УДК 004

Степан ГАЛИЛУЙКО, Володимир ДРАПАК, Людмила БАБАЛА

Західноукраїнський національний університет, м. Тернопіль, Україна

ПРОЄКТУВАННЯ СИСТЕМИ ВИЯВЛЕННЯ АНОМАЛІЙ У МЕРЕЖЕВОМУ ТРАФІКУ НА ОСНОВІ ШТУЧНОГО ІНТЕЛЕКТУ

Вступ. Система виявлення аномалій у мережевому трафіку, заснована на штучному інтелекті (ШІ), є важливою складовою сучасних заходів кібербезпеки. У зв'язку зі зростанням кіберзагроз та складністю інфраструктур, впровадження інноваційних методів для автоматизованого моніторингу та аналізу трафіку є вкрай актуальним [1].

Мета: розробка системи виявлення аномалій у мережевому трафіку на основі глибокого навчання, яка інтегрує різні технології обробки даних для аналізу поведінкових патернів, прогнозування аномальних подій та оперативного реагування на загрози.

1. Аналіз предметної області та технологій

В сучасних умовах стрімкого зростання обсягів мережевого трафіку та складності кіберзагроз, забезпечення безпеки мереж стало одним із ключових завдань. Однією з найважливіших задач у цій сфері є виявлення аномалій у мережевому трафіку, які можуть сигналізувати про можливі атаки, збої або несанкціоновану активність. Традиційні системи виявлення вторгнень (IDS) здебільшого базуються на сигнатурних методах, що обмежує їх ефективність перед новими, невідомими загрозами [2].

Аномалії в мережевому трафіку можуть проявлятися у вигляді змін у розподілі даних, атипових пакетів або нехарактерної поведінки користувачів. Виявлення таких відхилень потребує використання складних методів аналізу, здатних ідентифікувати приховані патерни та тенденції. З цією метою дедалі частіше застосовуються технології штучного інтелекту (ШІ), зокрема методи машинного навчання та глибинного навчання.

Наразі існують такі підходи до виявлення аномалій:

1. Сигнатурні методи - базуються на порівнянні трафіку з базою відомих загроз. Їхня обмеженість полягає у нездатності розпізнавати нові або модифіковані атаки.
2. Поведінкові методи - аналізують аномалії на основі статистичних моделей або правил, однак вони можуть генерувати значну кількість хибнопозитивних спрацьовувань.
3. Методи машинного навчання - включають класифікацію, кластеризацію та побудову прогнозів на основі навчання на великих наборах даних.

Популярними інструментами для аналізу трафіку є такі платформи, як Wireshark, Snort, а також сучасні рішення на основі ШІ, зокрема Zeek та Darktrace, які інтегрують машинне навчання для автоматизованого аналізу аномалій.

Основними викликами є:

- обробка великих обсягів даних у реальному часі;

- висока частота хибнопозитивних спрацьовувань;
- складність інтерпретації результатів роботи моделей ШІ.

Таким чином, аналіз предметної області дозволяє визначити необхідність у розробці ефективних систем, які поєднують автоматизацію, масштабованість та здатність до самонавчання для виявлення аномалій у мережевому трафіку.

2. Розробка системи виявлення аномалій у мережевому трафіку на основі штучного інтелекту

Розробка моделі для виявлення аномалій у мережевому трафіку на основі технологій штучного інтелекту спрямована на подолання обмежень традиційних методів виявлення, таких як висока частота хибнопозитивних спрацьовувань і обмежена здатність до адаптації до нових типів загроз. Цей підхід має на меті створення гнучкої та ефективної системи для моніторингу мережевого трафіку в реальному часі, здатної до самонавчання та вдосконалення в умовах постійно змінюваного середовища кіберзагроз.

Модель виявлення аномалій у мережевому трафіку базується на використанні методів машинного навчання для побудови прогностичної моделі, яка вивчатиме поведінку трафіку та виявлятиме відхилення від нормального стану.

На першому етапі зібрані дані з мережі очищуються від шуму та непотрібної інформації, застосовуються методи нормалізації та кодування категоріальних ознак для приведення всіх даних до єдиного формату. Для зменшення розмірності даних використовуються методи, такі як техніка головних компонент (PCA), що дозволяє виділити найбільш значущі характеристики трафіку [3].

Для вирішення задачі виявлення аномалій застосовується метод глибинного навчання - автокодери. Цей алгоритм здатен навчатися на звичайному (неанотованому) трафіку, вивчаючи його структуру і виявляючи відхилення, які не відповідають знайденим патернам. Автокодери мають здатність до самонавчання та адаптації до нових типів аномалій без необхідності значних зусиль на підготовку даних [4].

Модель тренується на великому наборі нормального трафіку, що дозволяє їй створити сприйнятливість до типових поведінкових патернів. Після навчання система може виявляти відхилення на основі обчислення реконструкційної помилки - різниці між оригінальним і реконструйованим трафіком. Якщо помилка перевищує поріг, система сигналізує про наявність аномалії.

Модель перевіряється на тестовому наборі даних, що містить як нормальний трафік, так і відомі аномалії. Оцінка ефективності здійснюється через показники точності, чутливості, специфічності та хибнопозитивних спрацьовувань. Для підвищення ефективності можна застосовувати техніки перенавчання на нових аномальних даних, що дозволяє адаптувати систему до нових типів атак.

Алгоритм виявлення аномалій передбачає збір даних через інтеграцію з мережевими пристроями та системами моніторингу. Дані проходять через фільтрацію, нормалізацію та кодування. Навчання моделі здійснюється через багатошарову нейронну мережу, де вхідні дані (мережеві пакети) проходять через

шари кодування і декодування. Модель намагається мінімізувати різницю між вхідними і вихідними даними.

Після отримання результату декодування обчислюється реконструкційна помилка, і якщо вона перевищує заданий поріг, система сигналізує про потенційну аномалію. Для зниження рівня хибнопозитивних спрацьовувань застосовуються додаткові стратегії на основі кластеризації або аналізу часових рядів, що дозволяє фільтрувати маловажливі аномалії, які можуть бути викликані незначними змінами в мережевому середовищі.

Для реалізації цієї моделі використовуються сучасні бібліотеки машинного навчання, такі як:

- TensorFlow;
- PyTorch.

А також інструменти для збору та аналізу даних з мережі, зокрема:

- Wireshark;
- Zeek;
- Suricata.

Очікуваним результатом є система, здатна ефективно виявляти аномалії в реальному часі та адаптуватися до нових типів атак, знижуючи кількість хибнопозитивних спрацьовувань. Перспективи подальшого вдосконалення цієї системи включають інтеграцію з іншими системами безпеки та застосування методів глибинного навчання для обробки складних залежностей у трафіку. Такий підхід дозволяє значно підвищити точність виявлення загроз і зменшити необхідність у ручному втручанні, що є критичним для сучасних автоматизованих систем безпеки [5].

Висновок. Використання штучного інтелекту для виявлення аномалій у мережевому трафіку є ефективним методом підвищення кібербезпеки, дозволяючи виявляти нові загрози та зменшувати хибнопозитивні спрацьовування. Хоча існують певні обмеження, зокрема необхідність великих даних для навчання, ці технології є потужним інструментом для захисту мереж. Подальші дослідження можуть зосередитись на оптимізації алгоритмів та інтеграції з іншими системами безпеки.

Перелік використаних джерел.

1. Zhang, Y., Wang, L. (2020). Network Anomaly Detection with Machine Learning. Journal of Cybersecurity and Privacy, 5(3), 101-115.

2. Ahmed, M., Mahmood, A. N., Hu, J. (2016). A survey of network anomaly detection techniques. International Journal of Computer Applications, 133(3), 1-15.

3. Chen, J., Li, X. (2019). Deep Learning for Anomaly Detection in Network Traffic. Proceedings of the IEEE International Conference on Computer Networks, 45-52.

4. Nguyen, T., Tran, T. (2018). Deep Autoencoders for Network Anomaly Detection: A Review. Computational Intelligence and Neuroscience, 2018, 1-9.

5. Bishop, Christopher. (2006). Pattern Recognition and Machine Learning. 10.1117/1.2819119.