DOI 10.37882/2223-2966.2024.7.20

УСОВЕРШЕНСТВОВАННЫЙ МЕТОД АВТОМАТИЧЕСКОГО АКТИВНОГО АНАЛИЗА ЗАЩИЩЕННОСТИ КОРПОРАТИВНОЙ СЕТИ

Колесников Антон Александрович
Санкт-Петербургский Политехнический университет Петра Великого, anton.kolesnikov.science@mail.ru

Аннотация. В контексте быстрого развития информационных технологий, приложений, облачных вычислений, больших данных и Интернета вещей вопросы обеспечения защиты корпоративных сетей становятся все более сложными. Это в свою очередь предполагает необходимость проведения углубленного анализа обнаружения вторжений и оценки безопасности сетей. С учетом отмеченного, в статье предложен усовершенствованный метод автоматического активного анализа защищенности корпоративной сети, основу которого составляет модель ситуационной осведомленности в области кибербезопасности и оптимизированная модель гауссовой матрицы — рекуррентные нейронные сети.

Ключевые слова: корпоративная сеть, защита, анализ, модель, риск.

AN ADVANCED METHOD OF AUTOMATIC ACTIVE ANALYSIS OF CORPORATE NETWORK SECURITY

A. Kolesnikov

Summary. In the context of rapid development of information technologies, applications, cloud computing and big data, and the Internet of Things, the issues of securing corporate networks are becoming increasingly complex. This in turn implies the need for in-depth analysis of intrusion detection and network security assessment. In view of the above, the paper proposes an improved method of automatic active analysis of corporate network security, the basis of which is a model of situational awareness in the field of cybersecurity and an optimized model of Gaussian matrix — recurrent neural networks.

Keywords: corporate network, protection, analysis, model, risk.

Безопасность корпоративной сети имеет первостепенное значение для современных предприятий. Она служит основой современных бизнес-операций, облегчая поток информации и обеспечивая работу критически важных приложений и услуг. Однако корпоративные сети не застрахованы от таких проблем, как задержки, нарушения безопасности и снижение производительности. В условиях постоянно меняющегося ландшафта угроз и увеличения числа кибератак крайне важно внедрять надежные меры сетевой безопасности для защиты конфиденциальных данных, информации клиентов и обеспечения бесперебойной работы всех систем.

В тоже время, существует распространенное заблуждение, что сетевая безопасность — это то, о чем должны беспокоиться только глобальные корпорации. При этом, статистика кибератак на корпоративные сети свидетельствует о том, что малый и средний бизнес на самом деле являются наиболее распространенной мишенью для киберпреступников: на их долю приходится 43 процента всех утечек данных. Одной из причин того, почему хакерам удается взламывать корпоративные сети, является ложное чувство безопасности, присущее многим предприятиям. Отдельный отчет BullGuard показывает, что 32 процента предприятий используют только бесплатные решения для безопасности сетевых устройств, а 23 процента вообще не используют никаких средств безопасности [1].

На сегодняшний день разработаны различные системы анализа защищенности корпоративной сети. В рамках большинства из них используется пользовательское программное обеспечение (ПО) или пользовательские варианты готового ПО, при этом подобный анализ представляет собой полуавтоматический процесс, поскольку полученные данные приходится вручную сравнивать с отчетами других организаций. Такой медленный темп изучения сетевого трафика на предмет угроз безопасности ставит организации в затруднительное положение, когда речь заходит о том, чтобы противостоять современным векторам атак: новому поколению ботнетов и DDoS-инструментов, растущему присутствию ненадежных персональных устройств в корпоративных сетях и ограниченной защите, обеспечиваемой технологиями Интернета вещей.

Таким образом, учитывая вышеупомянутые недостатки, набирает весомости вопрос автоматизации процесса активного анализа защищенности корпоративных сетей, что подтверждает актуальность, теоретическую и практическую значимость темы данной статьи.

Практические аспекты, касающиеся усиления активной защиты безопасности корпоративных сетей, улучшения классификации и прогнозирования событий в сетях раскрывают в своих публикациях Панков А.В., Елфимов А.В., Ушаков И.А., Горбатов В.С., Malaram Kumhar, Jitendra Bhatia.

Над разработкой методов оценки безопасности корпоративной сети с помощью иерархического анализа трудятся такие ученые как: Сердечный А.Л., Айдаркин А.В., Тарелкин М.А., Lennart Jaeger, Andreas Eckhardt, Ali Kamil Abed.

Возможности модели скрытой цепи Маркова и модели графа атак для прогнозирования намерений атаки в корпоративных сетях рассматриваются Азовцевой А.А., Мазниным Д.Н., Дворниковой О.А., Попковым Г.В., Pimal Khanpara, Kruti Lavingia, Rajvi Trivedi, Sudeep Tanwar, Amit Verma.

Положительно оценивая имеющиеся на сегодняшний день труды и наработки, отметим, что ранние и более зрелые технологии безопасности в основном сосредоточены на защите сетевой информации, но соответствующие технологии все еще нуждаются в оптимизации. Так, отдельного внимания заслуживают вопросы, связанные с обоснованием стратегии выбора вычислительного масштаба в процессе разработки модели мониторинга безопасности системы. Кроме того, более детальной проработки требует перспектива использования методов интеллектуального анализа для обучения моделей на больших размерностях данных сетевого трафика, чтобы различать поведение сети.

Итак, цель статьи заключается в рассмотрении возможностей усовершенствования метода автоматического активного анализа защищенности корпоративной сети.

В рамках проводимого исследования представляется целесообразным отметить, что метод активного мониторинга безопасности корпоративной сети предполагает активную генерацию тестового трафика или зондов с помощью специальных инструментов или программного обеспечения для мониторинга степени защиты [2]. Активный мониторинг обычно основывается на синтетических операциях, таких как пинг устройства или выполнение тестов по сценарию, для оценки производительности сети и выявления потенциальных проблем. Он позволяет в режиме реального времени получить представление о доступности сети, времени отклика и производительности приложений. Кроме того, активный анализ дает возможность измерять производительность сети с помощью различных метрик и ключевых показателей эффективности. Результатом активного анализа является способность идентифицировать, понимать и предвидеть элементы, относящиеся к сетевой среде в заданном пространстве и времени [3]. Модельное понимание ситуации с сетевой безопасностью показано на рисунке 1.

Учитывая изложенное, считаем, что основу усовершенствованного метода автоматического активного анализа защищенности корпоративной сети должна составлять модель ситуационной осведомленности в области кибербезопасности и оптимизированная модель гауссовой матрицы — рекуррентные нейронные сети. Рассмотрим эти составляющие более подробно.

Модель ситуационной осведомленности в области кибербезопасности

Модель ситуационной осведомленности в области кибербезопасности в основном соответствует жизненному циклу данных о безопасности. В этом жизненном цикле данные принимают различные формы, начиная с исходных данных датчиков, включая очистку данных,

Рис. 1. Модель осознания ситуации с корпоративной сетевой безопасностью

объединение данных, восприятие событий, и заканчивая сценарием. Жизненный цикл данных безопасности включает в себя предварительную обработку данных, распределенное хранение данных, агрегирование данных и обработку событий. В свою очередь анализ жизненного цикла предполагает оценку ситуации, моделирование, анализ сценарием, базовую линию, управление и визуализацию ситуации [4]. Для того чтобы извлечь из данных безопасности высокий уровень ценности, модель должна проводить многоуровневый процесс анализа. На рисунке 2 показан информационный поток системы многоуровневого анализа.

Рис. 2. Многоуровневая структура анализа безопасности корпоративной сети

В соответствии с моделью, представленной на рис. 2, поток информации от датчиков безопасности к конфигурационным файлам образует ценную информационную сеть для реализации ситуационной осведомленности в области кибербезопасности. Датчики получают оперативную, топологическую информацию от объектов системной инфраструктуры и информационных активов. Данные датчиков должны быть очищены и нормализованы, когда они попадают в распределенное хранилище данных.

В базах данных хранятся ключевые документы об истории и текущем состоянии сетевой инфраструктуры. Однако достаточно проблематичным является объединение данных, поступающих от различных датчиков в разных форматах. От журналов сетевого трафика до использования статистики и топологических карт — преобразование данных из разрозненных источников в общий формат представления на соответствующем уровне синтаксиса является сложной и дорогостоящей задачей. Более практичный подход, по мнению автора, заключается в переходе к семантической интеграции данных или интеграции данных на уровне сервисов, таких как обмен данными, виртуализация данных и использование данных как услуги. Эти общие описательные элементы подключаются к распределенной базе данных. После обработки данных основной процесс ситуационной осведомленности в области кибербезопасности заключается в оценке и прогнозировании сценариев, формировании понимания и представления текущей ситуации, а также прогнозировании развития ситуации в ближайшем будущем.

Для агрегирования данных представляется целесообразным использовать алгоритм случайного леса. Все классификаторы деревьев решений определяются как $h(X,\theta_k)$, а результат классификации каждой троичной модели решений рассчитывается по следующей формуле [5]:

$$H(x) = \max_{Y} \sum_{i=1}^{k} I(h_i(x) = Y)$$

$h(X,\theta_k)$ — измеряет связь между X и целевой переменной θ_k; $I(h_i(x)$ — количественно определяет избыточность между X и Y.

Преимуществами модели случайного леса являются сходимость и верхняя граница обобщенной ошибки. Для групповых классификаторов используется функция маржи, которая измеряет, превышает ли среднее число правильных классификаторов среднее число неправильных классификаторов. Чем выше значение маржи, тем надежнее предположение о классификации. Функция маржинальности имеет следующий вид:

$$mg(X,Y) = av_k I\left(h(X,\theta_k) = Y\right) - \max_{j \neq Y} av_k I\left(h(X,\theta_k) = j\right)$$

Ошибки в общих выводах должны определяться в соответствии с характеристиками обычных подсчетов голосов, как показано в формуле:

$$PE^* = PX, P(mgX, Y), 0)$$

По мере увеличения количества деревьев решений в случайном лесу все последовательности $\theta_1,\ldots,\theta_k$.PE^* сходятся почти везде, а сходимость ошибки обобщения показана в формуле:

$$\lim_{k\to\infty} PE^* = P_{X,Y}\left(\begin{array}{c} P_\theta\left(h(X,\theta_k) = Y\right) \\ -maxP_\theta\left(h\left(X, \begin{array}{c}\theta_k \\ j \neq Y\end{array}\right) = j\right) \prec 0 \end{array}\right)$$

На пике общей ошибки выполняется неравенство Чебышева:

$$PE^* \leq \frac{var_{X,Y}\left(mg(X,Y)\right)}{E_{X,Y} mg(X,Y)^2}$$

Согласно выше представленному уравнению можно определить классификационную силу одного дерева решений, как показано в формуле:

$$s = E_{X,Y} mg(X,Y)$$

В свою очередь функция общего верхнего предела погрешности будет иметь следующий вид:

$$PE^* \leq \frac{\rho(1-s^2)}{s^2}$$

Таким образом, можно отметить, что многоуровневая модель ситуационной осведомленности в области кибербезопасности, построенная в данном исследовании, согласует процесс ситуационной осведомленности о защищенности корпоративной сети с жизненным циклом данных о безопасности.

Оптимизированная модель гауссовой матрицы — рекуррентные нейронные сети

Точность активного анализа защищенности корпоративной сети зависит от того, могут ли данные о ситуационной осведомленности отражать динамические характеристики реального приложения, а также от предсказательной способности модели. Известно, что для управления компьютерной сетью с различными фазами работы одна гауссовская инженерная модель может не описать условия ее работы, а глобальное моделирование подвержено влиянию различных значений распределения. Поэтому для решения проблемы дисбаланса предлагаем использовать оптимизированную модель гауссовой матрицы — рекуррентные нейронные сети.

Исследование начинается с функциональной вероятностной регрессионной модели, использующей локальные индикаторные переменные. В этой модели алгоритм обучается на наборе данных $X \in RN$, полученном из одновершинного многомерного гауссовского распределения размера N. Если предположить, что плотность в k-м кластере равна $f_k = (X,\theta)$, то функция распределения гауссовской модели имеет вид, представленный в уравнении:

$$p(X,\psi) = \sum_{k=1}^{k} \pi_k f_k (X,\theta_k)$$

В уравнении π_k представляет собой вес каждого компонента k в наборе данных $\psi = (\pi_1,\ldots,\pi_k,\theta_1,\ldots,\theta_k)$. $\theta_k = (\mu_k,\Sigma_k),\mu_k,\Sigma_k$ распределены как среднее и ковариационная матрица каждого компонента k. Для оценки параметров модели гауссовской матрицы ψ обычно используется алгоритм максимизации ожиданий (EM), который включает в себя (E-шаг) и (M-шаг). Трудные или сложные задачи правдоподобия решаются путем итерации двух более простых шагов для получения оценки максимального правдоподобия. На первой итерации E-шага определяется условное ожидание log $L(\psi)$ для заданного набора данных X в виде уравнения:

$$E_{\psi(t)}\{\log L(\psi)|X\} = \\ = \sum_{k=1}^{k}\sum_{i=1}^{N} \tau_k^{(t)}(x_i,\psi^{(t)})\{\log\pi_k + \log f_k(x_i,\theta_k)\}$$

В уравнении $L(\psi)$ представляет собой оценку правдоподобия ψ, а log$L(\psi)$ — его логарифмическое правдоподобие. $\tau_k^{(t)}(x_i,\psi^{(t)})$ — обозначает апостериорную вероятность отнесения i-го экземпляра к k-му компоненту.

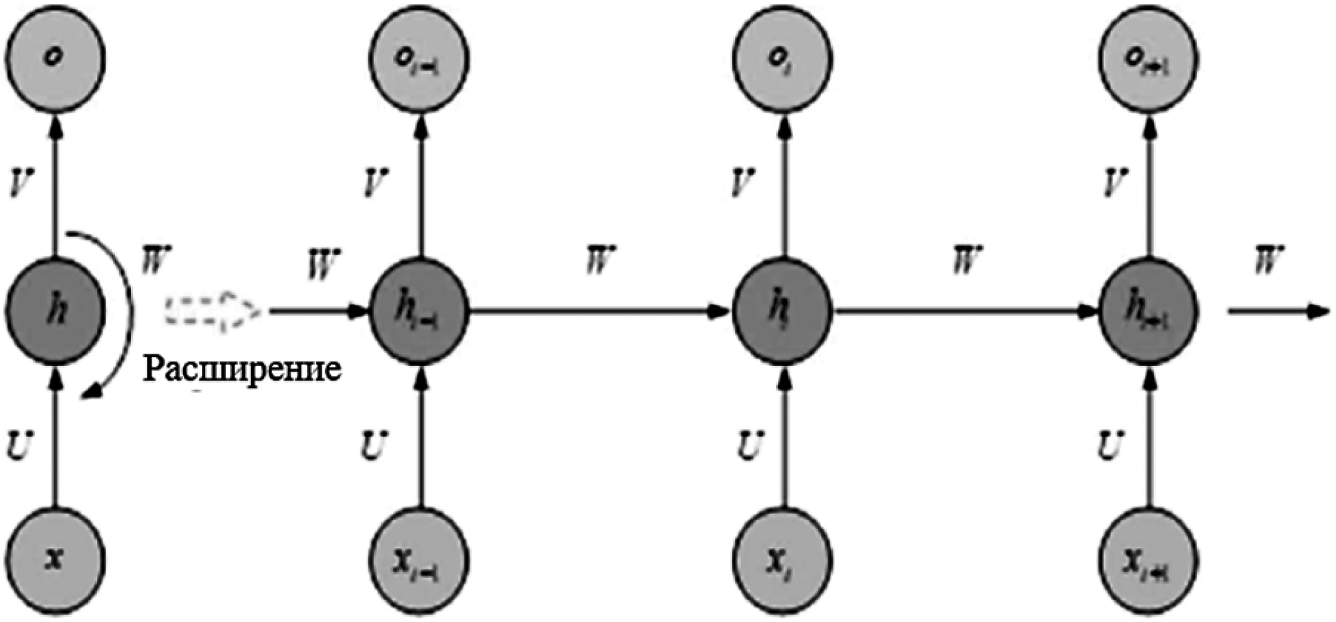

Рис. 3. Структура развертывания сети RNN

πk обозначает вес вероятности P на первой итерации с использованием k. Средний вектор μ и ковариация обновляются за M шагов. Итерация завершается, когда параметры сходятся или достигается максимальный размер шага, в результате чего получается средний вектор и ковариация гауссовской модели.

Рекуррентные нейронные сети (RNN) обладают свойством рекурсивной связности. В отличие от сетей с прямой передачей, RNN широко используются, поскольку они принимают только входные векторы фиксированной длины и выдают выходные векторы фиксированной длины. Как видно из рисунка 3, выход предыдущего временного шага является его входом, позволяя не только манипулировать последовательностью входных векторов, но и генерировать последовательность выходных векторов, сохраняя временную информацию в памяти и фиксируя долгосрочные зависимости между входными переменными.

Набор данных пропускается через сеть RNN для прогнозирования рисков и оценки безопасности, что позволяет создать оптимизированную модель анализа защищенности компьютерной сети и обнаружения вторжений.

Таким образом, в процессе исследования предложен усовершенствованный метод автоматического активного анализа защищенности корпоративной сети, основу которого составляет модель ситуационной осведомленности в области кибербезопасности и оптимизированная модель гауссовой матрицы — рекуррентные нейронные сети.

ЛИТЕРАТУРА

1. Сагдеева Л.А. Организация защиты корпоративной сети // Информационные технологии и системы: управление, экономика, транспорт, право. 2022. № S1. С. 11–12.
2. Jiao Wang Full-scene network security protection system based on ubiquitous power Internet of things // International Journal of Communication Systems. 2021. Volume 35, Issue 5. P. 87–93.
3. Hong-Tao Sun Neural-network-based event-triggered adaptive security path following control of autonomous ground vehicles subject to abnormal actuator signal // International Journal of Robust and Nonlinear Control. 2023. Volume 33, Issue 14. P. 23–29.
4. Тростянская Д.Н. Виртуальная частная сеть как средство защиты информации в корпоративных сетях // Известия Института менеджмента СГЭУ. 2022. № 1 (25). С. 131–133.
5. Зиненко О.А. Главные угрозы безопасности корпоративных сетей и как от них защититься // Защита информации. Инсайд. 2021. № 3 (99). С. 4–7.

Журнал «Современная наука: актуальные проблемы теории и практики»