Резултат (Result) 2.1

Отчет със сравнителен анализ на изследването за откриване и анализ на аномалии в софтуерни продукти

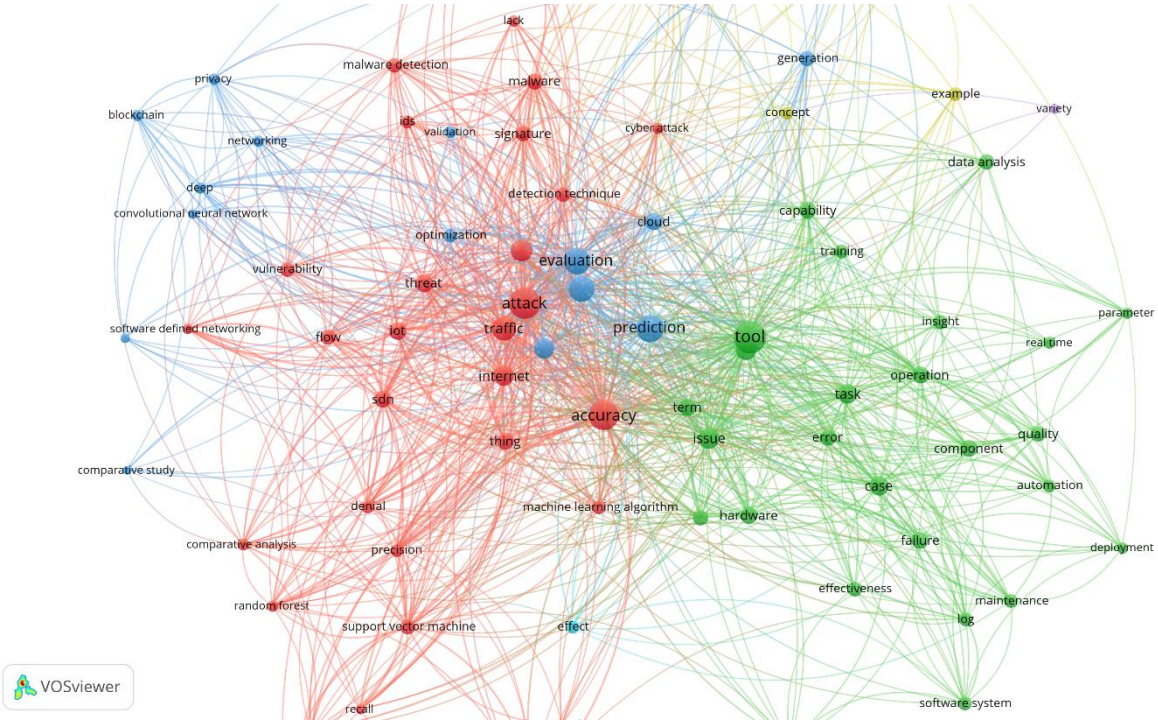
1. Методология на изследването.

Методологията на изследването за откриване и анализ на аномалии в софтуерни продукти включва няколко стъпки:

1.) Търсене на научни публикации в различни източника с търсене по тема, анотация и ключови думи чрез използване на дигитални библиотеки. Заявката за търсене беше:

```
TITLE-ABS-KEY((''detection'') AND (''analysis'') AND
(''anomalies'') AND (''machine learning'')
```

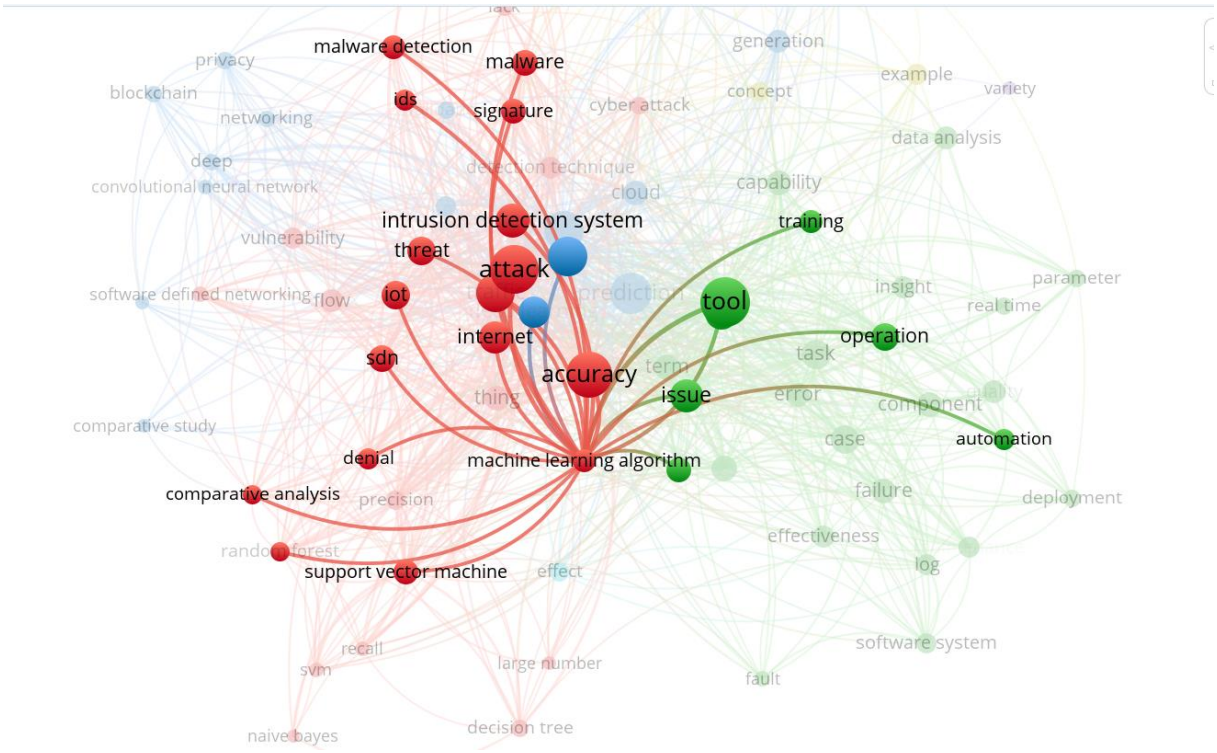
2.) След това създаваме подобна карта въз основа на индексираните ключови думи (фиг. 1). Те се различават от ключовите думи, предоставени от авторите, и предоставят повече възможности за наблюдение и заключения по време на анализа.



Фиг. 1 Индексна карта на ключовите думи.

Открити са 11102 понятия, от които 189 се повтарят повече от 10 пъти. За всеки от 189-те понятия се изчислява оценка на релевантността. Въз основа на тази оценка са избрани най-подходящите термини. По подразбиране се избират 60% от най-подходящите термини (113 понятия).

При изследване на индексната ключова дума „machine learning algorithm“ се изследват връзките въз основа, на който ще се направи анализа (фиг. 2).



Фиг. 2 Индексна карта на ключовата дума "machine learning algorithm" .

2. Откриване и анализ на аномалии в софтуерни продукти

Софтуерната сигурност се отнася до процеса на създаване и разработване на софтуер, който гарантира целостта, поверителността и наличността на неговия код, данни и услуги[1]. Откриването на аномалии в софтуерните продукти е важен аспект от поддържането на целостта и производителността, както и от осигуряването на сигурността и надеждността на софтуерните системи. Откриването и анализът на аномалии могат да помогнат на разработчиците на софтуер да идентифицират и отстранят грешки, да подобрят производителността на системата и да подобрят мерките за сигурност, за да предотвратят кибератаки и други неочаквани събития.

Аномалия в софтуерна система е всяко неочаквано или необичайно поведение, модел или събитие, което се отклонява от обичайната или очакваната работа на системата. В сравнение със софтуерната грешка, дефинирана като отклонение на системата от доставката, услугата или очаквания резултат[2], аномалиите могат да се проявят в различни форми, като необичайно поведение на системата, грешки в дневниците или базите данни, неочаквани изходи, бавна работа на системата или необичайна дейност на потребителя, и могат да показват потенциални проблеми, грешки или заплахи за сигурността в софтуера. Аномалиите могат да бъдат причинени от различни фактори, като грешки в софтуера, хардуерни неизправности, промени в

[1] R. Khan, S. Khan, H. Khan, M. Ilyas, Systematic mapping study on security approaches in secure software engineering, IEEE Access, Vol. 9, pp. 19139-19160, 2021.
[2] P. Kumar, A Wahid, Investigation of software reliability prediction using statistical and machine learning methods, Cognitive Analytics: Concepts, Methodologies, Tools, and Applications, Information Resources Management Association, IGI Global, 2020, pp. 1640-1660

околната среда, кибератаки или други неочаквани събития. Някои често срещани видове аномалии в софтуерните системи включват аномалии на сигурността, аномалии на производителността, функционални аномалии, аномалии на данните, аномалии на използването, аномалии на комуникацията. Аномалиите могат да имат сериозни последици, като например прекъсване на работата на системата, нарушаване на сигурността на данните, загуба на чувствителна информация и финансови загуби.

През последните години нарастващ брой платформи и софтуерни програми използват съхранени в хранилища набори от данни и отдалечен достъп. Откриването на аномалии в извличането на данни се отнася до идентифицирането на събития или наблюдения, които не следват очакван модел или други елементи в колекцията. Следователно наборите от данни са по-изложени на злонамерени атаки. Така че мрежовата сигурност придобива все по-голямо значение като изследователска област. Добре известен метод за защита на компютърните мрежи е използването на системи за откриване на проникване (IDS)[3]. Един от проблемите на тези системи за сигурност е фалшивата аларма за проникване в мрежата и точността на откриване на проникване, което се случва поради големия обем мрежови данни.

Откриването на аномалии във високоразмерни данни се превръща във фундаментален изследователски проблем, който има различни приложения в реалния свят. Много съществуващи техники за откриване на аномалии не успяват да поддържат достатъчна точност, тъй като големите данни се характеризират с голям обем и висока скорост, генерирани от различни източници[4].

Увеличаването на обема и разнообразието от данни, както и увеличеният трафик в интернет, водят до потенциално намаляване на устойчивостта на киберфизическите системи, особено на критичните инфраструктури. Внедряването на ефективна система за откриване на проникване за събиране на данни от сензори е изключително важно. Поради това са проведени сравнителни проучвания за използването на управлявани от изкуствен интелект системи за откриване на проникване за безжично свързани сензори, които проследяват важни приложения.

Система, която може да се учи от данни, може да бъде внедрена с помощта на техники за машинно обучение. Например, система за машинно обучение може да бъде научена да прави разлика между нетипични и типични пакети, използвайки пристигащи пакети като данни за обучение. След това може да се използва за категоризиране на новопристигащите пакети в нетипични и редовни пакети след обучение.

Представен е задълбочен анализ на използването на решения за машинно обучение, дълбоко обучение и обучение за подсилване за разпознаване на злонамерено поведение в събрания трафик[5]. Предложените механизми се оценяват чрез извършване

[3] K. Khan, A. Mehmood, S. Khan, M. A. Khan, Z. Iqbal and W. K. Mashwani, "A survey on intrusion detection and prevention in wireless ad-hoc networks", J. Syst. Archit., vol. 105, May 2020.

[4] S. Thudumu, P. Branch, J. Jin, et al. “A comprehensive survey of anomaly detection techniques for high dimensional big data”. J Big Data 7, 42 (2020). https://doi.org/10.1186/s40537-020-00320-x.

[5] S. Otoum, et al. “A Comparative Study of AI-Based Intrusion Detection Techniques in Critical Infrastructures”, ACM Transactions on Internet Technology (TOIT) 21 (2020): 1 - 22.

на симулация с помощта на масиви от данни за атака. Представени са показатели за ефективност за три различни IDS за откриване на злонамерено поведение.

Наблюдават се различни текущи изследователски проекти, използващи дълбоко обучение и машинно обучение за създаване на ефективни IDS. Популярни контролирани и неконтролирани алгоритми за машинно обучение се прилагат за идентифициране на ефективни и ефективни IDS в мрежи и компютри[6]. Контролираните алгоритми включват изкуствена невронна мрежа, дърво на решенията, k-най-близък съсед, наивен Бейс, произволна гора, поддържаща векторна машина и конволюционна невронна мрежа, докато неконтролираните алгоритми включват максимизиране на очакванията, k-средни стойности и алгоритми за самоорганизиращи се карти.

Авторите представят хибриден метод за базирана на аномалия мрежова IDS, използваща изкуствена пчелна колония (ABC) и AdaBoost алгоритми за постигане на висока степен на откриване с нисък процент на фалшиви положителни резултати[7]. Алгоритъмът ABC се използва за избор на характеристики, а AdaBoost се използва за оценка и класификация на функции.

На базата на множество различни техники за машинно обучение са проектирани множество системи за откриване на аномалии[8]. Изследват се проблемите на конвенционалните масиви от данни за проникване в мрежа (UNSW-NB15, TUIDS и NSLKDD)[9]. От друга страна, някои системи, като хибридни или ансамблови техники, се основават на смесване на различни техники за обучение. Тези методи са специално проектирани като класификатори, които се използват за категоризиране или идентифициране дали входящата интернет връзка е нормална или има атака.

Представен е подходът за машинно обучение за IDS като стратегия за киберсигурност за малки и средни предприятия[10]. Алгоритмите, които се тестват чрез реален набор от данни, са Naive Bayes, Sequential minimal optimization (SMO), C4.5 дърво на решенията и Random Forest. Проектирана е система за откриване на аномалии в реално време, базирана на YOLOv4, за автоматизиране на откриването на дефектни продукти в реални производствени обекти[11]. Системата допринася за събирането на данни за

[6] Z. K. Maseer, R. Yusof, N. Bahaman, S. A. Mostafa and C. F. M. Foozy, "Benchmarking of machine learning for anomaly based intrusion detection systems in the CICIDS2017 dataset", IEEE access, vol. 9, pp. 22351-22370, 2021.

[7] M. Mazini, B. Shirazi and I. Mahdavi, "Anomaly network-based intrusion detection system using a reliable hybrid artificial bee colony and AdaBoost algorithms", J. King Saud Univ. Comput. Inf. Sci., vol. 31, no. 4, pp. 541-553, Oct. 2019

[8] M. Verkerken, D'hooge, L., Wauters, T. et al. "Towards Model Generalization for Intrusion Detection: Unsupervised Machine Learning Techniques", J Netw Syst Manage 30, 12 (2022). https://doi.org/10.1007/s10922-021-09615-7.

[9] R. Chapaneri, S. Shah, "A Comprehensive Survey of Machine Learning-Based Network Intrusion Detection", In: Satapathy, S., Bhateja, V., Das, S. (eds) Smart Intelligent Computing and Applications . Smart Innovation, Systems and Technologies, vol 104. 2019, Springer, Singapore. https://doi.org/10.1007/978-981-13-1921-1_35.

[10] N. Baci, K. Vukatana, M. Baci, "Machine Learning Approach for Intrusion Detection Systems as a Cyber Security Strategy for Small and Medium Enterprises", WSEAS Transactions on Business and Economics, vol. 19, pp. 474-480, 2022.

[11] D. Kim, Y.-H. Han, J. Jeong, "Design and Implementation of Real-time Anomaly Detection System based on YOLOv4", WSEAS Transactions on Electronics, vol. 13, pp. 130-136, 2022.

производството и изграждането на система за интелигентна фабрика чрез използване на установената система.

Два основни класа методи за откриване на аномалии се прилагат за проектиране на системи за откриване на проникване. Методът за откриване на аномалия, базиран на правила, използва специфични правила, за да уточни какво е типично и повдига флаг, когато едно или повече от правилата са нарушени. Този метод е лесен за разсъждение (какво причинява аномалията), прост и разбираем, може да бъде динамичен/адаптивен, но обикновено не е адаптивен към промените в трафика.

Методът за откриване на аномалии, базиран на машинно обучение, прави модел на поведение, на който може да се вярва, и след това оценява новото поведение спрямо него. Използва се, за да научи класификаторите да разпознават нормалното поведение. Методът е адаптивен към промените и изисква минимална човешка намеса, но има семантични пропуски.

Правилата често се правят от хора и не могат да бъдат модифицирани, за да отчетат промените в трафика. Техниките за машинно обучение изискват обучение на модела и оценка на входните данни. Предизвикателствата пред използването на ML за откриване на аномалии са свързани с:

- Системите за сигурност са много нетърпими към грешки;
- Семантични пропуски;
- Липса на данни за обучение;
- Трудности при оценяването;
- Много изследвания, но не са налични много успешни системи.

2. Анализ на системи за откриване на аномалии, базирани на машинно обучение

А. Методи на системи за откриване на проникване

1) Откриване на базата на сигнатура

Мрежовият трафик се наблюдава чрез техники, базирани на сигнатура, за да се определи дали конкретен модел съответства на сигнатурата на атаката, включена в пакетите (фиг. 3). Наблюдаваните сигнатури се съпоставят със сигнатури за атака - колекция от разпознати сигнатури за атака - или атрибути на известни вредни заплахи в базата данни. Извършва се бързо търсене, проверка и сравнение на съдържанието на уловените мрежови пакети за сигнатури на известни заплахи.

Фиг. 3. Откриване на базата на сигнатура

2) Откриване на базата на аномалия

Основата на базираното на аномалии откриване е категоризирането на данните чрез сравняване на поведението на приложението с тези, които се считат за типични. Техниките за машинно обучение осигуряват основата за идентифициране на аномалии в мрежовия трафик в реално време. Вместо само да съпоставя предварително съществуващите сигнатури, моделът на машинно обучение се адаптира, за да разпознава сложни модели на трафик и анализира поведението, свързано с определени атаки, като прави информирани преценки. Откриването на базата на аномалии може да се използва за идентифициране на атаки, които се отклоняват от очакваното поведение. Архитектурата включва предварителна обработка на информация, след това разработване на модел (фиг. 4).

Фиг. 4. Откриване на базата на аномалия

Откриването на аномалии при извличането на данни се разглежда и като проблем за откриване на отклонения. Аномалиите/извънредните стойности могат да бъдат категоризирани в няколко категории[12]: (1) точкови аномалии се появяват, когато отделна инстанция на данни може да се разглежда като необичайна в сравнение с други данни; (2) контекстуални аномалии се появяват, когато отделна инстанция на данни е необичайна в определен контекст, но не и по друг начин; (3) колективни аномалии се появяват, когато свързани инстанции на данни са необичайни по отношение на другите разглеждани данни.

Системите за откриване на аномалии обикновено работят, като следват няколко стъпки:

- Събиране на данни: системата събира данни от различни източници;
- Предварителна обработка на данните: системата обработва данните, за да ги изчисти и да отстрани всякакви шумове;
- Обучение на модела: системата обучава модела, за да научи как изглежда нормалното поведение на данните;
- Откриване на аномалии: системата използва обучения модел, за да открива аномалии в данните, когато те се появят.

Методите и алгоритмите, които могат да се използват за идентифициране на аномалии, се класифицират като контролирани, неконтролирани и полуконтролирани в зависимост от наличието на етикети на данните и анотации на екземплярите от данните[13]. Контролираното обучение включва обучение на модел върху маркирани данни, при което аномалиите се идентифицират и маркират. След това моделът може да се използва за прогнозиране на аномалии в нови данни въз основа на моделите, които е научил от

[12] V. Chandola, A. Banerjee, V. Kumar, Anomaly detection: a survey, ACM Computing Surveys, Vol. 41, No. 3, 2009.

[13] A. Karale, Outlier detection methods and the challenges for their implementation with streaming data, Journal of Mobile Multimedia, Vol. 16, No. 3, pp. 351-388, 2020

маркираните данни. Неконтролираното обучение включва обучение на модел върху немаркирани данни и идентифициране на модели и аномалии въз основа на структурата на данните. Този подход може да бъде особено полезен, когато аномалиите са редки или когато данните са силно променливи. Полунаблюдаваното обучение включва обучение на модел върху комбинация от маркирани и немаркирани данни. Този подход може да бъде полезен, когато има ограничени налични маркирани данни, но моделът все пак може да научи модели и аномалии от немаркираните данни.

С оглед на проблема с откриването на аномалии в софтуерните системи подходите за автоматизирано откриване на аномалии се основават главно на данни от мониторинг, които включват метрики, логове, сигнали и следи[14], като данните от няколко източника са най-значими за постигане на съответната точност и задоволителни резултати[15]. Метриките са времеви редове с реална стойност, измерващи състоянието на системата, напр. използване на процесора и паметта, латентност и пропускателна способност, време за отговор, брой нишки и др. Журнали са полуструктурирани текстови съобщения, извеждани от протоколи за регистриране, за да се записва състоянието на системата по време на изпълнение[16].

Подходите за откриване на аномалии, базирани на метрики, са в състояние да откриват инциденти, свързани със здравето на приложенията. Откриването на аномалии, базирано на метрики, се основава на обучение на модел с помощта на набор от данни за обучение, който се състои от стойности на метрики, измервани на редовни интервали. По този начин моделът ще може да открива аномалии в нормалния модел на метриките и съответно да подава сигнали, когато поведението се променя значително от очакваното. Основните етапи на системите за откриване на аномалии, базирани на метрики, са представени на фиг. 5.

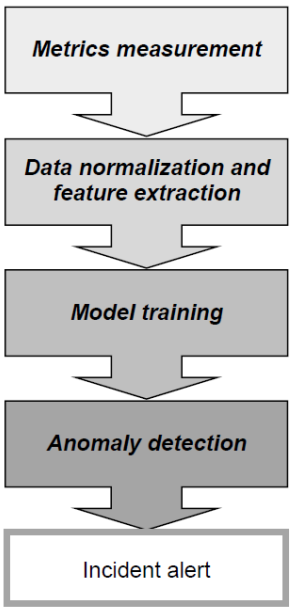
В сравнение с общите етапи на системата за откриване на аномалии, в системите, базирани на метрики, етапът на събиране на данни включва измерване на метрики на редовни интервали, за да се събере набор от данни за обучение под формата на времеви ред. Етапът на предварителна обработка на данните изисква подходящо нормализиране на данните, което е особено важно, когато се използват множество метрики. На етапа на предварителна обработка на данните може да се приложи и подходящо извличане на признаци, така че измерените данни да бъдат подходящо трансформирани. Етапът на нормализиране на данните и избор на признаци се различава в зависимост от алгоритъма за обучение на модела, използван на следващия етап от системата за откриване на аномалии. Основното предизвикателство на базираните на модели подходи за откриване на аномалии е точността на откриване на отклонения в зашумени данни, както и невъзможността да се открият неизвестни аномалии. Освен това тези подходи трябва да

[14] Z. Chen, et. al., Towards intelligent incident management: why we need it and how we make it, Proc. of the 28th ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering, ACM, pp. 1487–1497, 2020.

[15] J. Bogatinovski, S. Nedelkoski, Multi-source anomaly detection in distributed it systems, in H, Hacid, et al. (eds.), Service-Oriented Computing Workshops, Lecture Notes in Computer Science, Vol. 12632, Springer, Cham, 2020.

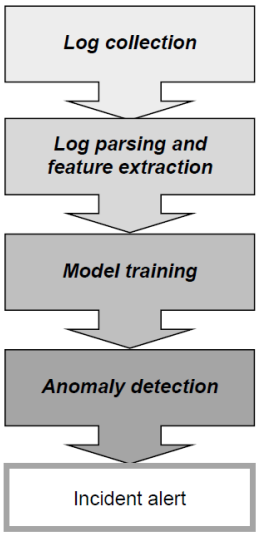
[16] C. Lee, T. Yang, Z. Chen, Y. Su, Y. Yang, M. Lyu, Heterogeneous anomaly detection for software systems via semi-supervised cross-modal attention, Proc. of the 45th International Conference on Software Engineering, IEEE Press, pp. 1724–1736, 2023.

използват многоизмерния и мултимодален характер на инцидентното влияние на няколко времеви редици от показатели.



Фиг. 5 Етапи на откриване на аномалии въз основа на метрики

Събитията от дневника се генерират синхронно с изпълнението на определени софтуерни пътища и представляват ценен източник на данни за откриване на инциденти. Събитията в дневника обикновено се съхраняват като текстово базирани лог-файлове, като по този начин общите етапи на системите за откриване на аномалии също се променят съответно на изискването за обработка на входни данни, базирани на естествен език (фиг. 6). На етапа на предварителна обработка на данните се прилага парсинг и токенизация на логовете, като суровото и неструктурирано лог съобщение се преобразува в текст и пунктуационни данни в основано на признаци представяне.



Фиг. 6 Етапи на откриването на аномалии на базата на регистри

Допълнителен подбор на признаци и извличане на признаци може да се приложи според изискванията на съответния алгоритъм за обучение на модела. На следващия етап се обучава модел за откриване на аномалии, така че да идентифицира данни, които не отговарят на очакваните модели.

Тъй като наборите от данни за метрики и логове, използвани при откриването на софтуерни аномалии, обикновено не са маркирани, основните използвани подходи се основават на обучение без наблюдение. За да се използва контролирано обучение за откриване на аномалии, на допълнителен етап след събирането на данни е необходимо анотиране на случаите в набора от данни като нормални или аномални.

Б. Техники за машинно обучение

IDS, базиран на аномалии, може да изгради система, която може да се учи от данни и да предлага преценка за ненаблюдавани данни чрез използване на техники за машинно обучение. Основната функция на техниките за машинно обучение е да правят разлика между нормално и злонамерено поведение.

Support Vector Machine

Най-често срещаната и харесвана техника за машинно обучение за приложения в класификацията и регресията е SVM[17]. Една от двете категории се избира за всеки пример в поредица от примери за обучение. След това техниката SVM се използва за създаване на модел, който може да предвиди дали нов пример попада в една или друга категория. Обучителните проби трябва да бъдат избрани преди извличане на атрибути от данни за SVM класификация. Обикновено мрежова връзка се избира като представителен или референтен набор от данни, който включва набор от характеристики на мрежовата връзка, получени от мрежи с отклонения.

SVM е специфичен за обучение от малки извадки и често се използва в редица случаи, включително разпознаване на лица, идентификация на интернет страница и откриване на проникване в мрежа. Високи нива на вземане на решения и обучение, нечувствителност към измерението на входните данни, непрекъсната корекция на различни параметри с нарастващи данни за обучение, което предоставя на системата потенциала да се учи сама, и други предимства са някои от предимствата на използването на SVM при откриване на проникване. Той може също така да отговори на широк спектър от проблеми с класификацията, които възникват в ситуации от реалния свят. В резултат на това SVM печели популярност в мрежовата сигурност. Предимствата включват повишена способност за учене от малки извадки, високи нива на обучение и преценка и нечувствителност към измерението на входните данни. Обучението обаче отнема време и по-голямата част от бинарните класификатори, които се използват днес, не са в състояние да дадат допълнителни подробности за типа атака, която е открита.

[17] I.H. Sarker, “Machine Learning: Algorithms, Real-World Applications and Research Directions”, SN COMPUT. SCI. 2, 160 (2021). https://doi.org/10.1007/s42979-021-00592-x.

Невронни мрежи

Невронната мрежа може да екстраполира заключения от оскъдни, противоречиви и непълни данни. Това прави възможно намирането на модели, които или са били неизвестни преди това, или не съвпадат напълно и се различават от дефинираните структури на по-ранните входни модели. Невронната мрежа е обещаваща технология за откриване на аномалии, тъй като идеалният детектор за проникване трябва да може да разграничава настоящите и бъдещи атаки в допълнение към записаните атаки. Не изисква специализирани познания, за да може да прави заключения от оскъдни, шумни и непълни данни и може да идентифицира неочаквани или необичайни прониквания. Възможността за по-късно намиране на неидентифицирани досега модели също е добра. Неговата скромна скорост на обучение може да се счита за недостатък, защото го прави предизвикателство за откриване в реално време. Пренастройването може да се извърши докато невронната мрежа се обучава.

Дърво на решенията

Проблемите с класификацията обикновено се решават с помощта на метода на дървото на решенията. Наборът от данни се научава и моделира с помощта на този алгоритъм. В резултат на това, когато се подадат нови данни за класифициране, получените знания от предишни набори от данни се използват за класифициране на новия набор от данни. Откриването на проникване е друг случай на използване подхода на дървото на решенията. Алгоритъмът на дървото на решенията създава модели на данни и се учи от данните за обучение. Капацитетът му да управлява големи колекции от данни е едно от неговите предимства. Това е важно, защото има постоянен поток от данни през компютърни мрежи. Дърветата на решенията се представят добре при откриване на проникване в реално време, защото предлагат най-висока производителност на откриване и са лесни за конструиране и оценка. Друга изгодна характеристика е точността на обобщаване на дървото на решенията в модела за откриване на проникване. Това се дължи на факта, че обобщената точност на дървото на решенията прави възможно разпознаването на нови атаки, които е вероятно постоянно да се появяват. Дърветата на решенията предлагат голяма прецизност на откриване и са полезни при големи набори от данни. Създаването на дърво на решенията изисква много изчисления.

Генетични алгоритми

Кръстосването, наследяването, мутацията и селекцията са само някои от еволюционните алгоритмични техники, използвани от генетичните алгоритми. Те използват евристични търсения и са повлияни от биологията. В резултат на това генетичните алгоритми могат да генерират правила за класификация и да изберат оптималните параметри на техниката за откриване. Като цяло прилагането на генетични алгоритми към мрежови данни включва следните фази. Системата за откриване на проникване следи трафика, преминаващ през определена мрежа. След това системата за откриване на проникване използва генетични алгоритми, които са обучени с помощта на критериите за категоризиране, идентифицирани от информацията, получена по време на мрежовото разследване на системата за откриване на проникване. Системата за откриване на проникване използва набор от критерии, за да класифицира входящите данни като необичайни или редовни въз основа на техния модел. GA успешно генерира

необходимите IDS атрибути за висок истински процент на откриване и нисък процент на фалшиви положителни резултати. Клъстерирането на GA е потенциална техника за откриване на вредни проникванията в компютърни системи и е използвано успешно в IDS за разграничаване между нормални действия и действия с проникване. Постоянните времена за реакция на оптимизация не могат да бъдат гарантирани от генетични алгоритми.

Размита логика

Размитата логика произлиза от теорията на размитите множества, която по-скоро приближава, отколкото точно прилага традиционната предикатна логика. Следователно размитите техники се използват в областта на откриването на аномалии, тъй като характеристиките, които трябва да се вземат предвид, могат да се разглеждат като размити променливи. Използвайки размити пространства, размитата логика позволява на даден обект едновременно да принадлежи към множество класове. Тази концепция е полезна, когато има несигурност по отношение на класификацията на класовете. Разликите между нормалните и необичайните класове не са ясно дефинирани и това важи за задачата за откриване на проникване.

Всяко развито размито правило използва не повече от пет качества, така че не са особено сложни. Предоставено е просто определение както на нормално, така и на аномално поведение. По-простите размити правила безспорно са по-полезни в приложения от реалния свят. Те първо създават правила, които са по-лесни за разбиране и имат висока степен на оперативна съвместимост. Освен това те създават правило за класификатор, което може да се използва по-бързо. Това е особено важно за данни с множество атрибути. Размитата логика се оказа полезна, особено когато се използва за блокиране на сканиране на портове и проби, но основните й недостатъци са високата консумация на ресурси и удължената крива на обучение. Обосновката е по-скоро приблизителна, отколкото точна. Предизвикателство е да се идентифицира малко, важно подмножество от правила. Динамичното актуализиране на правила по време на изпълнение е трудно.

Байесова мрежа

Вероятностните връзки между важните променливи са кодирани от байесовия мрежов модел. Обикновено тази техника се използва във връзка със статистически системи за откриване на проникване. Предимствата на този подход включват неговия капацитет за кодиране на взаимозависимости между променливи, прогнозни резултати и включва както данни, така и предишни знания. Наивната байесова техника се използва за задачи за обучение, при които е наличен набор за обучение с целеви клас. Предимствата на метода са, че той записва вероятностни връзки между съответните променливи и има способността да интегрира данни и минали знания. От друга страна, трудно е да се приложи, когато се работи с непрекъснати характеристики и има вероятност да пропусне полезни класификатори, ако предишните допускания са неверни.