розглядати як синхронні, якщо спостерігається велика кількість запитів для доменного імені в межах певного інтервалу часу, коли боти бот-мережі здійснюють запити – часу синхронізації ботів t_S. З метою перевірки синхронності DNS-запитів використаємо вектор щільності розподілу DNS-запитів в часі. Процедуру побудови вектора щільності розподілу DNS-запитів в часі визначимо кортежем:

$$\theta = \left\langle \chi_{j,k}, \chi_{j,k,i,TS}, t_{first}, t_{last}, t_S \right\rangle, \; j = d_1,...,d_{N_D}, k = \overline{1, N_T}, i = \overline{1, N_{\chi,j,k}}, \quad (11)$$

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.

Вектор щільності розподілу DNS-запитів в часі опишемо наступним чином:

$$\overline{W_d} = (\Omega_j)_{j=1}^{z}, \quad (12)$$

де Ω_j – кількість DNS-запитів в межах z-го інтервалу; $z = (t_{last} - t_{first}) / \frac{1}{3} t_S$.

Функцію визначення синхронності DNS-запитів опишемо кортежем:

$$\varphi = \left\langle \overline{W_d}, \Omega_{max}, \Omega_{max\pm2}, Sum_s, Sum_r, \delta \right\rangle, \quad (13)$$

де Ω_{max} – елемент вектора $\overline{W_d}$ з максимальним значенням, max – його індекс; $\Omega_{max\pm2}$ – елементи вектора, суміжні з Ω_{max}; Sum_s – сума значень максимального та двох суміжних з ним елементів вектора з найбільшими значеннями, які описують розподіл DNS-запитів неперервного інтервалу часу; Sum_r – сума значень решти елементів вектора.

Визначимо функцію φ наступним чином: $\varphi(\overline{W_d}) = true$, якщо $(1-\delta) \cdot Sum_s > Sum_r$, інакше $\varphi(\overline{W_d}) = false$. Якщо $\varphi(\overline{W_d}) = true$, то груповий запит підлягає подальшому аналізу, інакше – відкидається. Такий поділ надає можливість мінімізувати кількість запитів, які не потрапили в інтервал t_S.

Якщо було виявлено повторні групові запити в межах TTL-періоду та $\varphi(\overline{W_d}) = true$ для цих DNS-запитів, множини MAC-адрес груп хостів, що здійснювали повторні запити, об'єднуються. Визначимо функцію об'єднання множин MAC-адрес груп хостів $\vartheta : \bigcup_{j=1}^{q} G_j \to G$, де q – кількість повторних запитів.

Для кожного визначеного інтервалу часу моніторингу t_m будується матриця спостереження M_m (рис. 3, а), де m – номер ітерації спостереження. Якщо було виявлено синхронність DNS-запитів, множини MAC-адрес груп хостів з матриці V_{MAC} переносяться до матриці спостереження M_m. Кожен рядок матриці містить доменне ім'я, запитане групою хостів, MAC-адреси групи хостів, що здійснювали DNS-запити щодо

цього доменного імені, та ознаки наявності для цих груп нетипових для звичайних користувачів особливостей поведінки, властивих для бот-мереж [9].

Представимо матрицю M_m у вигляді набору векторів:

$$\overline{W_{M,i}} = \left(d_i, G_i, N_{G,i}, S_i, F_i, R_i, M_i, N_i\right), i = \overline{1, N_{GQ}}, \quad (14)$$

де $N_{G,i}$ – кількість хостів у групі; S_i – ознака звертання до локальних/нелокальних DNS-серверів; F_i – ознака повторного запиту в межах TTL-періоду; R_i – ознака наявності в DNS-відповідях коду помилки NXDOMAIN; M_i – ознака "інфікований" чи "підозрілий" щодо групи хостів, отримана на проміжних етапах аналізу; N_i – номер ітерації спостереження, на якій зафіксовано ознаку "підозрілий" [9]; N_{GQ} – кількість групових DNS-запитів.

	h_1	h_2	…	h_k	N_0	S	F	R	M	N
d_3	1	1	…	1	5	1	1	0	0	0
d_1	1	0	…	1	4	1	0	0	0	0
d_2	1	0	…	1	6	0.5	0	0	0	0
…	…	…	…	…	…	…	…	…	…	…

а)

	d_3	d_5	d_1	d_2	d_4	…	N_0	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	0

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

б)

Рис.3. а) Матриця спостереження M_m; б) Матриця мір Браун-Бланке B_m та приклад побудови вектора ознак для двох групових DNS-запитів

Підпроцес побудови матриці спостереження M_m визначимо кортежем:

$$P_M = \left\langle G_i, \chi_j, R_b, \overline{W_{M,i}} \right\rangle, i = \overline{1, N_{GQ}}, j = d_1, \ldots, d_{N_D}, \quad (15)$$

де R_b – множина правил для заповнення комірок матриці M_m [9].

Підпроцес виявлення групової активності шляхом аналізу групових запитів щодо одного й того самого доменного імені опишемо кортежем:

$$P_E = \left\langle \overline{W_{M,i}}, \tau, \varpi, R_S, R_m \right\rangle, i = \overline{1, N_{GQ}}, \quad (16)$$

де τ – функція порівняння груп хостів з матриці спостереження M_m, що запитували однакові доменні імена, з використанням множини правил R_S [9]; ϖ – функція об'єднання рядків матриці M_m на основі множини правил R_m [9].

Визначимо функцію τ наступним чином: $\tau : G_1 \times \ldots \times G_n \to \{inf\ ected, suspicious, not\_suspicious\}$, $\tau(G_1, \ldots, G_n) = inf\ ected$, якщо $K_S(G_1, \ldots, G_n) \geq \delta$ або $\delta` \leq K_S(G_1, \ldots, G_n) < \delta$ та $\exists r = true, r \in R_S$; $\tau(G_1, \ldots, G_n) = suspicious$, якщо $\delta` \leq K_S(G_1, \ldots, G_n) < \delta$ та $\forall r = false, r \in R_S$; $\tau(G_1, \ldots, G_n) = not\_suspicious$, якщо $K_S(G_1, \ldots, G_n) < \delta`$, де $G_1, \ldots, G_n$ – групи хостів, що запитували однакові доменні імена; $K_S$ – коефіцієнт подібності (коефіцієнт подібності Браун-Бланке для порівняння двох груп або індекс дисперсності Коха для 3 і більше груп); $\delta`$ – порогове значення подібності, за якого групи хостів вважатимуться підозрілими. Якщо $\tau(G_1, \ldots, G_n) = not\_suspicious$, то групові запити для доменного імені видаляються з матриці спостереження M_m.

Визначимо функцію ϖ наступним чином: $\varpi : \bigcup_{j=1}^{n} \left\langle G_j, S_j, F_j, R_j \right\rangle \to \langle G, S, F, R \rangle$. Об'єднання значень комірок S, F, R матриці M_m здійснюється за множиною правил R_m [9].

На основі матриці спостереження M_m будується нижньотрикутна матриця мір Браун-Бланке B_m (рис. 3, б). Матриця B_m містить коефіцієнти Браун-Бланке, обчислені для пар груп хостів, що запитували різні доменні імена, та ознаки $N_{G,i}, S_i, F_i, R_i, M_i, N_i$, перенесені з матриці M_m. Матриця B_m формується за зростанням кількості MAC-адрес в групах $N_{G,i}$, по стовпчиках. Обчислення значень комірок для кожного стовпчика припиняється, якщо відношення розмірів порівнюваних груп стане меншим за порогове значення δ'.

Представимо матрицю B_m у вигляді набору векторів:

$$\overline{W_{B,i}} = \left(d_i, K_i, N_{G,i}, S_i, F_i, R_i, M_i, N_i\right), i = \overline{1, N_{GQ}}, \quad (17)$$

де K_i – множина коефіцієнтів Браун-Бланке, обчислених для пар груп хостів, $K_B \geq \delta', K_B \in K_i$.

Підпроцес формування векторів ознак для пар групових DNS-запитів на основі матриці мір Браун-Бланке визначимо кортежем:

$$P_P = \left\langle \overline{W_{B,i}}, R_p, \overline{W_{G_i,G_j}} \right\rangle, i = \overline{1, N_{GQ}}, j = \overline{1, N_{GQ}}, \quad (18)$$

де $\overline{W_{G_i,G_j}}$ – вектор ознак для пари групових DNS-запитів, для яких $K_B \geq \delta`$; $\overline{W_{G_i,G_j}} = \left(K_B\left(G_i,G_j\right), S_{G_i,G_j}, F_{G_i,G_j}, R_{G_i,G_j}, M_{G_i,G_j}\right)$, де $S_{G_i,G_j}, F_{G_i,G_j}, R_{G_i,G_j}, M_{G_i,G_j}$ – зведені поведінкові ознаки для двох груп [9], які можуть приймати наступні значення: "Unusual" (непритаманна ботам), "Neutral" (властива як звичайним користувачам, так і ботам), "Suspicious" (підозріла), "Dangerous" (небезпечна, притаманна ботам); R_p – множина правил для формування векторів ознак [9].

Висновок щодо наявності в мережі групової активності ботів здійснюється на основі аналізу векторів ознак $\overline{W_{G_i,G_j}}$ для пар групових DNS-запитів. Підпроцес аналізу векторів ознак визначимо кортежем:

$$P_A = \left\langle \overline{W_{G_i,G_j}}, f\left(\overline{W_{G_i,G_j}}\right) R_a \right\rangle, i = \overline{1, N_{GQ}}, j = \overline{1, N_{GQ}}, \quad (19)$$

де $f\left(\overline{W_{G_i,G_j}}\right)$ – функція аналізу вектора ознак $\overline{W_{G_i,G_j}}$ на основі множини правил R_a [9], яка може приймати чотири значення: "Not_Infected" (неінфіковані), "Not_Suspicious" (не підозрілі), "Suspicious" (підозрілі), "Infected"(інфіковані).

Одна й та сама група в межах ітерації спостереження може отримати декілька різних оцінок. В такому випадку пріоритет має оцінка з вищим ступенем небезпечності. Групи хостів, які було визначено як не інфіковані, відкидаються. Групи хостів 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 повторно разом з даними, що будуть зібрані на наступній ітерації спостереження (матриця спостереження M_{m+1}) з метою виявлення можливих повторних групових запитів в інтервалі часу t_{m+1}.

Процес виявлення застосування технологій ухилення від виявлення на основі DNS визначимо кортежем:

$$C_{ET} = \left\langle P_T, P_{SW}, P_{DM}, \wp \right\rangle, \quad (20)$$

де P_T – підпроцес аналізу полів TTL вхідних DNS-повідомлень; P_{SW} – підпроцес вилучення ознак, які можуть вказувати на застосування технологій ухилення від виявлення на основі DNS, з вхідних DNS-повідомлень щодо певного доменного імені та побудови вектора ознак; P_{DM} – підпроцес побудови матриці даних на основі векторів ознак для різних доменних імен; $\wp$ – функція здійснення нечіткої кластеризації c-середніх з частковим навчанням.

Підпроцес аналізу полів TTL вхідних DNS-повідомлень з метою відбору DNS-відгуків для подальшого аналізу визначимо кортежем:

$$P_T = \left\langle D, \chi_j, \chi_{j,k}, H_j, \chi_{j,k,i,TS}, \chi_{j,k,i,ANS,TTL} \right\rangle, j = d_1,...,d_{N_D}, k = \overline{1, N_T}, i = \overline{1, N_{\chi,j,k}}. \quad (21)$$

Підпроцес вилучення ознак з вхідних DNS-повідомлень щодо певного доменного імені та побудови вектора ознак визначимо кортежем:

$$P_{SW} = \left\langle D, \chi, \chi_j, \chi_{j,k}, T_{FF}, T_{CM}, T_{DF}, T_{DT}, \overline{W_{ET,j}} \right\rangle, j = d_1,...,d_{N_D}, k = \overline{1, N_T}, \quad (22)$$

де T_{FF}, T_{CM}, T_{DF}, T_{DT} – множини ознак, які можуть вказувати на застосування технологій «швидкозмінних» мереж, періодичної зміни IP-відображення для шкідливого домена, «потік доменів», DNS-тунелювання відповідно; $\overline{W_{ET,j}}$ – вектор ознак вхідних DNS-повідомлень щодо доменного імені.

Множину ознак, які можуть вказувати на застосування технології «швидкозмінних» мереж, визначимо кортежем:

$$T_{FF} = \left\langle T_{mod}, T_{med}, T_{aver}, N_A, S_A, N_{UA}, S_{UA} \right\rangle, \quad (23)$$

де T_{mod} – TTL-період, мода (значення у множині спостережень, яке зустрічається найбільш часто; застосуємо мінімальне значення у випадку, якщо множина є мультимодальною); T_{med} – TTL-період, медіана (значення ознаки, яке розділяє ранжовану сукупність на дві рівні частини, де 50% нижніх одиниць ряду даних мають значення ознаки, не більше, ніж медіана, а 50% верхніх – не менше, ніж медіана); T_{aver} – TTL-період, середнє арифметичне значення; N_A – кількість A-записів, що відповідають доменному імені, у вхідному DNS-повідомленні (ознака використовується, якщо $N_A>1$); S_A – середня дистанція між IP-

адресами в множині А-записів для доменного імені у вхідному DNS-повідомленні (ознака використовується, якщо $N_A>1$); N_{UA}– кількість унікальних IP-адрес в множинах А-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 А-записів, що відповідають доменному імені, у вхідних DNS-повідомленнях (ознака використовується, якщо $N_A>1$).

Множину ознак, які можуть вказувати на застосування періодичної зміни IP-відображення для шкідливого домена, визначимо кортежем:

$$T_{CM} = \langle T_{mod}, T_{med}, T_{aver}, N_{IP}, S_{IP} \rangle, \tag{24}$$

де 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 $N_A=1$).

Множину ознак, які можуть вказувати на застосування технології «потік доменів», визначимо кортежем:

$$T_{DF} = \langle T_{mod}, T_{med}, T_{aver}, F_S, N_N \rangle, \tag{25}$$

де F_S– бінарна ознака успішності DNS-запиту ($F_S = 0$, якщо DNS-запит невдалий, $F_S = 1$, якщо DNS-запит успішний); N_N – кількість доменних імен, які спільно використовують IP-адресу.

Множину ознак, які можуть вказувати на застосування DNS-тунелювання, визначимо кортежем:

$$T_{DT} = \langle L_N, N_U, E_N, E_R, F_{UR}, L_P, f_{E_B} \rangle, \tag{26}$$

де L_N – довжина доменного імені; N_U – кількість унікальних символів в доменному імені; E_N – ентропія доменного імені; E_R– максимальне значення ентропії ресурсних записів DNS, які містяться в DNS-повідомленнях; F_{UR} – бінарна ознака використання рідковживаних типів записів DNS, або таких, які зазвичай не використовуються клієнтами; L_P – максимальний розмір DNS-повідомлень щодо доменного імені; f_{E_B} – функція залежності ентропії поля DNS-повідомлення від його довжини [11].

Вектор ознак $\overline{W_{ET,j}}$ вхідних DNS-повідомлень щодо доменного імені подамо наступним чином:

$$\overline{W_{ET,j}} = (L_{N,j}, N_{U,j}, E_{N,j}, T_{mod,j}, T_{med,j}, T_{aver,j}, N_{A,j}, N_{IP,j}, S_{IP,j}, S_{A,j}, N_{UA,j}, S_{UA,j}, N_{N,j}, F_{UR,j}, E_{R,j}, L_{P,j}, F_{S,j}), \tag{27}$$
$$j = d_1, \ldots, d_{N_D}.$$

Підпроцес побудови матриці даних на основі векторів ознак визначимо кортежем:

$$P_{DM} = \langle \overline{W_{ET,j}}, V \rangle, \tag{28}$$

де $V = \left(v_{ji}\right)_{j=1,i=1}^{N_D, N_Q}$ – матриця даних, $V(j,) = \overline{W_{ET,j}}$, N_Q – загальна кількість ознак, які можуть вказувати на використання технологій ухилення від виявлення на основі DNS.

Функцію $\wp$ здійснення нечіткої кластеризації с-середніх з частковим навчанням опишемо кортежем:

$$\wp = \langle V, R_e, X, \Psi \rangle, \tag{29}$$

де R_e – множина знань щодо ознак вхідних DNS-повідомлень до ботів, які застосовують технології ухилення від виявлення на основі DNS [10]; $X = \{x_i\}_{i=1}^{N_x}$ – промаркована вибірка даних, яка формується на основі множини знань R_e, для здійснення часткового навчання кластеризатора, N_x – кількість об'єктів в промаркованій вибірці; $\Psi = \{\psi_i\}_{i=1}^{N_\Psi}$ – множина наперед визначених кластерів, N_Ψ – кількість кластерів.

Функція нечіткої кластеризації с-середніх з частковим навчанням може бути визначена як $\wp : V \to \Psi$. Результатом здійснення кластеризації з частковим навчанням є ступені приналежності векторів ознак до п'яти кластерів з множини Ψ. Належність вектора ознак до кластера ψ_1 свідчить про застосування технології ухилення «cycling of IP mapping», ψ_2 – «domain flux», ψ_3 – «fast flux», ψ_4 – «DNS-tunneling», ψ_5 – кластер, який містить нормальні запити.

Процес локалізації хостів, інфікованих ботами, та блокування дій ботів опишемо кортежем:

$$C_L = \left\langle \mathrm{H}, f\left(\overline{W_{G_i}, G_j}\right), \wp, \zeta \right\rangle, i = \overline{1, N_{GQ}}, j = \overline{1, N_{GQ}}, \tag{30}$$

де ζ – множина заходів, які застосовуються до хостів, визначених як інфіковані, з метою ліквідації інфекції (блокування, усунення уразливостей систем, встановлення (оновлення) антивірусного ПЗ тощо).

Висновки

Запропоновано модель інформаційної технології виявлення бот-мереж в мережах на основі аналізу DNS-трафіка. Модель заснована на властивості групової активності ботів в DNS-трафіку та застосовує кластерний аналіз векторів ознак, вилучених з корисного навантаження DNS-повідомлень, які вказують на використання бот-мережами технологій ухилення від виявлення на основі DNS. Також враховуються особливості поведінки груп хостів, властиві бот-мережам.

Запропонована модель є основою методу, який дозволить виявляти ще невідомі боти, а також здійснювати виявлення на початковій стадії поширення інфекції в мережі.

Література

1. The Federal Bureau of Investigation. Demarest, J. (2014, July 15). Statement Before the Senate Judiciary Committee, Subcommittee on Crime and Terrorism, Washington, D.C. Retrieved from http://www.fbi.gov/news/testimony/taking-down-botnets.

2. 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.

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

4. Botnet Detection Using Adaptive Neuro Fuzzy Inference System. Roshna, R.S, Vinodh, E.: International Journal of Engineering Research and Applications (IJERA), Vol. 3, Issue 2, March-April 2013. – pp.1440-1445.

5. EXPOSURE: Finding Malicious Domains Using Passive DNS Analysis. Bilge, L., Kirda, E., Kruegel, C., Balduzzi, M.: NDSS, 2011– pp. 1-17.

6. Detecting DNS Tunneling. Farnham, G., Atlasis, A.: SANS Institute InfoSec Reading Room, 2013. – pp. 1-32.

7. As the Net Churns: Fast-Flux Botnet Observations. Nazario, J., Holz, T.: In: Conference on Malicious and Unwanted Software (Malware'08), 2008. – pp. 24-31.

8. RFC-1035. Domain names – implementation and specification. Mockapetris, P.: ISI, 1987 (http://www.ietf.org/rfc/rfc1035.txt?number=1035)

9. Pomorova O. A Technique for the Botnet Detection Based on DNS-Traffic Analysis / Oksana Pomorova, Oleg Savenko, Sergii Lysenko, Andrii Kryshchuk and Kira Bobrovnikova // Computer Networks 22th International Conference, CN 2015, Brunow, Poland, June 16-19, 2015. Proceedings, pp.127-138.

10. Lysenko S. DNS-based Anti-evasion Technique for Botnets Detection / Sergii Lysenko, Oksana Pomorova, Oleg Savenko, Andrii Kryshchuk and Kira Bobrovnikova // Proceedings of the 2015 IEEEE 8th International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS), IDAAACS'2015, Warsaw, Poland, September 24-26, 2015, Vol.1, pp.453-458.

11. On Botnets that use DNS for Command and Control. Dietrich, C.J., Rossow, C., Freiling, F. C., Bos, H., van Steen, M., Pohlmann, N.: In: Proceedings of European Conference on Computer Network Defense, 2011 – pp. 9-16.

Рецензія/Peer review : 5.11.2015 р. Надрукована/Printed :6.12.2015 р.
Рецензент: д.т.н., проф. Мартинюк В.В.