УДК 004.056

DOI: 10.31673/2412-9070.2021.062630

А. Д. КОЖУХІВСЬКИЙ, доктор техн. наук, професор;
О. Ю. ІЛЬЇН, доктор техн. наук;
В. А. САВЧЕНКО, студент;
А. Г. ЗАХАРЖЕВСЬКИЙ, канд. техн. наук,
Державний університет телекомунікацій, Київ

ПРОГНОЗУВАННЯ ДИНАМІКИ ПІДОЗРІЛОЇ АКТИВНОСТІ В МЕРЕЖІ НА ОСНОВІ АНАЛІЗУ МЕРЕЖНОГО ТРАФІКУ

Розглянуто можливість завчасного прогнозування кібератак на основі аналізу підозрілої активності в мережі, що дасть додаткові можливості службам захисту інформації у протидії таким атакам. Загальна проблема полягає у виявленні часу початку такої атаки, оскільки параметри трафіку не мають різких змін. Використовуючи методи машинного навчання на основі аналізу подібних ситуацій у минулому, є можливість створення інтегрованої системи для трансформації великих обсягів загальнодоступних даних, аби передбачати поведінку зловмисників у мережі.

Ключові слова: кібербезпека; підозріла активність; прогнозування нестаціонарних процесів; машинне навчання.

ВСТУП

Легкість пошуку вразливостей мережі зробила DDoS-атаки доволі поширеними. Для ефективної протидії таким атакам потрібні два основних заходи:

1) діагностування нападу на найбільш ранніх стадіях;

2) відокремлення шкідливого трафіку від звичайного після виявлення.

Розуміючи, які запити користувачів є результатом DDoS-атаки, можна виконати відповідні налаштування для брандмауерів, маршрутизаторів або застосувати інші заходи безпеки в мережі. Параметри повільних DDoS-атак здебільшого залежать від поведінки конкретного користувача, і для ефективного пом'якшення така поведінка має бути детально спрогнозована. Отже, прогнозування поведінки користувача на основі параметрів трафіку стає основним кроком у боротьбі з повільними DDoS-вторгненнями.

Постановка проблеми. Особливістю повільної DDoS-атаки є використання вразливості в протоколі TCP, де переривання можуть бути спровоковані навмисно або ненавмисно через затримки в каналі зв'язку. Оскільки повільні DDoS-атаки не спричиняють різке збільшення трафіку, що призводить до миттєвої відмови сервера в обслуговуванні, то виявити момент початку атаки практично неможливо, а відокремлення шкідливого трафіку від звичайного є значною проблемою. Тому для виявлення та розпізнавання повільних DDoS-атак необхідно розробити інші підходи та методи. Основною проблемою виявлення повільних DDoS-атак є неможливість їм запобігти, оскільки процес визначення базується на вивченні наявного трафіку без можливості його прогнозування залежно від активності користувачів. Безперечно, прогнозування поведінки користувачів дало б змогу виявити аномальну поведінку та запобігти появі повільних DDoS-атак.

Огляд літератури. Існує значна кількість публікацій про виявлення повільних DDoS-атак. У статті [1] автори пропонують новий метод виявлення повільних HTTP-атак у хмарі. Вирішення дає можливість виявляти атаки Slow HTTP Header Attacks (Slowloris), Slow HTTP Body Attacks (RUDY) або Slow HTTP Read Attacks. Інша їх робота [2] пропонує нову модель класифікації для пом'якшення атак у хмарі. Водночас такі підходи не гарантують ефективного виявлення атак на ранніх стадіях їх розвитку. У публікації [3] автори репрезентували систему, здатну виявляти та пом'якшувати атаки в межах мережної інфраструктури. Цю роботу продовжує стаття [4], яка досліджує модель захисту побічного каналу. Основними ідентифікаційними параметрами в обох моделях є швидкість передавання пакетів і рівномірна відстань між пакетами, що унеможливлює попередження дій зловмисників.

У дослідженні [5] розглядається вибірка даних для створення різних розподілів класів для протидії ефектам дуже незбалансованих повільних наборів даних HTTP DoS. Водночас значної кількості семплів (автори використовують 1,89 млн копій атак) реально досягти досить складно. Автори в праці [6] розробляють систему на основі метрики для виявлення традиційних повільних атак, які можуть бути ефективними з обмеженими ресурсами на основі дослідження подібності та впровадження евклідової метрики. Цей підхід досить ефективний лише за наявності великої кількості таких зразків повільних атак, але за великої різноманітності він навряд чи буде ефективним. Видання [7] визначає параметри якості TCP-з'єднань, характерні для повільних HTTP-атак. Здобуті формули оцінюють імовірність і час переходу вебсервера в режим перевантаження. Незважаючи на детальне дослідження, виявлення таких атак базується на статистиці спостережень і не стосується прогнозування. Стаття [8] пропонує

алгоритм виявлення повільних DDoS-атак на основі шаблонів трафіку залежно від стану завантаження сервера. Водночас процес прийняття рішень не розглядається. У роботі [9] аналізуються різні сценарії та пропонується гібридна нейронна мережа для виявлення DDoS-атак. Але метод і загальна методика виявлення DDoS-атак низької інтенсивності не обговорюються.

У [10] автори досліджують інтервальне прогнозування на основі ймовірнісної нейронної мережі з динамічним оновленням параметра згладжування. Але проблема динаміки моделі залишається нерозв'язаною. У статті [11] подано новий метод виявлення RUDY DDoS-атаки на основі самоподібності мережного трафіку, але не враховано різноманітність навчальних зразків і процес отримання навчальної множини. Автори [12] представили систему виявлення HTP DTP-атак у хмарі, засновану на показниках інформаційної ентропії та випадкових деревах. Такий підхід досить ефективний, хоча і не вирішує питання прогнозування розвитку нападу. Отже, більшість праць, присвячених протидії повільним DDoS-атакам, не розглядають питання прогнозування поведінки користувачів і тому недостатньо ефективні для виявлення атак на ранніх стадіях.

Метою статті є формування системи виявлення повільних DDoS-атак на основі прогнозування підозрілої активності в мережі. Для успішного розв'язання виявленої проблеми потрібно побудувати модель і технологію прогнозування поведінки користувачів з огляду на історію їх взаємодії з сервером, а також запропонувати топологію розпізнавання повільних DDoS-атак.

ОСНОВНА ЧАСТИНА

Щоб здобути універсальний метод прогнозування підозрілої активності в мережі, доцільним вбачається підхід, заснований на поданні досліджуваного часового ряду виразом [13; 14]

$$X(t)=m(t)+\sum_{\nu} V_\nu \varphi_\nu(t), \tag{1}$$

де $m(t)$ — математичне сподівання процесу; $\varphi_\nu(t)$ — невипадкові (координатні) функції часу; V_ν — випадкові некорельовані між собою коефіцієнти ($M[V_\nu]=0$, $M[V_\nu, V_\mu]=0$, $\nu\neq\mu$).

Таке подання дає змогу застосовувати його для будь-якого набору даних зокрема даних із пропусками та незбалансованих за часом. При цьому розглядається деякий скалярний випадковий процес $X(t)$, заданий випадковою послідовністю $X(t_i)=X(i)$, $i=\overline{1,I}$ на дискретному ряді спостережень t_i.

З самого початку такий процес можна подати у вигляді

$$X(i)=m(i)+\sum_{\nu=1}^{i} V_\nu \varphi_\nu(i),\ i=\overline{1,I}, \tag{2}$$

де V_ν — випадковий коефіцієнт з характеристиками ($M[V_\nu]=0$, $M[V_\nu, V_\mu]=0$, $\nu\neq\mu$); $M[V_\nu^2]=D_\nu$; $\varphi_\nu(i)$ — невипадкова координатна функція, $\varphi_\nu(\nu)=1$, $\varphi_\nu(i)=0$ при $\nu>1$.

Визначення оптимальних значень прогнозу на наступних кроках здійснюється на основі таких рекурентних співвідношень:

$$V_1=\overset{\circ}{X}(1),\ V_i=\overset{\circ}{X}(i)-\sum_{\nu=1}^{i-1} V_\nu \varphi_\nu(i),\ i=\overline{2,I}; \tag{3}$$

$$D_1=D(1),\ D_i=D(i)-\sum_{\nu=1}^{i-1} D_\nu \varphi_\nu^2(i),\ i=\overline{2,I}; \tag{4}$$

$$\varphi_\nu(i)=\frac{1}{D_\nu}M\left[V_\nu \overset{\circ}{X}(i)\right],\ \nu=\overline{1,I},\ i=\overline{\nu,I}. \tag{5}$$

Формула (5) показує, що основним обмеженням, яке накладається на досліджуваний випадковий процес, є скінченність його дисперсії. Під час дослідження кібератак це, зазвичай, виконується, що і забезпечує універсальність методу. Отже, подання (2) здатне забезпечити розв'язання задачі прогнозу кібератаки.

Алгоритм роботи методу

Щоб дістати аналітичний апостеріорний випадковий процес на базі вибірки випадкових даних, потрібно визначити процес $X(t)$ у вигляді (2) на дискретному ряді точок t_i, $i=\overline{1,I}$. Передбачається, що у деякі моменти t_μ, $\mu=\overline{1,k}$, $k<I$, які збігаються з t_i для $i\le k$, стали відомі значення $x(\mu)$, $\mu=\overline{1,k}$, реалізації процесу $X(t)$. Необхідно здобути аналітичний опис прогнозу $X^{ps}(t)$, який виникає з апріорного $X(t)$ з урахуванням даних спостережень. Визначення прогнозованих значень здійснюється за таким алгоритмом:

1. Визначення значення $x(1)$ реалізації процесу, одержаної завдяки спостереженням. Для цього значення є справедливим (2), яке при $\mu = 1$ зводиться до

$$x(1)= m(1) + \nu_1. \tag{6}$$

Формула (6) конкретизує значення ν_1 випадкового коефіцієнта V_1, яке відповідає результату першого спостереження.

2. Визначення коефіцієнтів V_i, $i = \overline{1,I}$ подання (2), які дають змогу конкретизувати значення V_1, що зумовлює зміну щільності розподілу решти коефіцієнтів V_i, $i = \overline{2,I}$. Щоб дістати прогноз, припустимо, що коефіцієнти вихідного подання (2) попарно незалежні:

$$f_2(\nu_i,\nu_j)= f_1(\nu_i)f_1(\nu_j),\ \ i=\overline{1,I-1},\ \ j=\overline{i+1,I}. \tag{7}$$

3. Підставляючи одержане з (6) значення V_1 до формули (2), дістанемо вираз для прогнозованого процесу, який у момент $i = 1$ проходить через точку $x(1)$:

$$X^{(1)}(i)= m(i)+(x(1)-m(1))\varphi_1(i)+\sum_{\nu=2}^{i} V_\nu\varphi_\nu(i),\ \ i=\overline{1,I} \tag{8}$$

з математичним сподіванням

$$m^{(1)}(i)= m(i)+(x(1)-m(1))\varphi_1(i),\ \ i=\overline{1,I}. \tag{9}$$

4. Обчислення:

$$X^{(1)}(i)= m^{(1)}(i)+\sum_{\nu=2}^{i} V_\nu\varphi_\nu(i),\ \ i=\overline{1,I}. \tag{10}$$

Якщо на наступному етапі спостережень одержано значення $x(2)$ тієї самої реалізації процесу, то для цього значення є справедливим (10), де $x(2) = m^{(1)}(2) + \nu_2$.

5. Повторити операції для випадку $\mu = 1$, та здобути

$$m^{(2)}(i)= m^{(1)}(i)+\left(x(2)-m^{(1)}(2)\right)\varphi_2(i),\ \ i=\overline{1,I}. \tag{11}$$

$$X^{(2)}(i)= m^{(2)}(i)+\sum_{\nu=3}^{i} V_\nu\varphi_\nu(i),\ \ i=\overline{1,I}. \tag{12}$$

6. Повторити ітерації для довільного числа $k < 1$ моментів контролю за такими формулами:

$$m^{(0)}(i)= m(i),\ \ i=\overline{1,I},$$

$$m^{(k)}(i)= m^{(k-1)}(i)+\left(x(k)-m^{(k-1)}(1)\right)\varphi_k(i),\ \ i=\overline{1,I}; \tag{13}$$

$$X^{(k)}(i)= m^{(k)}(i)+\sum_{\nu=k+1}^{i} V_\nu\varphi_\nu(i),\ \ i=\overline{1,I}. \tag{14}$$

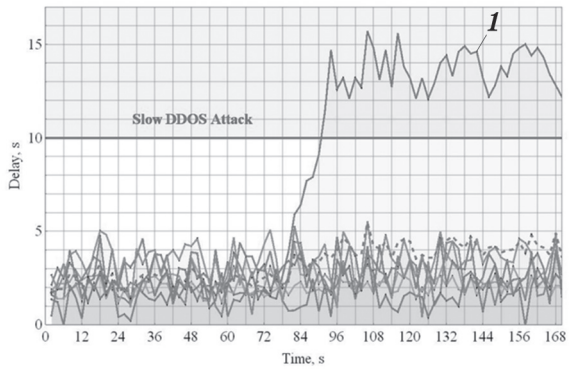

Рис. 1. Шаблони трафіку

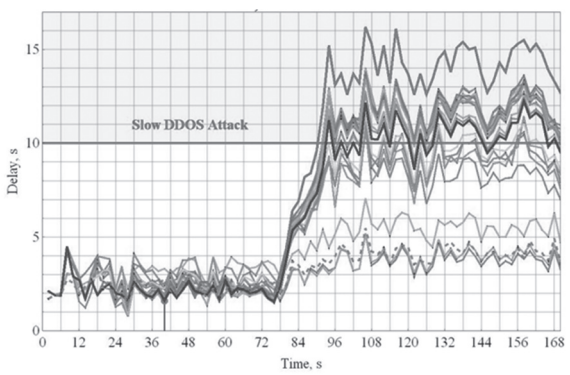

Рис. 2. Результати прогнозування підозрілої активності

Вирази (13) – (14) повністю описують лінію прогнозу, в якому (14) — математичне сподівання в точках t_i.

Приклад застосування методики

Моделювання виявлення повільної DDOS-атаки на основі прогнозування поведінки користувача було виконано для атаки RUDY. Для простоти розглядався лише один випадок атаки на фоні нормального трафіку, як зображено на рис. 1.

Середня затримка між переданими пакетами розглядається як досліджуваний параметр. RUDY — це атака на мережний сервер, спрямована на збій вебсервера через надсилання довгих запитів. Атака виконується за допомогою інструмента, який сканує цільовий вебсайт і виявляє вбудовані вебформи.

Після виявлення форм RUDY надсилає законні HTTP-запити POST з аномально довгим полем заголовка довжини вмісту, після чого починає вводити інформацію по одному байту на пакет. Цей тип атаки важко виявити через мізерні коливання

вхідного трафіку. До процесу, зображеного на рис. 1, було застосовано методику прогнозування (4) – (14), взявши за вихідні значення спостережень окремі точки часового ряду, які відповідають частковій траєкторії № 1 (рис. 1, *крива 1*). Беручи цю криву за контрольну, як вихідні дані спостережень брали перші значення часового ряду, котрі відповідають $t = 1, 40, 90$ с спостережень. На рис. 2 наведено результати прогнозування при $t = 40$ с. Невелика кількість початкових керуючих даних може лише відтворити процес у цілому (середня крива процесу), але конкретні значення прогнозованого трафіку будуть сильно відрізнятися від реальних (керуюча траєкторія). Так, знаючи середні параметри мережного трафіку і точки входу в прогноз, ви не можете точно передбачити подальшу поведінку системи. У такий спосіб метод «підбирає» необхідну траєкторію залежно від точки входу та середньої траєкторії.

Збільшення кількості спостережень до $t = 90$ с підвищує вірогідність подальшого прогнозу і при $t = 90$ с можна говорити про досить точний прогноз. Тобто ймовірність помилки у виборі правильної траєкторії залежить від кількості вихідних даних, що спостерігаються. Логічно припустити, що в цьому випадку точність прогнозу буде занадто сильно залежати від особливостей поведінки траєкторії, які призводять до аномального руху, а також від спостережуваної частоти аномалій.

ВИСНОВКИ

1. Повільні DDoS-атаки стають дедалі поширенішими через простоту реалізації та складність їх виявлення. Виявлення атаки сучасними методами є неефективним через відтермінований характер реакції на атаку в разі спостереження та аналізу параметрів трафіку. Більш перспективним підходом є прогнозування поведінки користувачів на основі аналізу підозрілого трафіку в мережі.

2. Прогнозування підозрілого трафіку забезпечує розв'язання проблеми виявлення повільних DDoS-атак на основі алгоритму пошуку невідомих майбутніх значень для часового ряду параметрів трафіку. Запропонований метод поєднує в собі переваги штучного інтелекту та статистичного аналізу, а також здатний до самонавчання в разі поповнення статистики атак. Такий підхід дає змогу точно визначати випадковий процес у контрольних точках і забезпечувати мінімум середньоквадратичної похибки апроксимації в інтервалах між цими точками.

3. Напрямком подальших досліджень у сфері протидії повільним DDoS-атакам може бути широкий спектр питань для вдосконалення методу для забезпечення можливості прогнозування з інтервалами, що виходять за межі доступної статистики, зокрема умови високого шуму даних або їх часткової відсутності.

Список використаної літератури

*1. **Dhanapal A., Nithyanandam P.** The Slow Http Distributed Denial of Service Attack Detection in Cloud // Scalable Computing: Practice and Experience. 2019. Vol. 20, N. 2. P. 285–298. URL:*
https://doi.org/10.12694/scpe.v20i2.1501

*2. **Dhanapal A., Nithyanandam P.** The Slow HTTP DDOS Attacks: Detection, Mitigation and Prevention in the Cloud Environment // Scalable Computing: Practice and Experience. 2019. Vol. 20, N. 4. P. 669–685. URL:*
https://doi.org/10.12694/scpe.v20i4.1569

*3. **Lukaseder T., Ghosh S., Kargl F.** Mitigation of Flooding and Slow DDoS Attacks in a Software-Defined Network. 16 August 2018. URL:*
https://arxiv.org/pdf/1808.05357.pdf

*4. **Abusaimeh H., Atta H., Shihadeh H.** Survey on Cache-Based Side-Channel Attacks in Cloud Computing // International Journal of Emerging Trends in Engineering Research. April 2020. Vol. 8, No. 4. P. 1019–1026.*

*5. **Calvert C. L., Khoshgoftaar T. M.** Impact of class distribution on the detection of slow HTTP DoS attacks using Big Data // Journal of Big Data. 2019. 6, 67. URL:*
https://doi.org/10.1186/s40537-019-0230-3

*6. **Cusack B., Tian Z.** Detecting and tracing slow attacks on mobile phone user service. In Valli, C. (Ed.) // The Proceedings of 14th Australian Digital Forensics Conference, 5-6 December 2016, Edith Cowan University, Perth, Australia. 2016. P. 4–10.*

*7. **Дуравкін Є. В., Карлссон А., Локтіонова А. С.** Метод виявлення повільної атаки // Системи обробки інформації. 2014. Вип. 8 (124). С. 102–106.*

*8. **Рубан І. В., Прибильнов Д. В., Лошаков Е. С.** Метод виявлення низькошвидкісної атаки типу «відмова в обслуговуванні» // Наука і техніка Повітряних Сил ЗС України. 2013. № 4(13). С. 85–88.*

*9. **Тарасов Я. В.** Дослідження застосування нейронних мереж для виявлення низькоінтенсивних DDoS-атак прикладного рівня // Питання кібербезпеки. 2017. №5(24). С. 23–29. URL: https://doi.org/10.21681/2311-3456-2017-5-23-29*

*10. **Краковський Ю. М., Лузгін А. Н.** Прогнозування інтенсивності кібератак на інформаційні системи критичних інфраструктур. Проблеми розумних міст та сталого розвитку територій // БЕЗПЕКА 2018. Єкатеринбург, 4-5 жовтня, 2018. 34-42. С. 180–187.*

*11. **Лисенко С., Ткачук В.** Методика та програмне забезпечення виявлення р.у.д.й. атака на основі використання алгоритму визначення самоподібності трафіку // Вісник Хмельн. нац. ун-ту. 2019. Вип. 3. С. 273.*

*12. **Idhammad M., Afdel K., Belouch M.** Detection System of HTTP DDoS Attacks in a Cloud Environment Based on Information Theoretic Entropy and Random Forest // Security and Communication Networks. Vol. 2018, Article ID 1263123. 13 p. URL: https://doi.org/10.1155/2018/1263123*

*13. **Network** traffic forecasting based on the canonical expansion of a random process / V. Savchenko, O. Matsko, O. Vorobiov [et al.] // Eastern European J. of Enterprise Technologies. 2018. V. 3, No 2 (93). P. 33–41. URL: https://doi.org/10.15587/1729-4061.2018.131471*

*14. **Detection** of Slow DDoS Attacks based on User's Behavior Forecasting / V. Savchenko, O. Ilin, N. Hnidenko [et al.] // International J. of Emerging Trends in Engineering Research. May 2020. Vol. 8, No. 5. P. 2019–2025.*

А. Д. Кожуховский, О. Ю. Ильин, В. А. Савченко, А. Г. Захаржевский

ПРОГНОЗИРОВАНИЕ ДИНАМИКИ ПОДОЗРИТЕЛЬНОЙ АКТИВНОСТИ В СЕТИ НА ОСНОВЕ АНАЛИЗА СЕТЕВОГО ТРАФИКА

Рассмотрена возможность предварительного прогнозирования кибератак на основе анализа подозрительной активности в сети, что даст дополнительные возможности службам защиты информации в противодействии таким атакам. Общая проблема заключается в выявлении времени начала такой атаки, поскольку параметры трафика не имеют резких изменений. Используя методы машинного обучения на основе анализа подобных ситуаций в прошлом, есть возможность создания интегрированной системы для трансформации больших объемов общедоступных данных для предсказания поведения злоумышленников в сети.

Ключевые слова: кибербезопасность; подозрительная активность; прогнозирование нестационарных процессов; машинное обучение.

A. D. Kozhukhivskyi, O. Yu. Ilyin, V. A. Savchenko, A.H. Zakharzhevskyi

PREDICTION OF DYNAMICS OF SUSPICIOUS NETWORK ACTIVITY BASED ON NETWORK TRAFFIC ANALYSIS

The article considers the possibility of early prediction of cyberattacks based on the analysis of suspicious activity in the network, which will provide additional opportunities for information protection services in countering such attacks. A feature of a slow DDoS attack is the use of a vulnerability in the TCP protocol, where interruptions can be caused intentionally or unintentionally as a result of delays in the communication channel. It is well known that detection of slow DDoS attacks is significantly different from volume-based attacks, as slow attacks do not increase network traffic. The general problem is to detect the start time of such an attack, since traffic parameters do not change dramatically. An assumption is made about the dependence of the slow attack on the user's behavior. Using machine learning methods based on the analysis of similar situations in the past, it is possible to create an integrated system for transforming large volumes of publicly available data to predict the behavior of attackers in the network. A method of detecting such attacks based on research and prediction of suspicious user activity is proposed. The possibilities of using this method have been proven on the basis of modeling RUDY attacks on HTTP services. The characteristics of forecasting accuracy depending on the accumulated traffic and attack statistics are given. It is concluded that this method can be used to detect different types of slow DDoS attacks. Predicting suspicious traffic provides a solution to the problem of detecting slow DDoS attacks based on an algorithm for finding unknown future values for a time series of traffic parameters. The proposed method combines the advantages of artificial intelligence and statistical analysis and is capable of self-learning in case of replenishment of attack statistics. This approach allows you to accurately determine the random process at the control points and ensure a minimum of the mean square error of approximation in the intervals between these points.

Keywords: cyber security; suspicious activity; prediction of non-stationary processes; machine learning.