Алтынников М.С. Обзор методов прогнозирования кибератак в образовательных учреждениях // Международный журнал информационных технологий и энергоэффективности. – 2022. – Т. 7 № 4(26) часть 1 с. 49–53

Международный журнал информационных технологий и энергоэффективности

Сайт журнала:
http://www.openaccessscience.ru/index.php/ijcse/

УДК 004.89

ОБЗОР МЕТОДОВ ПРОГНОЗИРОВАНИЯ КИБЕРАТАК В ОБРАЗОВАТЕЛЬНЫХ УЧРЕЖДЕНИЯХ

Алтынников М.С.
Иркутский Государственный Университет Путей Сообщения, Иркутск, Россия (664074, г. Иркутск, ул. Чернышевского, д.15), e-mail: ms@altynnikov.ru

Анализ научных исследований разных стран в сфере прогнозирования кибератак показывает, что прогнозирование кибератак до их возникновения - важная, но сложная задача, поскольку поиск ранних признаков атаки из большого объема данных не является тривиальной задачей. На протяжении нескольких лет научным сообществом ведутся исследовательские работы по прогнозированию кибератак различными методами с целью создания адекватных методов заблаговременной защиты от них. В работе проеден литературный обзор на тему методов прогнозирования кибератак образовательных учреждениях. Были рассмотрены следующие методы: машинное обучение; применение предварительно известных сигнатур или прототипов определенных процессов или событий; метод интервального прогнозирования интенсивности кибератак по средствам ВНС.

Ключевые слова: прогнозирование кибератак, кибератаки, методы прогнозирования.

REVIEW OF METHODS FOR CYBER ATTACK PREDICTION IN EDUCATIONAL INSTITUTIONS

Altynnikov M.S.
Irkutsk State Transport University, Irkutsk, Russia (664074, Irkutsk, st. Chernyshevsky, 15), e-mail: ms@altynnikov.ru

Analysis of research from various countries on cyber attack prediction shows that predicting cyber attacks before they occur is an important but challenging task, as finding early warning signs of an attack from large amounts of data is not a trivial task. For several years, the scientific community has been conducting research work on predicting cyber attacks using various methods in order to create adequate methods for early protection against them. The paper reviewed the literature on methods of predicting cyber attacks in educational institutions. The following methods were considered: machine learning; application of preknown signatures or prototypes of certain processes or events; method of interval prediction of cyber attack intensity by means of PNN

Keywords: personal data, information security, security in medical institutions.

В мире более 25% утечек информации происходит из медицинских учреждений, в России доля таких утечек составляет 7%. Обращает на себя внимание высокая (в сравнении с общемировой) доля утечек, которые пришлись на банки и финансовые организации (12%). Также высоки (в сравнению с мировыми показателями) доли образовательных учреждений (14%), госорганов и силовых структур (22%) [4].

В I квартале 2022 г. количество утечек конфиденциальной информации из образовательных учреждений во всем мире выросло более чем на 15% по сравнению с

аналогичным периодом прошлого года. Хакеры и внутренние злоумышленники похищали персональные данные и другую конфиденциальную информацию [5].

Сфера образования всегда была одной из самых атакуемых отраслей. Однако, видна тенденция повышения частоты кибератак в исследуемой сфере деятельности. Часто к моменту обнаружения взлома бывает уже слишком поздно, и ущерб уже нанесен. В результате такие события заставляют задуматься о том, можно ли было предугадать эти нарушения и избежать ущерба.

Поскольку риски кибератак продолжают расти, необходимы исследования и разработки, позволяющие прогнозировать атаки вместо пассивного обнаружения вторжения. В последние годы исследователи начали использовать предиктивную аналитику, которая помогает прогнозировать будущие киберинциденты против целевых организаций до того, как они произойдут.

Машинное обучение широко используется в области кибербезопасности, в основном для обнаружения различных вредоносных действий или субъектов, например, спама и фишинга. Для целей прогнозирования они используются гораздо реже, за исключением работы [7], где текстовые данные используются для обучения классификаторов, позволяющих предсказать, может ли доброкачественная в настоящее время веб-страница стать вредоносной в ближайшем будущем. Разница между обнаружением и предсказанием аналогична разнице между диагностикой пациента, который уже может быть болен (например, с помощью биопсии), и прогнозированием того, может ли в настоящее время здоровый человек заболеть, на основе множества соответствующих факторов. Первый вариант обычно основывается на определении известных характеристик объекта, который необходимо обнаружить, а второй - на факторах, которые, как считается, коррелируют с целью прогнозирования.

Предсказание интенсивности кибератак в концепции раннего определения и предотвращения несовершенство большинства современных систем обеспеченья кибербезопасности объектов заключается в том, что при идентификации кибератак применяются предварительно известные сигнатуры или прототипы определенных процессов или событий [6]. Например, так осуществляется работа антивирусных систем, межсетевых экранов, систем обнаружения и предотвращения вторжений. В работе [6] утверждается, что подобные системы результативны только в отношении начинающих злоумышленников, которые применяют стандартные приёмы и инструменты для организации кибератак. Против опытных злоумышленников настоящие системы, часто, оказываются малоэффективными. Тут одним из перспективных направлений исследований для решения поставленной проблемы является направление по прогнозированию интенсивности кибератак на объектах средством машинного обучения. Например, такой подход удачно интегрируется в изложенную в [2,3] концепцию раннего распознавания кибератак и предупреждения о них.

Так же в [1] подробно представлен метод интервального прогнозирования интенсивности кибератак на объекты критической информационной инфраструктуры. Для организации интервального прогнозирования была избрана вероятностная нейронная сеть [8] с динамическим обновлением параметра сглаживания [9] (ВНС).

Преимущества ВНС (применительно к прогнозированию интенсивности кибератак) превалируют над недостатками [10].

Например, ВНС:

• при обучении и прогнозировании устойчива к аномальным выбросам;

• модель причисляется к моделям «ленивого» обучения и обучается предельно быстро в сравнении с моделями прочих классов;

• модель устойчива к «дисбалансу» классов обучающего множества;

• результаты работы ВНС легко поддаются интерпретации, так как работа ВНС основана на выявлении «схожих» объектов;

• не просит априорных познаний о статистических характеристиках прогнозируемого показателя.

К недостаткам можно отнести:

• «неотделимость» процесса прогнозирования от обучающих данных (в отличие, например, от параметрических моделей, где «обучение» заключается в оценке параметров моделей);

• обучающая выборка должна быть репрезентативной.

В [11] изучают отчеты, собранные с помощью антивирусных агентов McAfee, установленных на 85 000+ узлах многонационального предприятия. Используя логистическую регрессию для прогнозирования риска столкновения узлов с вредоносным ПО, они обнаружили, что узлы с высоким рейтингом сталкиваются с вредоносным ПО в 3 раза чаще по сравнению с базовым показателем. Liu, Y, Sarabi A, Zhang J, Naghizadeh P, Karir M, Bailey M, Liu M [12] собирают 258 внешних измеряемых признаков из сети организации, которые основаны на неправильно настроенных DNS (или BGP) в сети и временных рядах вредоносной активности для спама, фишинга и сканирования. Обучив классификатор Random forest, используя собранные признаки и сообщения о киберинцидентах в базе данных сообщества VERIS, Hackmageddon и Web Hacking Incidents Database, они достигли 90% точности в прогнозировании нарушений против целевой организации. Также в [13] используют журналы появления двоичных файлов и маркированные данные из антивирусных продуктов и продуктов предотвращения вторжений антивирусной компании для прогнозирования того, какие машины подвержены высокому риску заражения. Используя классификатор Random Forest и подход полусамостоятельного обучения, они достигают высокой точности (коэффициент истинных и ложных срабатываний составляет 96% и 5% соответственно) в прогнозировании риска заражения для хостов.

Заключение

Анализ научных исследований разных стран в сфере прогнозирования кибератак показывает, что прогнозирование кибератак до их возникновения - важная, но сложная задача, поскольку поиск ранних признаков атаки из большого объема данных не является тривиальной задачей. На протяжении нескольких лет научным сообществом ведутся исследовательские работы по прогнозированию кибератак различными методами с целью создания адекватных методов заблаговременной защиты от них..

Список литературы

1. Ю.М. Краковский, Б.В. Курчинский, А.Н. Лузгин. «Интервальное прогнозирование интенсивности кибератак», Доклады ТУСУР ¬– 2018 – том 21 – № 1, 2005..
2. Петренко С.А. Концепция раннего распознавания и предупреждения компьютерного нападения / С.А. Петренко, А.С. Петренко // Матер. Всерос. науч.-практ. конф. «Информационные системы и технологии в моделировании и управлении». – 2016. – С. 82–86.

3. Петренко С.А. Национальная система раннего предупреждения о компьютерном нападении / С.А. Петренко, Д.Д. Ступин. – Инополис: Изд. дом «Афина», 2017. – 440 с.
4. Аналитический центр InfoWatch. «Утечки данных. Россия. 2016 год», 2017, С. 16.
5. Аналитический центр InfoWatch. «Утечки конфиденциальной информации из сферы образования». – 2022
6. Jones M. Cyber-Attack Forecast Modeling and Complexity Reduction Using a Game-Theoretic Framework / M. Jones, G. Kotsalis, J.S. Shamma // Tarraf D. (eds) Control of Cyber-Physical Systems. Lecture Notes in Control and Information Sciences. – Heidelberg: Springer, 2013. – 380 p.
7. SO SKA, K. , CHRISTIN, N . Automatically Detecting Vulnerable Websites Before They Turn Malicious.In Proceedings of the 23rd USENIX Security Symposium(San Diego, CA, August 2014).
8. Specht D.H. Probabilistic Neural Networks / D.H. Specht // Neural Networks. – 1990. – № 3. – P. 109–118.
9. Kargapoltsev S.K. A dynamic updating algorithm of smoothing parameter values of probabilistic neural networks / S.K. Kargapoltsev, Y.M. Krakovsky, A.V. Lukyanov, A.N. Luzgin // Far East Journal of Electronics and Communications. – 2017. – Vol. 17, № 4. – P. 909–914.
10. Probabilistic neural network [Электронный ресурс]. – Режим доступа: https://en.wikipedia.org/wiki/Probabilistic_ neural_network, свободный (дата обращения: 02.04.2018).
11. Yen, TF, Heorhiadi V, Oprea A, Reiter MK, Juels A (2014) An epidemiological study of malware encounters in a large enterprise In: Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security, CCS ’14, 1117–1130.
12. Liu, Y, Sarabi A, Zhang J, Naghizadeh P, Karir M, Bailey M, Liu M (2015) Cloudy with a chance of breach: Forecasting cyber security incidents In: Proceedings of the 24th USENIX Security Symposium (USENIX Security 15), 1009–1024.
13. Bilge, L, Han Y, Dell’Amico M (2017) Riskteller: Predicting the risk of cyber incidents In: Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security (CCS), 1299–1311.

References

1. Yu.M. Krakovsky, B.V. Kurchinsky, A.N. Luzgin. “Interval Prediction of Cyberattack Intensity”, TUSUR Reports ¬– 2018 – Volume 21 – No. 1, 2005..
2. Petrenko S.A. The concept of early recognition and prevention of a computer attack / S.A. Petrenko, A.S. Petrenko // Mater. Vseros. scientific-practical. conf. "Information systems and technologies in modeling and management". - 2016. - S. 82–86.
3. Petrenko S.A. National system of early warning about a computer attack / S.A. Petrenko, D.D. Stupin. - Innopolis: Ed. house "Athena", 2017. - 440 p.
4. Analytical center InfoWatch. “Data leaks. Russia. 2016”, 2017, p. 16.
5. Analytical center InfoWatch. Leaks of confidential information from the education sector. – 2022
6. Jones M. Cyber-Attack Forecast Modeling and Complexity Reduction Using a Game-Theoretic Framework / M. Jones, G. Kotsalis, J.S. Shamma // Tarraf D. (eds) Control of Cyber-Physical

Systems. Lecture Notes in Control and Information Sciences. – Heidelberg: Springer, 2013. – 380 p.

7. SO SKA, K. , CHRISTIN, N . Automatically Detecting Vulnerable Websites Before They Turn Malicious.In Proceedings of the 23rd USENIX Security Symposium(San Diego, CA, August 2014).
8. Specht D.H. Probabilistic Neural Networks / D.H. Specht // Neural Networks. - 1990. - No. 3. - P. 109–118.
9. Kargapoltsev S.K. A dynamic updating algorithm of smoothing parameter values of probabilistic neural networks / S.K. Kargapoltsev, Y.M. Krakovsky, A.V. Lukyanov, A.N. Luzgin // Far East Journal of Electronics and Communications. - 2017. - Vol. 17, No. 4. - P. 909–914.
10. Probabilistic neural network [Electronic resource]. – Access mode: https://en.wikipedia.org/wiki/Probabilistic_neural_network, free (date of access: 04/02/2018).
11. Yen, TF, Heorhiadi V, Oprea A, Reiter MK, Juels A (2014) An epidemiological study of malware encounters in a large enterprise In: Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security, CCS '14, 1117 –1130.
12. Liu, Y, Sarabi A, Zhang J, Naghizadeh P, Karir M, Bailey M, Liu M (2015) Cloudy with a chance of breach: Forecasting cyber security incidents In: Proceedings of the 24th USENIX Security Symposium (USENIX Security 15) , 1009–1024.
13. Bilge, L, Han Y, Dell'Amico M (2017) Riskteller: Predicting the risk of cyber incidents In: Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security (CCS), 1299–1311.