DOI 10.31891/2307-5732-2020-291-6-22-27
УДК 004.75:004.8:004.49

А.С. КАШТАЛЬЯН, О.С. САВЕНКО
Хмельницький національний університет

КЛАСТЕРИЗАЦІЯ ЗЛОВМИСНИКІВ У КОМП'ЮТЕРНИХ МЕРЕЖАХ ЗА ЇХ ПОВЕДІНКОВИМИ ХАРАКТЕРИСТИКАМИ

В статті запропоновано метод знаходження схожих зловмисників в мережі приманок за їх поведінковими характеристиками. Розглянуто частотні характеристики, які дозволяють визначити тип атаки, а також спрогнозувати її. Запропоновано розглядати поведінкові характеристики зловмисників щодо мережі приманок як багато до багатьох, багато до одного, один до багатьох та один до одного. За поведінкову характеристику для розгляду обрано часовий ряд, який представляє залежність кількості запитів від джерела атаки до приманки за одиницю часу. Розглянуто інформаційні критерії, які можуть бути використані для кластеризації поведінкових характеристик зловмисників.

Запропоновано процедуру агломеративної кластеризації поведінкових характеристик на основі інформаційного критерію. Визначено критерій відстані між кластерами та інтегральний критерій відстані між об'єктами всередині кластеру. Визначено три критерії зупинки процесу кластеризації: кількість кластерів, абсолютний мінімум та відносний градієнт. Кількість кластерів у якості критерію зупинки дозволяє визначати рівень небезпеки зловмисників. Критерій абсолютного мінімуму відповідає максимальній чистоті кластерів зловмисників. Критерій відносного градієнту забезпечує найвищу повноту кластерів.

На основі даних мережі приманок проведено експериментальні дослідження, на їх основі визначено інформаційний критерій, який забезпечує найвищу точність кластеризації, та критерій зупинки, який забезпечує оптимальне співвідношення точності та повноти кластерів.

Ключові слова: приманка, кластеризація часових рядів, профіль зловмисника, частотні характеристики, інформаційний критерій.

A. KASHTALIAN, O. SAVENKO
Khmelnytsky National University, Khmelnytsky, Ukraine

CLUSETRING OF INTRUDERS IN COMPUTER NETWORKS BY THEIR BEHEVIOUR CHARACTERISTIC

The method of finding similar intruders in a honeynet by their behavior characteristics is suggested. Frequency characteristics which allow to determine attack's type and to predict it is considered. It is proposed to consider behavior characteristic of intruders with regard to a honeynet as many to many, many to one, one to many and one to one. A time series which represents the dependence of the number of requests from an attack's source to a honeypot per time unit is chosen as a behavior characteristic to examine. It is considered the information criterions which can be used for clustering intruders' behavior characteristics.

It is proposed the procedure of agglomerative clustering behavior characteristics on the basis of the information criterion. It is defined distances between clusters and the integral criterion of a distance between objects inside a cluster. It is determined three criterions of stopping the clustering process: the number of clusters, the absolute minimum, the relative gradient. The number of clusters as a stop criterion allows determining the level of a damage of intruders. The absolute minimum criterion corresponds to the maximum purity of intruders clusters. The relative gradient criterion provides the highest cluster completeness.

The experimental researches were conducted based on data of the honeynet. It is defined the information criterion which provides the highest accuracy of clustering and the stop criterion which provides the optimal ratio of the accuracy and the completeness of clusters on their basis.

Keywords: honeypot, time series clustering, intruder's profile, frequency characteristics, information criterion.

Вступ. Постановка проблеми

Біля 20% зловмисних доменів є новими і використовуються в середньому 1 тиждень після реєстрації [1]. З постійно виникаючими новими загрозами ризики для безпеки інформації комп'ютерних мережах значно зростають. Для захисту комп'ютерних мереж розроблено і використовуються багато рішень, які дозволяють ефективно виявляти атаки в процесі їх дії та аналізувати атаки, які вже відбулися. Однак для більш ефективного захисту необхідний збір та аналіз інформації, що стосується вразливостей системи та характеристик зловмисників до того, як зловмисні дії завдадуть суттєвої шкоди. Саме для цих цілей призначені приманки та їх мережі [2]. Приманка – це засіб, що дозволяє «заманити» зловмисника та зібрати про нього інформацію до того, як зловмисник проникне в реальну мережу. Оскільки приманка не виконує реальних функцій, то увесь трафік, який надходить на неї, можна вважати зловмисним. Якщо розглядати мережу приманок, то її трафік представляє собою результат дій зловмисників. Зловмисники можуть бути різного типу та здійснювати атаки різного типу з різним рівнем небезпеки. Незважаючи на значну кількість зловмисників, їх дії є подібними за своєю метою та способом здійснення. Тому, важливою задачею є знаходження груп подібних зловмисників для подальшого використання цієї інформації в системах виявлення зловмисників в реальному часі.

В більшості випадків для систем виявлення втручань виконується кластеризація відносно зловмисних дій за традиційними та сучасними підходами. В роботі [3] пропонується використовувати кластеризацію за *k*-середніми. В сучасних системах використовуються нейромережеві моделі [4] та генетичні алгоритми [5]. Також використовуються підходи до кластеризації атак на основі дерев рішень [6] та методу опорних векторів [7]. В багатьох роботах не розглядається питання визначення оптимальної кількості кластерів та характеристики атак часто розглядаються як статичні. В даній роботі кластеризація

зловмисних дій проводиться групуючись на динамічних характеристиках джерел атак та кількість кластерів визначається автоматично.

1. Частотні характеристики джерел атак.

Інформація, яку збирає приманка, описує характеристики зловмисників. Серед таких характеристик важливе місце займають частотні характеристики, до яких можна віднести: кількість джерел атак за одиницю часу; кількість нових джерел, що додаються за одиницю часу; кількість джерел атак в певній IP-агломерації; кількість джерел на кількість атак; кількість запитів за одиницю часу; кількість переданих пакетів за одиницю часу. Спостереження свідчать, що у випадку використання TCP протоколу, розміри пакетів є практично однаковими [8].

Частотні характеристики атак в багатьох випадках дозволяють визначити тип атаки, а отже дають можливість її прогнозувати. А також на основі частотних характеристик можна визначити рівень небезпеки зловмисника в реальному часі та вжити відповідних заходів.

Відношення джерел атак до приманок можна описати як багато до багатьох, оскільки один зловмисник здійснює атаки на різні цілі, так само одна приманка піддається впливу атак багатьох джерел. Якщо розглядати відношення один до багатьох (одне джерело атаки, багато приманок), отримаємо профіль зловмисника. Якщо розглядати відношення багато до одного (багато джерел атак, одна приманка), отримаємо характеристики розподіленої атаки, а також профіль вразливостей приманки.

Профіль зловмисника, який може атакувати певну кількість приманок, може бути розглянутий як інтегральний профіль, що складається з профілів один до одного (одне джерело атаки, одна приманка). Для отримання інтегрального профілю зловмисника необхідно отримати характеристики атак на приманки з точки зору відношення один до одного. Якщо розглядати всі атаки на приманку в межах заданого часового інтервалу, то вони в загальному випадку можуть бути результатом різних типів атак.

Розглянемо в якості поведінкової характеристики зловмисника часовий ряд, що представляє собою кількість запитів від одного джерела атаки до однієї приманки за інтервал часу в 1 секунду. Врахуємо той факт, що всі запити представляють собою зловмисну активність.

Часовий ряд кількості запитів від *i*-го джерела атаки до *j*-ї приманки за одиницю часу:

$$X^{ij} = \left\{x_1^{ij}, x_2^{ij}, x_3^{ij}, \ldots, x_k^{ij}\right\},$$

де x_k^{ij} – кількість запитів від *i*-го джерела до *j*-ї приманки за 1 сек.; *k* – тривалість досліджуваного часового вікна; *i*=1,2,3,…,*m*, *m* – кількість джерел атак; *j*=1,2,3,…*n*, *n* – кількість приманок.

2. Інформаційні критерії, використані для кластеризації.

Завдання кластеризації часових рядів передбачає виявлення патернів, які представляють часовий ряд якомога інформативніше [9]. Особливість кластеризації часових рядів полягає в тому, що ознаки змінюються в часі. Методи кластеризації часових рядів можна розділити та три типи: методи, в яких використовуються безпосередньо необроблені дані; методи, які використовуються вилучені ознаки; методи, які ґрунтуються на моделях часових рядів. [10].

Методи, що ґрунтуються на моделях, передбачають попередню трансформацію даних часових рядів шляхом визначення параметрів моделі. Використовуються моделі різного типу, зокрема: динамічна регресія; ARMA, ARIMA, ARFIMA; GARCH, NGARCH; нейромережеві моделі. Метою використання параметрів моделі часового ряду замість його значень є перехід до простору ознак, в якому можна виконати кластеризацію. Для визначення відстаней між параметрами моделей часових рядів використовується ряд інформаційних критеріїв.

Інформаційний критерій визначається як подвоєне негативне значення зміщеної корекції

$$-2\log L\left(\hat{\Psi}\right)+2C,$$

де *L* – функція максимальної правдоподібності, $\hat{\Psi}$ – вектор параметрів, C – міра складності (штрафний член). Критерій використовується для вибору моделі, яка забезпечує його мінімум.

Інформаційний критерій Акаіке [11]

$$AIC = -2\log L\left(\hat{\Psi}\right)+2d,$$

де *d* –число параметрів моделі.

Модифікований критерій Акаіке у випадку числа параметрів моделі d достатньо великого відносно розміру вибірки *n*.

$$AIC_C = -2\log L\left(\hat{\Psi}\right)+2dn/\left(n-d-1\right).$$

Інформаційний критерій Кульбака [12]

$$KIC = -2\log L\left(\hat{\Psi}\right)+3\left(d+1\right).$$

Байєсівський інформаційний критерій [13]

$$BIC = -2\log L\left(\hat{\Psi}\right) + d\log(n).$$

Розглянемо більш детально розрахунок інформаційного критерію як міри подібності двох часових рядів на основі байєсівського інформаційного критерію. Нехай задані два часових ряди, представлені векторами $X = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{n_x}\}$ та $Y = \{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_{n_y}\}$. Необхідно оцінити наступні дві гіпотези [14]:

$$H_1 : \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{n_x}, \mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_{n_y} - N(\mu, \Sigma),$$
$$H_2 : \begin{array}{l} \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{n_x} - N(\mu_X, \Sigma_X); \\ \mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_{n_y} - N(\mu_Y, \Sigma_Y). \end{array}$$

Гіпотеза H_1 передбачає, що X та Y отримані з однієї моделі, гіпотеза H_2 означає, що X та Y формуються двома моделями. Нехай $Z = X \cup Y$ та $n = n_x + n_y$. Тоді, різниця між величинами байєсівського критерію для двох гіпотез може бути визначена таким чином:

$$\Delta BIC(X,Y) = BIC(H_1, Z) - BIC(H_2, Z) =$$
$$= \log p\left(X \mid \hat{\boldsymbol{\mu}}_X, \hat{\boldsymbol{\Sigma}}_X\right) + \log p\left(Y \mid \hat{\boldsymbol{\mu}}_Y, \hat{\boldsymbol{\Sigma}}_Y\right) - \log p\left(Z \mid \hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}}\right) - \frac{1}{2}\lambda\left(d + \frac{1}{2}d(d+1)\right)\log n =$$
$$= \frac{n}{2}\log\left|\hat{\boldsymbol{\Sigma}}\right| - \frac{n_x}{2}\log\left|\hat{\boldsymbol{\Sigma}}_X\right| - \frac{n_y}{2}\log\left|\hat{\boldsymbol{\Sigma}}_Y\right| - \frac{1}{2}\lambda\left(d + \frac{1}{2}d(d+1)\right)\log n,$$

де $\hat{\boldsymbol{\mu}}$, $\hat{\boldsymbol{\mu}}_X$ та $\hat{\boldsymbol{\mu}}_Y$ - вектори середніх значень вибірок *Z*, *X* та *Y* відповідно; $\hat{\boldsymbol{\Sigma}}$, $\hat{\boldsymbol{\Sigma}}_X$ та $\hat{\boldsymbol{\Sigma}}_Y$ - коваріаційні матриці вибірок; *d* – розмірність вектора ознак [15]. Чим менша величина Δ*BIC*, тим більш схожими є два часових ряди. За умови $\lambda = 0$ відстань Δ*BIC* є еквівалентною відношенню правдоподібності.

3. Агломеративна кластеризація зловмисників за поведінковими характеристиками.

Процедура кластеризації часових рядів, які характеризують профілі зловмисників, ґрунтується на процедурі агломеративної кластеризації та включає такі кроки:

1. Визначення початкових кластерів. На першому етапі кластеризації кожний часовий ряд представляє собою кластер. Відповідно множина кластерів має вигляд:

$$C = \{C_1, C_2, C_3, \ldots, C_l\}$$

де $C_1 = \{X^{11}\}, C_2 = \{X^{12}\}$, $C_3 = \{X^{13}\}$, і так далі, $C_l = \{x^{mn}\}$; $l = m \times n$ - кількість початкових кластерів.

2. Визначення відстаней між парами кластерів, які визначають на основі одного з інформаційних критеріїв, наведених вище. Якщо в якості такого інформаційного критерію використовується байєсівський інформаційний критерій, то відстань між двома кластерами $C_i = \{X_1, X_2, X_3, \ldots, X_m\}$ та $C_j = \{Y_1, Y_2, Y_3, \ldots, Y_n\}$ визначається як

$$D_{ij} = \Delta BIC\left(C_i, C_j\right) = \frac{\sum_{a=1}^{m}\sum_{b=1}^{n}\Delta BIC\left(X_a, Y_b\right)}{\left|C_i \times C_j\right|},\ a \neq b,$$

де *m* – кількість об'єктів (часових рядів) в кластері C_i; *n* – кількість об'єктів (часових рядів) в кластері; i,j=1,2,3…l – кількість кластерів поточної ітерації процесу кластеризації.

3. Визначення з множини відстаней $D = \{D_{12}, D_{13}, \ldots, D_{ij}, \ldots, D_{l-1,l}\}$ мінімальної для поточної ітерації $D_{\min iter} = \min(D) = D_{ij}$, вибір відповідної пари кластерів C_i та C_j та об'єднання множин об'єктів цих двох кластерів в один кластер $C_{join(ij)} = C_i \cup C_j$.

4. Визначення інтегрального критерію відстаней між об'єктами всередині кластера. Інтегральний критерій визначається на основі одного з інформаційних критеріїв. Якщо цей критерій визначається на основі байєсівського інформаційного критерію, то він має вигляд:

$$J = \frac{\sum_{i=1}^{l}\Delta BIC\left(C_i\right)}{l},$$

де $\Delta BIC(C_i)$ - відстань між об'єктами всередині одного кластеру; i=1,2,3,…,l; l – кількість кластерів. Ця відстань для кластера $C_i = \{X_1, X_2, X_3, \ldots, X_m\}$ визначається за співвідношенням:

$$\Delta BIC(C_i) = \frac{\sum_{a=1}^{m}\sum_{b=a}^{m} \Delta BIC(X_a, X_b)}{m}.$$

5. Перевірка умови зупинки процесу об'єднання кластерів. Якщо умова виконана, отримують остаточну множину кластерів $C_{res} = \{C_1, C_2, C_3, \ldots, C_{lres}\}$, де l_{res} – остаточна кількість кластерів. Виконання пунктів 2-5 повторюється до тих пір, поки умова зупинки не буде виконана.

Для зупинки процесу об'єднання кластерів використовується один з трьох критеріїв:

1. Кількість кластерів. Вимагає задання числа кластерів, що може потребувати попередньої оцінки. В загальному випадку кількість кластерів профілів зловмисників не є попередньо відомою. Тому рекомендується використовувати цей критерій зупинки для задачі визначення рівня небезпеки зловмисника.

2. Абсолютний мінімум функції залежності інтегрального критерію від числа ітерацій об'єднання кластерів $\min(J(k))$, де k – число ітерацій. У цьому випадку обирається множина кластерів, що відповідає ітерації, на якій спостерігається мінімум $J(k)$ (рис. 1(1)). Цей критерій зупинки забезпечує максимально чистоту отриманих кластерів. Недоліком є те, що частина кластерів може бути неповною, і зловмисників одного типу може містити більше ніж один кластер.

3. Градієнтний критерій. У цьому випадку визначається стрибок функції залежності відносного градієнта $\nabla J(k)/J(k)$ від числа ітерацій та обирається множина кластерів, відповідна цій ітерації (рис. 1(2)).

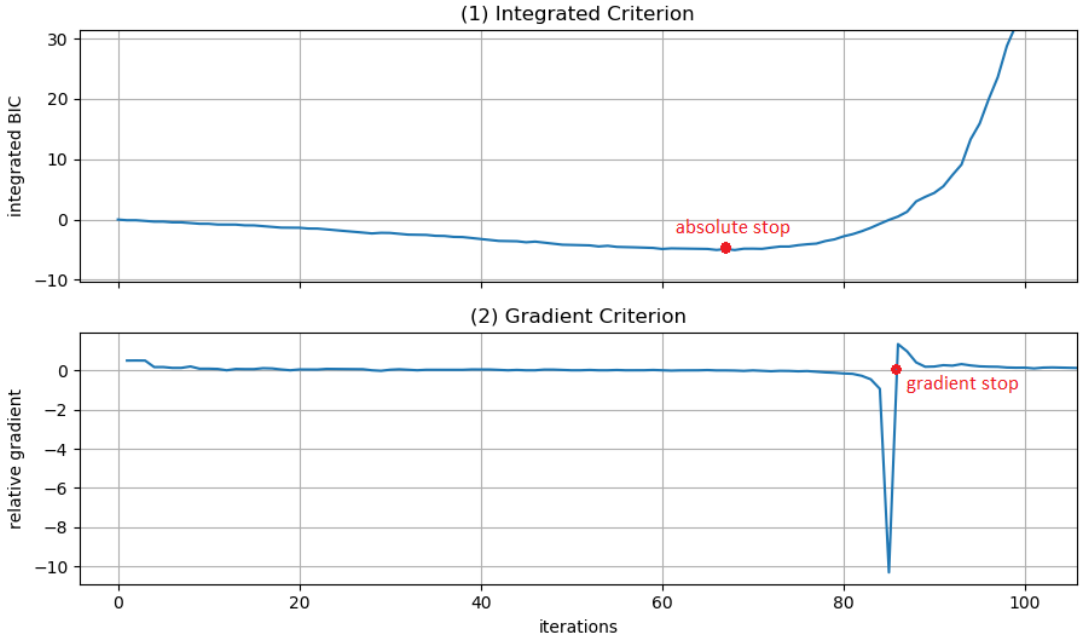

Рис. 1. Критерії зупинки процесу об'єднання кластерів. (1)Залежність інтегрального критерію відстаней між об'єктами всередині кластера від кількості ітерацій *J*(*k*). (2)Залежність відносного градієнту функції $\nabla J(k)/J(k)$ від кількості ітерацій.

4. Результати кластеризації зловмисників

Для проведення експериментальних досліджень було використано датасет мережі приманок [16]. З датасету були відібрані часові ряди кількості запитів за 1 секунду від одного джерела атаки до однієї приманки, які сформували набір часових рядів для кластеризації. Кластеризація проводилася із використанням трьох інформаційних критеріїв (AIC, KIC, BIC), також з використанням двох критеріїв зупинки процесу кластеризації (абсолютний мінімум та відносний градієнт). Найвищі показники точності показала кластеризація на основі BIC критерію із зупинкою за відносним градієнтом. Точність кластеризації перевищила 80%, а зупинка за відносним градієнтом забезпечила оптимальне співвідношення чистоти кластерів та їх повноти. Також, було проведено кластеризацію з числом кластерів 3 для визначення відносного рівня небезпеки зловмисника: високого, середнього та низького (рис. 2).

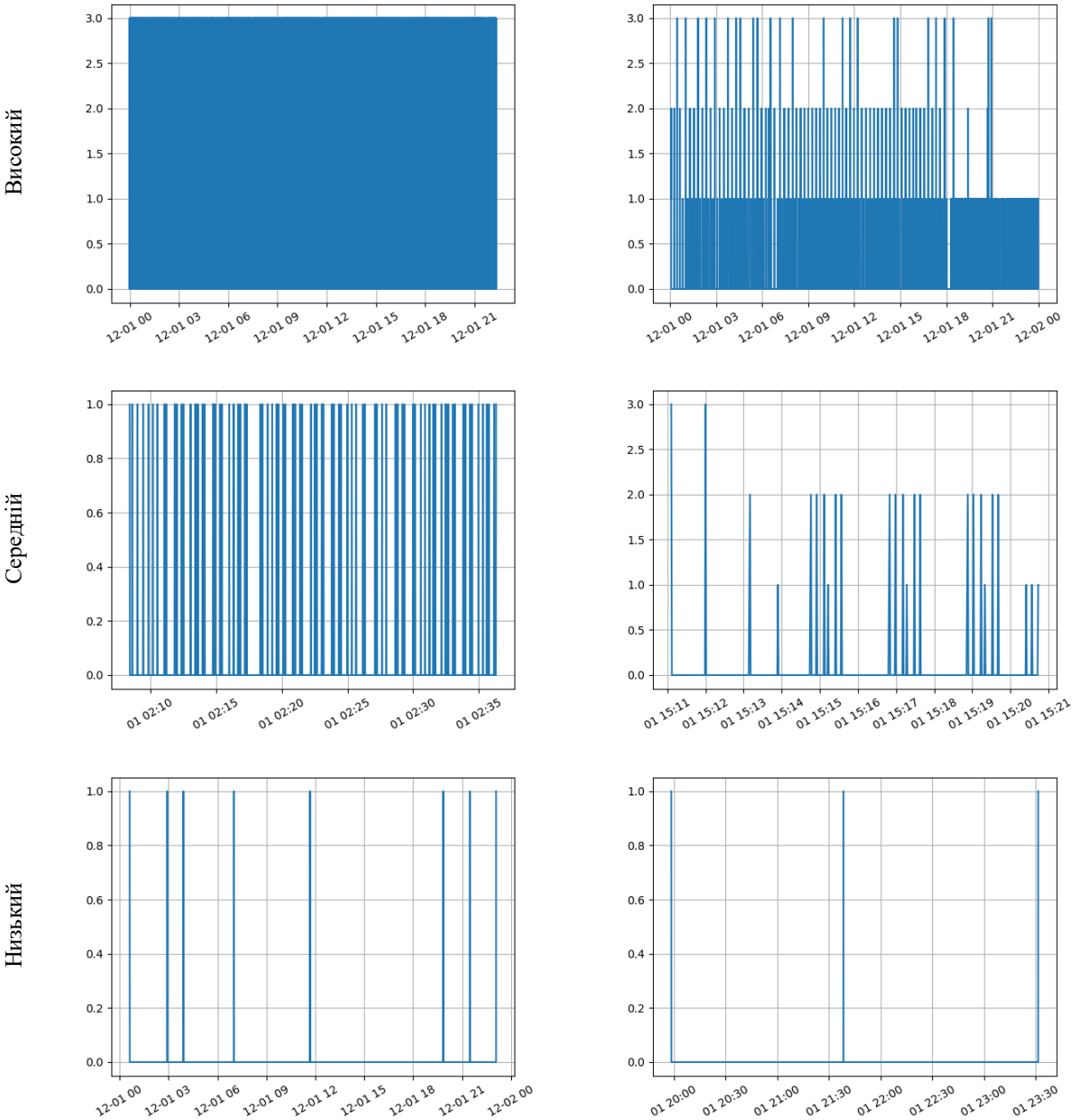

Рис. 2. Приклади частотних характеристик зловмисників з різними рівнями загрози.

Висновки

Збір та аналіз приманками поведінкових характеристик зловмисників дозволяє виявляти та групувати нові типи атак, що є надзвичайно важливим для систем виявлення втручань. Запропонований підхід дозволяє проводити неперервний пошук подібних зловмисників з врахуванням їх динамічних характеристик. Запропонована процедура агломеративної кластеризації на основі інформаційних критеріїв дозволяє отримати достатню точність кластеризації та визначити рівень небезпеки зловмисника.

Подальше дослідження проводиться в напрямку визначення поведінкових патернів зловмисників на основі відношення «один до багатьох», підвищення точності кластеризації та врахування додаткових поведінкових характеристик.

Література

1. Cisco 2019. Annual Report. Defining the Future of the Internet/ https://www.cisco.com/c/dam/en_us/about/annual-report/cisco-annual-report-2019.pdf. [Access 29.09.2020].

2. Каштальян А.С. Покращення безпеки та модель антивірусних інтелектуальних приманок в корпоративних комп'ютерних мережах/ А.С. Каштальян, О.С. Савенко// Вісник Хмельницького національного університету. Технічні науки, 2020, №4, с. 33-38.

3. Varaprasad Rao M. Algorithm for Clustering with Intrusion Detection using Modified & Hashed K-Means Algorithms/ M. Varaprasad Rao, A. Damodaram, N. Ch. Bhatra Charyulu// Proceedings of 2nd International

Conference on Computer Science Engineering Applications, New Delhi, India, Springer Publications Vol 2, May 25-27, 2012, pp 737-744.

4. Debar H. A neural network component for an intrusion detection system/ H. Debar, M. Becker, and D. Siboni// in Proceedings 1992 IEEE Computer Society Symposium on Research in Security and Privacy, 1992, pp. 240–250.

5. Bhattacharjee P.S. Intrusion detection system for nsl-kdd data set using vectorised fitness function in genetic algorithm/ P. S. Bhattacharjee, A. K. M. Fujail, and S. A. Begum// Adv. Comput. Sci. Technol., vol. 10, no. 2, 2017, pp. 235–246.

6. Sung S. Jo H. comparative study on the performance of intrusion detection using decision tree and artificial neural network models/ S. Jo, H. Sung, and B. Ahn// Journal of the Korea Society of Digital Industry and Information Management, vol. 11, no. 4, 2015, pp. 33–45.

7. Puri A. A novel technique for intrusion detection system for network security using hybrid svm-cart/ A. Puri and N. Sharma// IJEDR, vol. 5, no. 2, 2017, pp. 155–161.

8. Yegneswaran V. On the design and use of internet sinks for network abuse monitoring/ V. Yegneswaran, P. Barford, D. Plonka// Recent Advances in Intrusion Detection. Springer, 2004, pp. 146-165.

9. Iglesias Felix. Analysis of Similarity Measures in Times Series Clustering for the Discovery of Building Energy Patterns/ Felix Iglesias, Wolfgang Kastner/ Energies, №6, 2013, pp. 579-597.

10. Warren Liao T. Clustering of time series data—a survey/ T. Warren Liao// Pattern Recognition 38 (2005) 1857 – 1874 – pp. 1857-1874.

11. Akaike H. Information theory and an extension of the maximum likelihood principle. In Second International Symposium on Information Theory; Petrov, B.N., Csaki, F., Eds.; Akademiai Kiado: Budapest, Hungary, 1973; pp. 267–281.

12. Cavanaugh J.E. A large-sample model selection criterion based on Kullback's symmetric divergence/ J. E. Cavanaugh// Stat. Probab. Lett. 1999, 42, 333–343.

13. Schwarz G. Estimating the dimension of a model/ G. Schwarcz// Ann. Stat. 1978, 6, 461–464.

14. Chen S. Speaker, environment and channel change detection and clustering via the Bayesian information criterion/ S. Chen and P. S. Gopalakrishnan// in Proc. DARPA Broadcast News Transcription and Understanding Workshop, Lansdowne, VA, Feb. 1998, pp. 127–132.

15. Cettolo M. Evaluation of BIC-based algorithms for audio segmentation/ M. Cettolo, M. Vescovi, and R. Rizzi// Computer Speech and Language, vol. 19, 2005, pp. 147–170.

16. Traffic Data from Kyoto University's Honeypots, https://www.takakura.com/Kyoto_data/ [Access 01.10.2020].

References

1. Cisco 2019. Annual Report. Defining the Future of the Internet/ https://www.cisco.com/c/dam/en_us/about/annual-report/cisco-annual-report-2019.pdf. [Access 29.09.2020].

2. Kashtalian A.S. Security Improvement and the Model of Anti-Virus Intellectual Honeynet in Corporate Computer Networks/ A.S. Kashtalian, O.S. Savenko// Herald of Khmelnitsky National University, 2020, Issue 4, pp. 33-38.

3. Varaprasad Rao M. Algorithm for Clustering with Intrusion Detection using Modified & Hashed K-Means Algorithms/ M. Varaprasad Rao, A. Damodaram, N. Ch. Bhatra Charyulu// Proceedings of 2nd International Conference on Computer Science Engineering Applications, New Delhi, India, Springer Publications Vol 2, May 25-27, 2012, pp 737-744.

4. Debar H. A neural network component for an intrusion detection system/ H. Debar, M. Becker, and D. Siboni// in Proceedings 1992 IEEE Computer Society Symposium on Research in Security and Privacy, 1992, pp. 240–250.

5. Bhattacharjee P.S. Intrusion detection system for nsl-kdd data set using vectorised fitness function in genetic algorithm/ P. S. Bhattacharjee, A. K. M. Fujail, and S. A. Begum// Adv. Comput. Sci. Technol., vol. 10, no. 2, 2017, pp. 235–246.

6. Sung S. Jo H. comparative study on the performance of intrusion detection using decision tree and artificial neural network models/ S. Jo, H. Sung, and B. Ahn// Journal of the Korea Society of Digital Industry and Information Management, vol. 11, no. 4, 2015, pp. 33–45.

7. Puri A. A novel technique for intrusion detection system for network security using hybrid svm-cart/ A. Puri and N. Sharma// IJEDR, vol. 5, no. 2, 2017, pp. 155–161.

8. Yegneswaran V. On the design and use of internet sinks for network abuse monitoring/ V. Yegneswaran, P. Barford, D. Plonka// Recent Advances in Intrusion Detection. Springer, 2004, pp. 146-165.

9. Iglesias Felix. Analysis of Similarity Measures in Times Series Clustering for the Discovery of Building Energy Patterns/ Felix Iglesias, Wolfgang Kastner/ Energies, №6, 2013, pp. 579-597.

10. Warren Liao T. Clustering of time series data—a survey/ T. Warren Liao// Pattern Recognition 38 (2005) 1857 – 1874 – pp. 1857-1874.

11. Akaike H. Information theory and an extension of the maximum likelihood principle. In Second International Symposium on Information Theory; Petrov, B.N., Csaki, F., Eds.; Akademiai Kiado: Budapest, Hungary, 1973; pp. 267–281.

12. Cavanaugh J.E. A large-sample model selection criterion based on Kullback's symmetric divergence/ J. E. Cavanaugh// Stat. Probab. Lett. 1999, 42, 333–343.

13. Schwarz G. Estimating the dimension of a model/ G. Schwarcz// Ann. Stat. 1978, 6, 461–464.

14. Chen S. Speaker, environment and channel change detection and clustering via the Bayesian information criterion/ S. Chen and P. S. Gopalakrishnan// in Proc. DARPA Broadcast News Transcription and Understanding Workshop, Lansdowne, VA, Feb. 1998, pp. 127–132.

15. Cettolo M. Evaluation of BIC-based algorithms for audio segmentation/ M. Cettolo, M. Vescovi, and R. Rizzi// Computer Speech and Language, vol. 19, 2005, pp. 147–170.

16. Traffic Data from Kyoto University's Honeypots, https://www.takakura.com/Kyoto_data/ [Access 01.10.2020].

Надійшла / Paper received : 12.11.2020 р. Надрукована/Printed :04.01.2021 р.