DOI: https://doi.org/10.26565/2304-6201-2023-60-02

УДК 681.3

Деменкова Світлана Дмитрівна *старший викладач кафедри автоматизації хіміко-технологічних систем та екологічного моніторингу, Національний технічний університет Харківський політехнічний інститут, Харків, вул.. Кирпичова,2, 61002*
e-mail: svet1972232765@gmail.com
https://orcid.org/[0000-0003-0604-5456]

Демченко Катерина Вікторівна *к.т.н., доцент, доцент кафедри автоматизації та комп'ютерно-інтегрованих технологій Державний біотехнологічний університет, вул. Алчевських 44, м. Харків, Україна, 61002*
e-mail: yayaska@btu.kharkiv.ua
https://orcid.org/0000-0002-3168-5351

Корольова Яна Юріївна *к.т.н., доцент кафедри мультимедійних та інтернет технології і системи, Національний технічний університет Харківський політехнічний інститут, Харків, вул.. Кирпичова,2, 61002*
e-mail: yanakoroleva815@gmail.com
https://orcid.org/[0000-0003-0604-5456]

Пахомов Юрій Васильович *к.т.н., доцент кафедри комп'ютерних наук та інформаційних технологій Харківський національний університет міського господарства імені О.М. Бекетова, м. Харків, вул. Маршала Бажанова, 17, Україна, 61002*
e-mail: abc050073@gmail.com
https://orcid.org/0000-0002-2267-8600

Використання метода пошуку аномалій для виявлення мережевих атак

Стаття присвячена опису моделі виявлення мережевих вторгнень у трафіку мереж, побудованих з урахуванням стека протоколів TCP/IP. Аналізуються основні об'єкти локальної обчислювальної мережі. Описуються основні контрольовані параметри кожного типу об'єктів. У роботі розробляються методи пошуку аномалій, що ґрунтуються як на аналізі за правилами, так і на аналізі з використанням імовірнісних моделей.
Актуальність. У зв'язку з інтенсивним зростанням інформаційних технологій та їх впровадженням у різні галузі діяльності міського господарства питання інформаційної безпеки виходить на перше місце і стає дуже актуальним.
Методи дослідження. Під час вирішення поставлених завдань використовувалися методи теорії управління; методи побудови систем захисту; теорія графів; теорія ймовірності та математична статистика; методи аналізу часових рядів; методи прогнозної аналітики та обробки великих даних; методи побудови високонавантажених захищених програм. Основними методами дослідження є імовірнісне і верифікаційне моделювання.
Результати. Імовірнісне і верифікаційне моделювання мережевих атак підтвердило працездатність запропонованого підходу. Результати синтезу за допомогою САПР показали, що додаткові апаратні витрати не перевищують 20% порівняно з канонічною моделлю опису.
Висновки. Розроблена модель системи дозволяє виявляти атаки на ключові об'єкти, що моделюються. Отримані під час випробувань результати показали високу ефективність виявлення аномалій числових параметрів моделі. Для збільшення точності надалі планується перейти до моделювання поняття «сервіс» та моделювання протоколів HTTP, SMTP, POP3. Модель сесії також дозволяє виявляти існуючі та нові атаки на рівні сесії TCP, а також деякі види атак на відмову в обслуговуванні. Модель потоків мережевого трафіку дозволяє нам виявляти такі види атак, як: різні види сканування системи, установку троянських програм (бо зросте кількість байт у вихідному потоці), установку ICMP шелла (бо зросте кількість ICMP пакетів). Модель часових інтервалів дозволяє виявляти деякі види сканування системи, атаки на відмову в обслуговуванні та встановлення web-шеллу. Стаття присвячена опису моделі виявлення мережевих вторгнень у трафіку мереж, побудованих з урахуванням стека протоколів TCP/IP. Аналізуються основні об'єкти локальної обчислювальної мережі. Описуються основні контрольовані параметри кожного типу об'єктів. Розробляються методи пошуку аномалій, що ґрунтуються як на аналізі за правилами, так і на аналізі з використанням імовірнісних моделей.

***Ключові слова**: пошук аномалій, мережеві атаки, скінчені автомати, верифікація, моделювання, розподілених інформаційних мережах, імовірнісне моделювання, верифікаційне моделювання*

Як цитувати: Деменкова С. Д, Демченко К. В., Корольова Я. Ю., Пахомов Ю. В. Використання метода пошуку аномалій для виявлення мережевих атак. *Вісник Харківського національного університету імені В.Н. Каразіна, серія Математичне моделювання. Інформаційні технології. Автоматизовані системи управління*. 2023. вип. 60. С.15-26.
https://doi.org/10.26565/2304-6201-2023-60-02

How to quote: Demenkova S., Demchenko K., Koroleva Y., Pakhomov Y., "Using anomaly detection method to detect network attack", *Bulletin of V.N. Karazin Kharkiv National University, series Mathematical modelling. Information technology. Automated control systems*, vol. 60, pp.15-26, 2023. https://doi.org/10.26565/2304-6201-2023-60-02 [In Ukrainian].

1 Вступ

Сучасні методики знаходження та знешкодження кібератак у комп'ютерних системах та мереж, а також виявлення кібератак є неформальними щодо моделі такої атаки, тому для них складно оцінити обчислювальну складність, коректність, завершеність тощо. Зазвичай роздають методи виявлення атак, та зловживань. Більшість комерційних комп'ютерних систем та мереж використовують сигнатурні (експертні) методи виявлення. В промислових системах рідка використовуються системи академічних розробок в області виявлення аномалій, тому, що вони породжують дуже велику кількість помилкових спрацьовувань [1].

2 Постановка задачі

Побудова розподілених інформаційних мереж (РІМ), стійких до комп'ютерних атак, пов'язана зі значними витратами часу та ресурсів. Забезпечення працездатності РІМ залежить від їх здатності протистояти цілеспрямованим впливам, що порушують їх роботу. Для РІМ основною проблемою є низька ефективність виявлення невідомих атак та захисту від внутрішнього порушника.

У зв'язку з інтенсивним зростанням інформаційних технологій та їх впровадженням у різні галузі діяльності міського господарства питання інформаційної безпеки стає дедалі **актуальнішим**. Сьогодні більшість промислових виробництв для захисту своїх корпоративних мереж використовує якісь різні методики захисту від вторгнень, а саме міжмережеві екрани, антивірусні системи та системи виявлення вторгнень. Системи виявлення вторгнень (СВВ) – це програмні або апаратно-програмні засоби, які автоматизують події та процес їх контролю, які проходять у комп'ютерній системі або мережі, і також самі аналізують події та знаходять ознаки порушення політики безпеки. Системи виявлення вторгнень на сьогоднішній день є одним із необхідних компонентів інфраструктури безпеки корпоративної мережі для більшості організацій міського господарства.

Сучасні РІМ характеризуються надвисокими обсягами інтенсивна надходить різнорідного мережевого трафіку, у зв'язку з чим **актуальним** є розгляд систем виявлення вторгнень (СВВ), спрямованих на обробку великих даних. Математичні аспекти побудови СВВ та моніторингу РІМ для визначення внутрішнього порушника на сьогоднішній день розроблені недостатньо, а механізми прихованого моніторингу та аналізу Великих даних існують окремо. Тому побудова розподілених систем виявлення вторгнень (РСВВ) з використанням методів прихованого моніторингу та аналізу великих даних є **актуальним**.

Системи виявлення атак, залежно від використовуваної технології виявлення атак, поділяють на дві основні групи: системи, які виявляють зловмисну поведінку та аномальну поведінку. Перші покладаються на модель зловмисної поведінки (наприклад, шаблон атаки) і порівнюють модель з потоком подій. Другі покладаються на модель нормальної поведінки (наприклад профіль системи) і шукають аномальні входження в потік подій. Незважаючи на широке поширення СВВ та активні дослідження в даній галузі, існуючі системи мають ряд недоліків [1]. Наприклад, нездатність виявлення нових видів атак.

У роботі **розглядаються** системи виявлення аномальної поведінки, оскільки ця технологія дозволяє виявляти нові види атак. **Метою роботи** є розроблення моделей виявлення атак у мережевому трафіку на основі виявлення аномалії заголовків мережевих пакетів.

3 Огляд літератури

В [1] наведені таки підходи, які класифікують методи, які виявляють аномалії у системах, які виявляють атаки. Також проведено аналіз та розгляд популярних групи методів виявлення аномалій та зазначено, що методи виявлення аномалій в системах виявлення атак не дуже формалізовані для побудови моделі атаки, та в них досить складно оцінити обчислювальну складність, коректність та завершеність.

У [2] розглядається можливість автоматизації процесу побудови мережевих графіків за допомогою підходу мультипаралельної обробки, який дозволяє перетворювати алгоритми лінійної дії в паралельні алгоритми. Метою даного дослідження є скорочення часу мережевого планування для виконання ІТ-проектів за рахунок використання методів мультипаралельної обробки.

У [3] наведені методи побудови гетерогенних комп'ютерних систем і мереж критичного застосування також були проаналізовані вимоги до цих систем, наведено перелік та зміст комп'ютерних технологій, розроблено новий метод автоматизованого проектування гетерогенних комп'ютерних систем та мереж критичного застосування.

У [4] проведено аналіз методів тестового діагностування двовимірних однорідних мереж (ОМ), запропоновано метод модифікації автоматної моделі мережі, що забезпечує С - тестування рядків та L - тестування стовпців мережі

У [5] запропоновано використовувати системи запобігання та виявлення вторгнень для того, щоб захист мережи був комплексний. Шляхом порівняння систем було наведено виявлено їх недоліки та вимоги для наближення до так званої супер системи по запобіганню та виявленню кібератак та вторгнень.

У [6] проведено аналіз характеристик виявлення мережевих вторгнень в інформаційну систему і виявлення ознак комп'ютерних атак на підприємстві; аналіз можливих дій зі сторони зловмисників, досліджено методи та принципи встановлення оптимальної системи виявлення мережевих вторгнень; розглянуто можливості розробки та використання комп’ютеризованих систем для виявлення вторгнень до мережі і властивостей комп’ютеризованих атак на підприємстві в сучасних умовах; досліджено і розроблено рекомендації щодо впровадження таких систем виявлення атак, вторгнень та ознак кібератак для можливого подальшого встановлення їх в систему захисту інформації будь-якої організації.

У зв’язку з тім, що збільшується кількість різних комп'ютерних мережевих загроз, які призводять до фінансових втрат організацій, то в [7] показано, що дуже важний напряв у інформаційної безпеки – це розробка систем пошуку атак в комп’ютеризованих мережах. Також показані основні існуючи методи вирішення задач по виявленню атак в мережі. Розглядаються роботи, які присвячені методу вейвлет-аналізу по виявленню аномалій в мережевому трафіку та показані отримані тестові дані мережевого трафіку з аномаліями для практичного виконання, також показано практичне виконання шумоусунення сигналу для конкретизації та зменшення розміру даних, використано різноманіття методів вейвлет-аналізу для виявлення можливості появи аномалій.

У [8] показано, що ніякі методи захисту від кібератак в мережі не дають гарантію від проникнення зловмисника в комп’ютерну мережу систему та у випадку злому потрібно швидко виявити та перервати доступ, провести розслідування та виправити дірки в безпеці. Для цього треба використовувати методи, які виявляють зловживання та аномалії. Були досліджени можливості використання частотного методу, який виявляє аномалії в роботі системи шляхом аналізу ентропії журналу подій та використовується для виявлення аномалій у трафіку комп’ютерної мережі, при цьому аномалії в журналі подій на хостах можуть вказувати для перевірки на наявність несанкціонованих дій також. Дослідження були проведені на основі журналу подій в ОС Windows та показали, що можна виявити перевищення порогів безпеки щодо кількості різних повідомлень в журналі подій шляхом аналізу ентропії. Це вказує на аномалії в роботі комп’ютерної мережі. Метод, який був запропоновану можна інтегрувати у системи виявлення вторгнень, які будуть сповіщати адміністратора безпеки про зловживання та атаки.

У [9] показано, що всі комп’ютерні системи та мережи використовують системи виявлення кібератак та вторгнень до комп’ютерних мереж та інформаційних систем. Зазвичай це можуть бути або програмні, або апаратно-програмні засоби, які автоматизують процес контролю подій в комп’ютерної системі або мережі, і аналізують події для пошуку проблем безпеки.

У зв’язку зі збільшенням несанкціонованих атак у комп’ютерні мережі, системи виявлення кібератак (СВА) є важливою частиною системи безпеки організації. Існують багато методів виявлення кібератак та аномалій, але їх стійкість слабка, відсутня верифікація, дуже багато хибних спрацьовувань, характер не допомагають їх широкому використанню. Запропоновано аналіз, можливості найпоширеного типу атак – DDoS, тобто атак типу, а саме - відмовляє обслуговувати, шхяхом переривання або призупинення роботи, хост-сервера робить онлайн-сервіс недоступним для користувачів.

В [10] було запропоновано інтелектуальну систему, яка виявляє аномалій та ідентифікує пристрої розумних будинків, які застосовують колективну комунікацію. Сенс роботи запропонованої системи у отриманні результату від об'єднання розумних будинків в одну комп'ютерну мережу для підвищення безпеки для окремо розумного будинку і всієї мережі. Гарна властивість системи – це зв'язок кластерів розумних будинків один з одним для обміну інформації про профілі розумних пристроїв в білих списках кожного кластеру.

В [11] були систематизовані, узагальнені та розвинені поняття про методики і системи, які аналізують комп'ютерні мережі, які вимагають захисту. Наведений математичний апарат побудови моделі справного функціонування комп'ютерних систем, визначення загальної оцінки стану системи, якої потрібен захист. Надано загальні недоліки та напрямки подальшого розвитку комп'ютерних систем по виявленню кібератак та вторгнень. Інтелектуальна комп'ютерна система виявлення кібератак, аномалій та ідентифікації пристроїв розумних будинків із застосуванням колективної комунікації.

У [12] показано, що при розвитку комп'ютерних систем до вдосконалення деструктивного програмного забезпечення, спрямованого на різноманітні ресурси комп'ютерних систем. Дуже небезпечними є ті засоби, які підлаштовуються під реальне програмне забезпечення чи web-сервіс та знаходять шлях до персональних даних користувача, або можуть використовивать ці ресурси чи програмне забезпечення у своїх цілях. При активізації таких атак, активізується і засоби побудови спеціалізованих засобів виявлення та протидії кібератак, які будуть ефективні проти наявних та наступних кіберзагроз з невстановленими або нечітко визначеними властивостями, тобто функціонують у нечіткому, слабоформалізованому середовищі. Методики, моделі та комп'ютерні системи, основані на нечітких множинах використовуються для розробки та переробки існуючих засобів виявлення вторгнень та аномалій у комп'ютерних системах та мережах, які виникають при реалізації кіберзагроз. Метод формування лінгвістичних еталонів для комп'ютерних систем по виявленню вторгнень використовується при виявленні однією з ознак.

4 Загальна модель мережевого трафіку

Методи виявлення аномалій можна класифікувати за різними критеріями, а саме: по джерелам даних (Network-based IDS, Host-based IDS, Application-based IDS, Hybrid), по аналізу даних (експертна оцінка та машинне навчання), по способу отримання даних (сенсори, журнал подій, комбінований), по характеру отриманих даних (структурний аналіз, поведінковий аналіз).

В процесі порівняння конкретних груп методів виявлення аномалій: статистичний метод, кластерний аналіз, нейронні мережі, імунні мережі, експертні системи, метод опорних векторів, сигнатурні методи, було виявлено, що методи, які використовують підходи визначення аномалій використовують адаптивність, яка залежить від того, який метод був реалізовано. Тому вони виявляють тільки аномальну поведінку системи, але вони не класифікують цю аномалію. В перспективі при необхідності розробки нових методів виявлення аномалій, треба як можна раніше виявляти нові ще невідомі атаки, коли система знаходиться в робочому стані та в статичному, який відмінний від норми, яка задається розробленою моделлю нормальної поведінки [1].

Розглянемо трафік у звичайній мережі на основі стека протоколів TCP/IP, коли програма відправляє пакет за протоколом TCP. Типовий шлях пакета по локальній мережі виглядає наступним чином: пакет протоколу TCP з мережі Internet потрапляє у захищену локальну мережу, потім пересилається до конкретного комп'ютера у мережі (хосту). Далі пакет відправляється на конкретний порт, пов'язаний із конкретною сесією TCP.

Щоб точно змоделювати «нормальний» трафік мережі, введемо модель кожного з п'яти ключових значень (об'єктів): пакет (D), мережа (A), хост (H), порт (P), сесія (S). Кожне ключове значення K_1, K_2, K_3, K_4, K_5 можна охарактеризувати за допомогою набору параметрів $K_i(\kappa_1^i, \kappa_2^i, ..., \kappa_n^i),\ i \in (1,5).$ Пакет, що потрапляє в мережу, передається для аналізу в кожну з моделей. Спочатку у модель пакета, де перевіряється коректність полів заголовка пакета. Потім пакет передається у модель мережі, де він перевіряється на відповідність профілю мережі. Модель мережі вибирає відповідну модель нижнього рівня – модель хосту тощо. Таким чином пакет передається від моделей верхнього рівня моделям нижнього рівня. Після аналізу, кожна модель видає значення аномальності пакета (рисунок 4.1).

Кожна з моделей має два типи параметрів: прості параметри; складові параметри.

Прості параметри описують одну з характеристик моделі, наприклад, обсяг трафіку за годину або нормальну довжину пакета. Прості настройки не можуть вказувати на ключові об'єкти. Складові параметри – вказівники на логічна структуровані множини ключових об'єктів нижнього рівня, наприклад, список портів або дерево хостів.

До кожного ключового значення складається профіль. Профіль являє собою набір параметрів $P_i(p_1^i, p_2^i,..., p_j^i, q_1^i, q_2^i,..., q_k^i)$, $i \in (1,5)$. Параметри p_j^i відповідають параметрам ключових значень k_j^i, n ≤ l, а q_i – це додаткові параметри, що характеризують профіль.

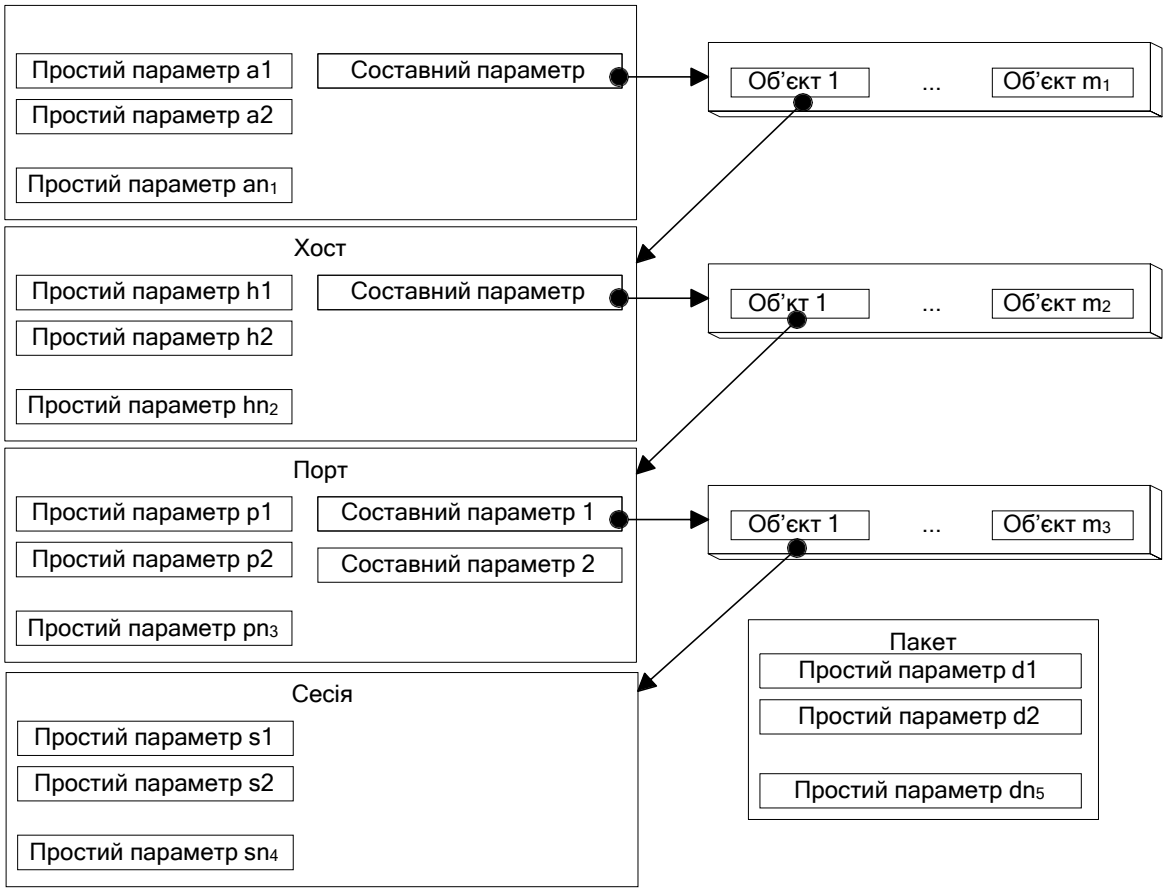

Рис.4.1 Взаємозв'язок параметрів ключових значень.

Як правило, це загальні параметри для всіх профілів: q_1 - тимчасова мітка створення профілю; q_2 – кількість пакетів у навчальній вибірці профілю; q_3 – тимчасова мітка останньої зміни профілю; q_4 – тимчасова мітка останнього аномального значення у профілі.

5 Загальна характеристика параметрів

За способом виявлення аномальних значень всі параметри всіх об'єктів *Ki* поділяються на два основні класи: чисельні; категоріальні.

До числових параметрів належать параметри, для перевірки яких використовуються статистичні та імовірнісні моделі. Наприклад, довжина пакета, обсяг переданих даних за одиницю часу, кількість пакетів у сесії тощо. Вони обчислюються за описаним нижче імовірнісним алгоритмом.

До категоріальних параметрів належать такі параметри, що можуть бути верифіковані за правилами. Ці параметри не можна оцінити у числовій формі. Наприклад, прапори TCP сесії.

Розглянемо докладніше числові параметри та методи їх аналізу. Пошук аномальних значень числових параметрів виконується з використанням імовірнісних методів. Для кожного числового параметра враховуються такі характеристики: мінімальне значення (V); максимальне значення (W); математичне очікування (M); дисперсія (D); порогове значення (Q).

Нехай є два класи пакетів: Ω_1 – клас нормальних пакетів та Ω_2 – клас пакетів, що належать мережевим атакам. Умовні щільності розподілу ймовірностей $f_1(x)$ та $f_2(x)$. Апріорна ймовірність появи об'єктів $P(\Omega_1)$ та $P(\Omega_2)$. Умовні ймовірності помилок 1-го та 2-го роду Q_1 та Q_2 можна визначити за формулами:

$$Q_1 = \int_{x_0}^{\infty} f_1(x)dx;\ Q_2 = \int_{-\infty}^{x_0} f_2(x)dx. \tag{5.1}$$

Вважатимемо, що щільності ймовірностей наших функцій підпорядковуються нормальному закону розподілу, тобто:

$$f_1(x) = (1/\sigma_1\sqrt{2\pi})e^{-(x-m_1)^2/(2\sigma_1^2)};\quad f_2(x) = (1/\sigma_2\sqrt{2\pi})e^{-(x-m_2)^2/(2\sigma_2^2)}. \tag{5.2}$$

Рішення про виявлення аномального значення приймається з використанням критерію Байєса та інтервалів довіри:

$$\frac{f_1(x)}{f_2(x)} > Q \text{ або } \frac{(x-m_1)^2/(2\sigma_1^2)}{(x-m_2)^2/(2\sigma_2^2} > Q, \tag{5.3}$$

де Q приймається експертно на етапі виявлення.

Оскільки критерій Байєса згладжує граничні значення, то додатково до критерію виконується перевірка на належність значення x, що спостерігається, до інтервалу можливих значень:

$$V \le x \le W \tag{5.4}$$

Ця модель параметра використовується для оцінки таких параметрів, як «нормальна» довжина пакета для поточної моделі та інших. Для оцінки інших числових параметрів цієї моделі недостатньо, тому були розроблені додатково дві більш складніші моделі числових параметрів: модель параметра потоків мережевого трафіку і модель параметра часових інтервалів мережевого трафіку. Розглянемо їх докладніше.

6 Модель параметра потоків мережевого трафіку

Майже у всіх об'єктів, що моделюються, можна виділити параметр, що характеризує обсяг вхідного і вихідного трафіку в одиницю часу. Побудуємо для цього параметра окрему модель і назвемо її «моделлю активності трафіку».

Трафік стосовно моделі може бути двох типів: вхідний потік (in flow) та вихідний (out flow).

З одного боку, сам собою підрахунок кількості переданих байт нам нічого не дасть. З іншого боку, при використанні інтелектуальних чи статистичних механізмів аналізу кількість переданих байт за одиницю часу набагато інформативніша. Наприклад, якщо раптом порт, який зазвичай активний тільки в робочі дні, раптом починає передавати дані у вихідні ввечері. Виходячи з викладених вище міркувань у профіль моделі активності трафіку включені наступні параметри:

–параметри вхідного потоку: - кількість переданих байт на годину, кількість переданих байт на день тижня, кількість переданих байт за день, кількість переданих байт за весь тиждень, кількість переданих байт на місяць;

- параметри вихідного потоку: - кількість переданих байт на годину, кількість переданих байт на день, кількість переданих байт на день тижня, кількість переданих байт на тиждень, кількість переданих байт на місяць.

Кожен із цих параметрів можна розглядати як числовий параметр та оцінювати одним із статистичних методів, які описані вище.

За допомогою побудованої моделі можна оцінювати активність як обсягів трафіку, так і активність кількості пакетів із заданим значенням параметра. Наприклад, збільшення кількості пакетів ICMP у мережі, в залежності від часу доби, може вказувати на встановлення ICMP-шелла.

Тому за допомогою цієї моделі числового параметра оцінюються такі параметри ключових понять: обсяг трафіку в байтах; кількість пакетів TCP; кількість пакетів ICMP; кількість пакетів UDP; загальна кількість пакетів за всіма протоколами.

7 Модель параметра часових інтервалів мережевого трафіку

Є ряд параметрів, які можна виділити у всіх об'єктів, що моделюються. Одним із таких параметрів є параметр, що характеризує час прибуття останнього пакета. Використовуючи цей параметр, можна знайти деякі види сканування портів. При їх використанні відправляються пакети на різні порти хосту, що атакуються через певні інтервали часу. Атака на відмову в обслуговуванні на web-сервері відкриває одночасно велику кількість з'єднань, тим самим переповнюючи чергу сервісу, що дозволяє виявляти такі атаки, з використанням часу прибуття останнього пакета. Під час встановлення веб-шелла інтервали між пакетами також будуть змінюватися.

Для виявлення атак, які описані вище, було запроваджено модель, що описує інтервали між пакетами: – модель часових інтервалів пакетів. Кожен із пакетів, що надходить по мережі, забезпечений тимчасовою міткою. Для цього в профілях об’єктів «Хост», «Порт» та «Сесія» введено два додаткові поля:

– час попереднього пакета (t_n) – це поле зберігає час прибуття останнього пакета і використовується для виявлення аномалій;

– інтервал до попереднього пакета (I_n) – це поле зберігає модель розподілу інтервалів часу між пакетами.

Для максимальної точності виявлення аномальних значень кожне поле зберігає відлік часу з точністю до мілісекунд.

Важливою частиною моделі є інтервали часу до попереднього пакета. Тимчасовий інтервал у свою чергою можна розглядати як числовий параметр і аналізувати за алгоритмом, який описаний вище, тобто, з урахуванням математичного очікування, дисперсії та інтервалу допустимих значень.

8 Категоріальні параметри

Категоріальні параметри не можна перевірити статистичними чи ймовірнісними методами, тому для них застосовується верифікація за правилами: – верифікаційне розпізнавання аномалій. Верифікаційне розпізнавання кожному параметру k_i^j ставить у відповідність правило R_i^j, де верхній індекс – індекс моделі (1 – мережа, 2 – хост, 3 – порт, 4 – сесія, 5 – пакет), нижній індекс – індекс параметра. Правило перевіряє належність параметра діапазону допустимих значень. У кожній ключовій моделі, крім моделі сесії, є хоча б один категоріальний параметр – це складовий параметр, з урахуванням правила якого вибирається відповідна модель нижнього рівня.

У моделі мережі є два категоріальні параметри. Перший категоріальний параметр описує діапазон адрес, що захищаються. Правило R_1^1 перевіряє, якому потоку належить пакет: вхідному, вихідному, чи це транзитний пакет. Другий параметр характеризує дерево хостів мережі, що захищається. Правило R_2^1 шукає відповідний профіль хоста, і якщо не знаходить його, повідомляє про аномалію.

Модель хоста має два категоріальні параметри. Відповідно до специфікації протоколу TCP у одному хості існує велика кількість різних портів.

Очевидно, що недоцільно зберігати профілі всіх портів. Тому модель поділяє порти на дві множини:

- серверні порти (порти, які постійно відкриті в системі та відповідають прикладним сервісам);

– клієнтські порти (порти, що динамічно генеруються системою).

На кожний серверний порт у системі заводиться окремий профіль. Для клієнтських портів моделі хоста формується загальний профіль, що характеризує загальну поведінку клієнтських портів. Перший параметр характеризує перелік серверних портів. Таким чином, при обробці серверного з’єднання правило R_1^2 шукає відповідний профіль порту для серверних з’єднань, і якщо не знаходить, генерує повідомлення про аномалію. Другий параметр характеризує загальний

профіль поведінки портів клієнтів. Правило R_2^2 ставить у відповідність клієнтським з'єднанням загальний профіль клієнтського порту.

Модель сесії, на відміну від моделей мережі і хоста, є динамічною, тобто, сесія не присутня в мережі постійно. Тому в моделі порту не зберігається профіль кожної сесії, а зберігається загальний профіль сесії. Модель порту включає два категоріальні параметри. Перший параметр – це список сесій, відкритих на даному порті на даний момент часу. Правило R_1^3 аналізованого пакету ставить у відповідність профіль сесії. Другий параметр – це загальний профіль сесії цього порту. Правило R_2^3 перетворює завершені сесії у загальному профілі сесії.

У моделях сесії та пакету є кілька категоріальних параметрів, тому розглянемо ці моделі докладніше.

9 Моделі сесії, пакета

Більшість сучасних протоколів прикладного рівня працюють з урахуванням протоколу TCP. Тому точна модель TCP сесії одна із ключових компонентів у описі «нормального поведінки» мережі. Як відомо, сесія однозначно ідентифікується наступним набором параметрів [3]: – IP адреса відправника; – IP адреса одержувача; – порт відправника; – порт отримувача.

Цей набір параметрів надалі і використовується як ідентифікатор сесій. Для обліку моделі сесії у системі до моделі порту було додано два складових параметри : список TCP сесій, аналізованих з першого пакета (таким чином, є можливість аналізу послідовності прапорів); список інших сесій. При побудові моделі сесії враховуються такі параметри: S_1 – послідовність прапорів; S_2 – вікно TCP сесії; S_3 – наявність опцій; S_4 – клієнт чи сервер; S_5 – кількість пакетів у сесії; S_6 – довжина пакетів; S_7 – кількість переданих даних у бітах за сесію; S_8 – співвідношення кількості переданих біт між клієнтом та сервером; S_9 – інтервали між пакетами.

Параметри сесії також розбиті на дві групи: категоріальні та числові. p align="justify"> Категоріальні параметри (S_1, S_2, S_3, S_4) аналізуються за допомогою набору верифікаційних правил, отриманих експериментально для кожного параметра: $R_i\,(s_i)$, i = 1...4. Числові параметри (S_5, S_6, S_7, S_8, S_9) аналізуються статистичними методами з використанням моделей, сформованих під час навчання та збережених у профілях P_i, i = 5...9.

Статистичні методи аналізу числових параметрів було описано вище, тому розглянемо докладніше аналіз категоріальних параметрів.

Верифікаційні правила сформовані експертно виходячи з аналізу специфікації протоколу TCP. Розглянемо їх докладніше.

Правило для аналізу послідовності прапорів R_1^4 ґрунтується на тому факті, що сесія TCP зазвичай має чітко визначену послідовність прапорів.

Під час встановлення сесії це SYN → SYN/ACK → ACK. Надсилання даних ACK →ACK/PUSH → …→ ACK. Завершення сесії FIN→ACK→FIN→ACK. Є й інші, допустимі стандартом послідовності прапорів, але вони у нормальній роботі мережі зустрічаються дуже рідко. Тому в побудованій моделі для аналізу параметра прапорів сесії S_1 був реалізований автомат, який перевіряє відповідність прапорів поточного пакета стану прапорів поточної сесії.

$$R_1^4(S_1 \mid S_1^{j-1}, S_1^{j-2}, ..., S_1^{j-k}, \tag{5.5}$$

де, S_1 – поточне значення першого параметра сесії для j-го пакета у цій сесії, а $R_1^4(S_1 \mid S_1^{j-1}, S_1^{j-2}, ..., S_1^{j-k}$, – значення першого параметра сесії для попередніх пакетів цієї сесії.

Правило R_2^4 здійснює аналіз параметра вікна TCP-сесії. При невдалій атаці, спрямованої на перехоплення чужої TCP-сесії, а також за деяких інших видів атак зловмиснику не вдається потрапити у вікно цієї сесії. При нормальній роботі більшості програм це зустрічається вкрай рідко. Тому пакети, які не потрапляють у вікно сесії, ми вважаємо аномальними.

$$R_2^4(S_2 \mid S_2^{j-1}, S_2^{j-2}, ..., S_2^{j-k}, \tag{5.6}$$

де, $k \in (3...5)$.

Правило призначене для аналізу параметрів опцій TCP-сесії. Це правило ґрунтується на знаннях про те, що опції зазвичай присутні лише в 1-му або 3-му пакеті сесії.

Параметр – клієнтська це сесія або серверна – сам по собі не аналізується, але він використовується при верифікаційному аналізі інших параметрів моделі. Коли у модель порту потрапляє пакет з прапором SYN і ідентифікатор цієї сесії відсутній у обох складових параметра порту, то тоді вона додається у відповідний список і сесія вважається клієнтською. Зокрема, цей параметр використовується для аналізу параметра s_1 послідовності прапорів (де враховується у який стан може перейти сервер і в яке – лише клієнт), а також, при аналізі числового параметра s_8 (співвідношення кількості переданих біт між клієнтом і сервером).

Верифікаційні правила моделі пакета перевіряють коректність полів заголовків пакетів IP та TCP. Правила формуються експертно з урахуванням специфікацій. Зокрема, перевіряються такі параметри: правило R_1^5 перевіряє, щоб IP адреса джерела не дорівнювала IP адресі одержувача; правило R_2^5 перевіряє, щоб порти джерела та одержувача не дорівнювали нулю; правило R_3^5 призначене для виявлення TCP пакетів з неприпустимими значеннями прапорів; правило R_4^5 перевіряє, щоб зарезервовані поля були заповнені нульовими значеннями; правило R_5^5 перевіряє, щоб довжина пакета була більшою або дорівнювала 40 байтам.

10. Висновки

Розроблена модель системи дозволяє виявляти атаки на ключові об’єкти, що моделюються. Отримані під час випробувань результати показали високу ефективність виявлення аномалій числових параметрів моделі. Для збільшення точності надалі планується перейти до моделювання поняття «сервіс» та моделювання протоколів HTTP, SMTP, POP3. Модель сесії також дозволяє виявляти існуючі та нові атаки на рівні сесії TCP, а також деякі види атак на відмову в обслуговуванні. Модель потоків мережевого трафіку дозволяє нам виявляти такі види атак, як: різні види сканування системи, установку троянських програм (бо зросте кількість байт у вихідному потоці), установку ICMP шелла (бо зросте кількість ICMP пакетів). Модель часових інтервалів дозволяє виявляти деякі види сканування системи, атаки на відмову в обслуговуванні та встановлення web-шеллу.

Стаття присвячена опису моделі виявлення мережевих вторгнень у трафіку мереж, побудованих з урахуванням стека протоколів TCP/IP. Аналізуються основні об'єкти локальної обчислювальної мережі. Описуються основні контрольовані параметри кожного типу об'єктів. Розробляються методи пошуку аномалій, що ґрунтуються як на аналізі за правилами, так і на аналізі з використанням імовірнісних моделей.

СПИСОК ЛІТЕРАТУРИ

1. Рубан І. В. Класифікація методів виявлення аномалій в інформаційних системах / Рубан І. В., Мартовицький В. О., Партика С. О. // Системи озброєння і військова техніка. – 2016. – №. 3. – С. 100-105. https://openarchive.nure.ua/server/api/core/bitstreams/7c434471-942c-40a7-b70c-0cc2655a42fe/content

2. Мірошник М.А. Модель мережевого планування із застосуванням методів мультипаралельної обробки інформації / Мірошник М.А., Толстолузький Є.Д. // тези 23 Міжнародна науково-технічная конференція "Проблеми інформатики та моделювання", Харків: НТУ "ХПІ", 2023. с. 75-76. https://repository.kpi.kharkov.ua/items/480d4c7b-d463-49dc-8521-c1162b16db88

3. Мирошник, М. А. Методы автоматизированного проектирования гетерогенных компьютерных систем и сетей критического применения / М. А. Мирошник, А. А. Можаев // Інформаційно-керуючі системи на залізничному транспорті. 2019. № 4. С.40-46.

DOI: https://doi.org/10.18664/ikszt.v0i4.178719.4.

4. Мирошник М. А. Синтез легкотестируемых двумерных сетей / М. А. Мирошник, Я. Ю. Королева // Інформаційно-керуючі системи на залізничному транспорті : тези стендових доповідей та виступів учасників 31-ї міжнародної науково-практичної конференції "Інформаційно-керуючі системи на залізничному транспорті" (Харків, 24-26 жовтня, 2018 р.). – 2018. – № 4 (додаток). – С. 20-21. http://lib.kart.edu.ua/handle/123456789/11689

5. Коробейнікова Т.І. Аналіз сучасних відкритих систем виявлення та запобігання вторгнень / Т.І. Коробейнікова, О.О. Цар // Національний університет «Львівська політехніка», Україна. May 2023, Грааль науки. С.317-325. DOI:10.36074/grail-of-science.12.05.2023.050, License, CC BY-SA 4.0

6. Лук'яненко Т. Ю. Методика виявлення мережевих вторгнень і ознак комп'ютерних атак на основі емпіричного підходу. / Лук'яненко Т. Ю., Поночовний П. М., Легомінова С. В. // Сучасний захист інформації. – № 2 (2022). – с.15-21. DOI: 10.31673/2409-7292.2022.021521

7. Чемерис К. М., Дейнега Л. Ю. Застосування методу вейвлет-аналізу для виявлення атак в мережах. Наука і техніка Повітряних Сил Збройних Сил України. 2022. № 1(46). С. 99-107. https://doi.org/10.30748/nitps.2022.46.14.

8. Панченко М.В. Виявлення аномалій інформаційної безпеки на основі аналізу ентропії інформаційної системи / М. В. Панченко, А. М. Бігдан, Т. В. Бабенко, Д. С. Тимофєєв. Енергетика і автоматика", №1, 2022 р. DOI 10.31548/energiya

http://journals.nubip.edu.ua/index.php/Energiya/article/viewFile/energiya2022.01.072/14743

9. Толюпа С. Засоби виявлення кібернетичних атак на інформаційні системи / С. Толюпа, Н. Лукова-Чуйко, Я. Шестак. Інфокомунікаційні технології та електронна інженерія. - №2 (2). 2021, стр. 19-31. https://science.lpnu.ua/sites/default/files/journal-paper/2022/mar/27268/stattya3stolyupanlukova-chuykoyashestak.pdf

10. . Нічепорук А.О. Інтелектуальна система виявлення аномалій та ідентифікації пристроїв розумних будинків із застосуванням колективної комунікації / А.О. Нічепорук, А.А. Нічепорук, О.С. Савенко, А.Д. Казанцев. Хмельницький національний університет // ISSN 2221-3805. Електротехнічні та комп'ютерні системи. 2021. № 34 (110) Інформаційні системи та технології Users/Administrator/Downloads/3196-Article Text-2350-1-10-20210904.pdf

11. Аналіз систем та методів виявлення несанкціонованих вторгнень у комп'ютерні мережі [Електронний ресурс] / В. В. Литвинов [та ін.] // Математичні машини і системи. К : ІПММС НАН України, 2018. № 1. С. 31-40. http://dspace.nbuv.gov.ua/handle/123456789/132008 .

12. Терейковський І. Моделі еталонів лінгвістичних змінних для систем виявлення email-спуфінг-атак / І. Терейковський, А. Корченко, П. Вікулов, І. І. Дж. Ірейфідж // Безпека інформації. - 2018. - Т. 24, № 2. - С. 99-109.

Режим доступу: http://nbuv.gov.ua/UJRN/bezin_2018.

REFERENCES

1. Ruban I. V. Classification of anomaly detection methods in information systems / Ruban I. V., Martovytskyi V. O., Partika S. O. // Armament systems and military equipment. – 2016. – no. 3. – pp. 100-105. https://openarchive.nure.ua/server/api/core/bitstreams/7c434471-942c-40a7-b70c-0cc2655a42fe/content [in Ukrainian]

2. . Miroshnyk M.A. A model of network planning with the use of multiparallel information processing methods / M.A. Miroshnyk, E.D. Tolstoluzkyi. // theses 23 International Scientific and Technical

Conference "Problems of Informatics and Modeling", Kharkiv: NTU "KhPI", 2023. – pp. 75-76. https://repository.kpi.kharkov.ua/items/480d4c7b-d463-49dc-8521-c1162b16db88 [in Ukrainian]

3. Miroshnyk M.A.. Methods of automated design of heterogeneous computer systems and networks of critical application / Miroshnyk M.A., A.A. Mozhaev // Information and control systems on railway transport. – 2019. – No. 4. – P.40-46. DOI: https://doi.org/10.18664/ikszt.v0i4.178719.4 [in Ukrainian]
4. Miroshnyk M.A. Synthesis of easily testable two-dimensional networks / M. A. Myroshnyk, Y. Yu. Koroleva // Information and control systems in railway transport: abstracts of poster reports and speeches of the participants of the 31st international scientific and practical conference "Information and control systems in railway transport" (Kharkiv, October 24-26, 2018). – 2018. – No. 4 (appendix). - pp. 20-21. http://lib.kart.edu.ua/handle/123456789/11689 [in Ukrainian]
5. Korobeynikova T.I. Analysis of modern open intrusion detection and prevention systems / T.I. Korobeynikova, O.O. Tsar // Lviv Polytechnic National University, Ukraine. May 2023, the grail of science. pp. 317-325.

   DOI:10.36074/grail-of-science.12.05.2023.050, License, CC BY-SA 4.0 [in Ukrainian]
6. Lukyanenko, T. Yu. Methodology for detecting network intrusions and signs of computer attacks based on an empirical approach. / Lukyanenko T. Yu., Ponochevny P. M., Legominova S. V. // Modern protection of information. – No. 2 (2022). - pp. 15-21.DOI: 10.31673/2409-7292.2022.021521 [in Ukrainian]
7. Chemeris K. M., Deinega L. Yu. Application of the wavelet analysis method to detect attacks in networks. Science and technology of the Air Force of the Armed Forces of Ukraine. 2022. No. 1(46). pp. 99-107. https://doi.org/10.30748/nitps.2022.46.14 [in Ukrainian]
8. M.V. Panchenko Identification of information security anomalies based on information system entropy analysis / M. V. Panchenko, A. M. Bigdan, T. V. Babenko, D. S. Timofeev. Energy and automation", No. 1, 2022. DOI 10.31548/energiya [in Ukrainian]

   http://journals.nubip.edu.ua/index.php/Energiya/article/viewFile/energiya2022.01.072/14743
9. Tolyupa S. Means of detecting cybernetic attacks on information systems / S. Tolyupa, N. Lukova-Chuiko, Ya. Shestak. Information communication technologies and electronic engineering. - #2 (2). 2021, pp. 19-31. [in Ukrainian]

   https://science.lpnu.ua/sites/default/files/journal-paper/2022/mar/27268/stattya3stolyupanlukova-chuykoyashestak.pdf [in Ukrainian]
10. Nicheporuk A.O. An intelligent system for detecting anomalies and identifying devices of smart buildings using collective communication / A.O. Nicheporuk, A.A. Nicheporuk, O.S. Savenko, A.D. Kazantsev. Khmelnytskyi National University // ISSN 2221-3805. Electrical and computer systems. 2021. No. 34 (110) Information systems and technologies Users/Administrator/Downloads/3196-Article Text-2350-1-10-20210904.pdf [in Ukrainian]
11. Analysis of systems and methods for detecting unauthorized intrusions into computer networks [Electronic resource] / V.V. Litvinov [et al.] // Mathematical machines and systems. K: IPMMS of the National Academy of Sciences of Ukraine, 2018. No. 1. P. 31-40.

    http://dspace.nbuv.gov.ua/handle/123456789/132008 . [in Ukrainian]
12. I. Tereykovskyi. Models of standards of linguistic variables for email-spoofing-attack detection systems / I. Tereykovskyi, A. Korchenko, P. Vikulov, I. I. J. Ireifidge // Information security. - 2018. - Vol. 24, No. 2. - P. 99-109. - Access mode: http://nbuv.gov.ua/UJRN/bezin_2018. [in Ukrainian]

Demenkova Svitlana *Senior lecturer of the department of automation of chemical-technological systems and environmental monitoring, National Technical University, Kharkiv Polytechnic Institute, Kharkiv, Kirpychova St.,2, 61002*
e-mail: svet1972232765@gmail.com
https://orcid.org/0000-0003-0604-5456

Demchenko Kateryna *associate professor of the Department of Automation and Computer-Integrated Technologies, State Biotechnology University, str. Alchevsky 44, Kharkiv, Ukraine, 61002*

e-mail: yayaska@btu.kharkiv.ua
https://orcid.org/0000-0002-3168-5351

Koroleva Yana *associate professor of the department of multimedia and Internet technologies and systems, National Technical University, Kharkiv Polytechnic Institute, Kharkiv, Kirpychova St.,2, 61002*
e-mail: yanakoroleva815@gmail.com
https://orcid.org/0000-0003-0604-5456

Pakhomov Yurii *associate Professor of the Department of Computer Sciences and Information Technologies, Beketov Kharkiv National University of Urban Economy Kharkiv, str. 17, Marshala Bazhanov, Ukraine, 61002*
e-mail: abc050073@gmail.com
https://orcid.org/0000-0002-2267-8600

Using anomaly detection method to detect network attacks

The article is focused on the description of a model for detecting network intrusions in the network traffic based on the TCP/IP protocol stack. The main objects of a local area network have been analyzed. The main controlled parameters of each type of object have been described. The methods of anomaly detection based on both rule-based and probabilistic model analysis have been developed.

Relevance. Due to the intensive growth of information technologies and their implementation in various sectors of the municipal economy, the issue of information security becomes very relevant.

Research methods. In solving the tasks, the methods of control theory; methods of building security systems; graph theory; probability theory and mathematical statistics; methods of time series analysis; methods of predictive analytics and big data processing; methods of building high-load secure programs have been used. The main research methods used are probabilistic and verification modeling.

The results. Probabilistic and verification modeling of network attacks has confirmed the effectiveness of the proposed approach. The results of synthesis using CAD showed that the additional hardware costs do not exceed 20% compared to the standard description model.

Conclusions. The developed system model allows detection of attacks on key simulated objects. The results obtained during the tests showed a high efficiency of detecting anomalies in the numerical parameters of the model. In order to increase accuracy, it is planned to move on to the modeling of the concept of "service" and the modeling of HTTP, SMTP, and POP3 protocols. The session model also allows detecting existing and new TCP session-level attacks, as well as some types of denial-of-service attacks. The model of network traffic flows allows us to detect such types of attacks as: various types of system scanning, installation of Trojan programs (because the number of bytes in the output stream will increase), installation of ICMP shell (because the number of ICMP packets will increase). The time interval model allows detecting some types of system scanning, denial-of-service attacks, and web shell installation.

Keywords: *anomaly detection, network attacks, finite state machines, verification, modeling, distributed information networks, probabilistic modeling, verification modeling.*

Надійшла у першій редакції 01.11.2023, в останній – 02.12.2023.
The first version has been received on 01.11.2023, the final version – on 02.12.2023.