УДК 004.75

СИСТЕМА ВИЯВЛЕННЯ ВТОРГНЕНЬ НА ОСНОВІ МАШИННОГО НАВЧАННЯ

Тимощук Ю.В.
Тернопільський національний економічний університет, магістрант

I. Постановка проблеми

У зв’язку зі збільшенням обсягів інформації, що поширюється в обчислювальних мережах та розширенням кількості завдань, які вирішуються за допомогою інформаційних систем, виникають проблеми, пов'язані із ростом кількості загроз та підвищенням вразливості інформаційних ресурсів.

Система виявлення вторгнень (Intrusion Detection System, IDS) це пристрій або програмне забезпечення, яке здійснює моніторинг діяльності мережі або системи та виявляє, чи відбувається та чи інша зловмисна активність. Значне зростання використання мережі Інтернет викликає занепокоєння щодо безпечного спілкування та захисту цифрової інформації. В даний час хакери використовують різні типи атак для отримання необхідної інформації. Багато із цих методів та алгоритмів допомагають виявляти атаки.

Комп’ютерна та мережева безпека набуває все більшого значення із збільшенням кількості атак, орієнтованих на конфіденційність, цілісність та доступність даних. Вторгнення націлюються на окрему мережу чи організацію для крадіжки даних. Для виявлення вторгнення у ту чи іншу систему, було зроблено багато схем та витрачено чимало зусиль [1].

Вторгнення означає будь-який несанкціонований доступ або шкідливе використання інформаційних ресурсів. Зловмисник є об'єктом реального світу, який намагається знайти засоби для отримання несанкціонованого доступу до інформації, заподіяння шкоди чи інших шкідливих дій.

Процеси, пов'язані із машинним навчанням, подібні до процесів інтелектуального аналізу даних та інтелектуального моделювання. Обидва вимагають пошуку даних та регулювання дій програми. Багато людей знайомі із машинним навчанням від покупок у мережі Інтернет та подачею оголошень, пов'язаних із їх придбанням.

Це відбувається тому, що програмні модулі із рекомендаціями використовують машинне навчання, щоб персоналізувати доставку оголошень у режимі реального часу. Крім персонального маркетингу, інші поширені випадки використання машинного навчання включають виявлення шахрайства, фільтрацію спаму, виявлення загроз мережної безпеки, інтелектуальне обслуговування та створення каналів новин.

II. Мета роботи

Основна мета даної роботи полягає у розробці системи виявлення вторгнень на основі машинного навчання.

III. Виявлення вторгнень на основі машинного навчання

Алгоритми машинного навчання широко застосовуються для виявлення аномалій. Ці алгоритми будують модель виявлення або модель прогнозування у фазі навчання алгоритмом на основі даних тренінгу. Після чого модель прогнозування випробовується на нових даних на етапі тестування, щоб розмежувати доброякісні дані та зловмисні атаки.

Системи виявлення вторгнень класифікуються на дві категорії: мережеву систему виявлення вторгнень та систему виявлення вторгнень на основі хоста, яка базується на джерелі даних [2]. Аналізуючи джерело набору даних, IDS сигналізує тривогу, коли виявляє вторгнення або атаку. Система виявлення вторгнень на основі хоста (Host-basedintrusiondetectionsystem, HIDS) налаштована на єдину систему, яку також називають цільовою системою, що схильна до атаки.

Вхідні дані потребують попередньої обробки, щоб зрозуміти алгоритми машинного навчання. Вхідні дані містять приклади, які також називаються екземплярами, спостереженнями або записами. Кожен запис представлений деякими атрибутами чи ознаками, які також називаються функціями векторів. Існує декілька примірників, які позначені типом класу.

Алгоритм, якому потрібні дані, повністю позначається міткою класу та називається контрольованим алгоритмом, хоча є деякі алгоритми, які не потребують даних, позначених класом, і називаються алгоритмами навчання без нагляду (неконтрольованими) [3].

Непідконтрольні алгоритми не потребують, щоб усі записи були позначені класом. Такий метод навчання знаходить приховану закономірність у даних, тоді як алгоритм керованого навчання

знаходить зв’язок між даними та його класом. У даній роботі розглянуто алгоритми навчання, що контролюються.

Як було сказано раніше, контрольований алгоритм навчається на наборі прикладів, які позначаються його класовим типом. Алгоритм будує модель виявлення аномалії на фазі навчання, а потім вона перевіряється на нових даних, для того щоб побачити продуктивність цієї моделі [4, 5], як показано на рисунку 1. На етапі тестування дана модель присвоює мітку класу новим даним.

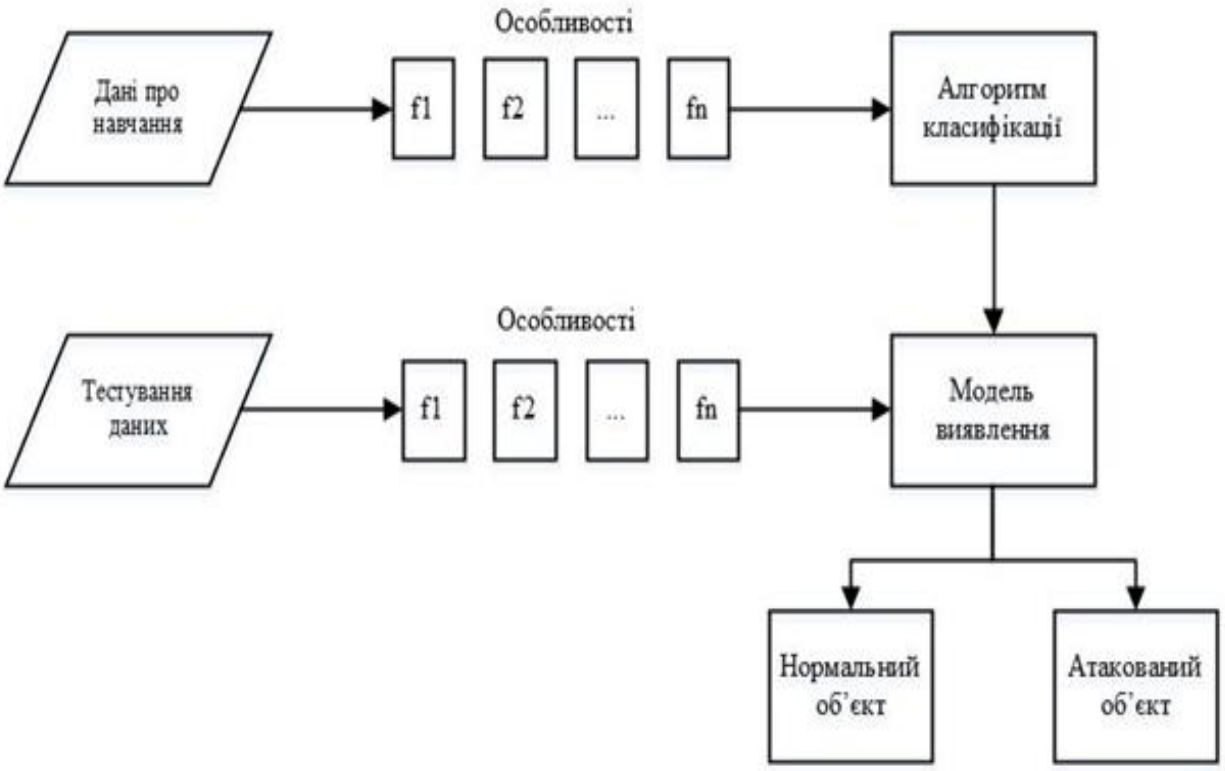

Рисунок 1 – Контрольований підхід до виявлення аномалії

Безпека є ключовим питанням як для комп’ютера, так і для комп'ютерної мережі. Недосконалі системи виявлення вторгнень є однією з основних проблем дослідження мережевої безпеки.

Висновок

У роботі було проведено аналіз алгоритму для систем виявлення та запобігання вторгнень. Багато ресурсів використовуються у різних техніках машинного навчання. Ці прийоми працюють дуже добре для IDS, але відомо, що не існує єдиного методу, який дозволяє виявити всі типи атак.

Тому для підвищення продуктивності роботи все ще потрібно удосконалювати методи машинного навчання для виявлення всіх видів атак. Помилкових сигналів тривоги повинно бути менше і також потрібно працювати над підвищенням ефективності роботи алгоритму вибору функцій.

Список використаних джерел

1. Ganapathy S., Kulothungan K., Muthurajkumar S., Vijayalakshmi M., Yogesh P., and Kannan A., “Intelligent feature selection and classification techniques for intrusion detection in networks: a survey,” EURASIP Journal on Wireless Communications and Networking, vol. 2013, pp. 16, 2013/11/27, 2013.
2. Davis and A. J. Clark, “Data preprocessing for anomaly based network intrusion detection: A review,” Computers & Security, vol. 30, pp. 353, 2011.
3. Maloof M. A., Machine learning and data mining for computer security: Springer, 2006.
4. Tsai J. J. and Philip S. Y. Machine learning in cyber trust: security, privacy, and reliability: Springer Science & Business Media, 2009.
5. Chio, C., Freeman, D.MachineLearningandSecurity: ProtectingSystemswithDataandAlgorithms." O'ReillyMedia, Inc.", 2018.