УДК 004.056.5:004.738.2

Лебедев Иван Владимирович,
аспирант кафедры отечественные сети магистральной связи,
Московский технический университет связи и информатики,
г. Москва,

МЕТОДЫ ОБНАРУЖЕНИЯ АНОМАЛИЙ В ТРАФИКЕ СИСТЕМ УПРАВЛЕНИЯ НА ОСНОВЕ АВТОЭНКОДЕРОВ

Аннотация. *Статья посвящена исследованию методов обнаружения аномалий в сетевом трафике систем управления критической инфраструктурой на основе нейросетевых автоэнкодеров. Актуальность работы определяется тем, что традиционные сигнатурные системы обнаружения вторжений не справляются с выявлением неизвестных атак и тонких аномалий в управляющем трафике систем экстренного оповещения населения, компрометация которых в условиях чрезвычайной ситуации может привести к катастрофическим последствиям. Рассмотрены принципы работы автоэнкодера как детектора аномалий, основанные на гипотезе о том, что модель, обученная исключительно на нормальном трафике, будет воспроизводить его с малой ошибкой реконструкции, тогда как аномальный трафик породит существенно большую ошибку. Описана архитектура свёрточного автоэнкодера для анализа временны́х рядов сетевых метрик, методология формирования обучающей выборки и процедура выбора порога обнаружения. Проведено сравнение с методами Isolation Forest и One-Class SVM. Результаты эксперимента демонстрируют превосходство автоэнкодера по точности обнаружения аномалий при приемлемом уровне ложноположительных срабатываний.*

Ключевые слова: *автоэнкодер, обнаружение аномалий, сетевой трафик, системы управления, критическая инфраструктура, глубокое обучение, ошибка реконструкции, Isolation Forest, One-Class SVM, кибербезопасность.*

Lebedev Ivan Vladimirovich,
Postgraduate student of the Department of Domestic Backbone Communications Networks,
Moscow Technical University of Communications and Informatics,
Moscow,
e-mail: ivan_ivan_lebedev@mail.ru

METHODS FOR ANOMALY DETECTION IN CONTROL SYSTEM TRAFFIC BASED ON AUTOENCODERS

Abstract. *The article examines methods for detecting anomalies in network traffic of critical infrastructure control systems based on neural network autoencoders. The relevance of the work is determined by the fact that traditional signature-based intrusion detection systems fail to identify unknown attacks and subtle anomalies in the control traffic of emergency public notification systems, the compromise of which during an emergency can lead to catastrophic consequences. The principles of operation of an autoencoder as an anomaly detector are considered, based on the hypothesis that a model trained exclusively on normal traffic will reproduce it with a small reconstruction error, while anomalous traffic will produce a substantially larger error. The architecture of a convolutional autoencoder for analyzing time series of network metrics, the methodology for forming a training dataset, and the procedure for selecting a detection threshold are described. A comparison with Isolation Forest and One-Class SVM methods is conducted. The experimental results demonstrate the superiority of the autoencoder in anomaly detection accuracy at an acceptable level of false positive triggerings.*

Keywords: *autoencoder, anomaly detection, network traffic, control systems, critical infrastructure, deep learning, reconstruction error, Isolation Forest, One-Class SVM, cybersecurity.*

Введение

Системы управления каналами связи в инфраструктуре экстренного оповещения населения относятся к объектам критической информационной инфраструктуры и являются потенциальными целями кибератак. Специфика этих систем состоит в том, что управляющий трафик между модулями мониторинга, прогнозирования и исполнения носит строго регулярный характер: фиксированные интервалы опроса, предсказуемые форматы сообщений, ограниченный набор допустимых операций. Именно эта регулярность, будучи уязвимостью с точки зрения предсказуемости, одновременно открывает возможность для эффективного обнаружения отклонений от нормального поведения [1, 2].

Традиционные системы обнаружения вторжений (IDS) строятся на сигнатурном принципе: база правил описывает известные паттерны атак, и любое совпадение трафика с паттерном порождает алерт. Этот подход хорошо работает против известных угроз, но принципиально бессилен перед атаками нулевого дня и медленными, тщательно замаскированными воздействиями на управляющую логику системы. В контексте систем оповещения особую опасность представляют именно тонкие манипуляции — например, незначительные целенаправленные искажения телеметрии, которые не нарушают ни один из известных сигнатурных правил, но систематически смещают алгоритм выбора канала в сторону заведомо деградированного пути [3, 4].

Методы обнаружения аномалий, основанные на моделировании нормального поведения системы, предлагают принципиально иной подход. Вместо описания того, как выглядит атака, они описывают то, как выглядит норма, — и маркируют всё, что от неё отклоняется. Среди таких методов

нейросетевые автоэнкодеры занимают особое место: в отличие от статистических методов, они способны улавливать сложные нелинейные зависимости в многомерных временны́х рядах сетевых метрик, что делает их пригодными для анализа реального управляющего трафика [5, 6].

Цель данной работы — разработать и экспериментально оценить метод обнаружения аномалий в трафике систем управления каналами связи на основе свёрточного автоэнкодера, а также провести его сравнение с методами Isolation Forest и One-Class SVM по метрикам точности и оперативности обнаружения.

Результаты исследования

Принцип работы автоэнкодера как детектора аномалий

Автоэнкодер представляет собой нейронную сеть, обученную воспроизводить входные данные на выходе через узкое скрытое представление — «бутылочное горлышко». Архитектурно он состоит из двух частей: энкодера E, отображающего входной вектор $x \in \mathbb{R}^n$ в латентное пространство $z \in \mathbb{R}^k$ ($k \ll n$), и декодера D, восстанавливающего вход из латентного представления:

$$\hat{z} = E(x), \quad \hat{x} = D(\hat{z}), \quad \mathscr{L} = \|x - \hat{x}\|^2.$$

Ключевая гипотеза детектора аномалий состоит в следующем. Если модель обучена исключительно на образцах нормального трафика, она научится эффективно сжимать и восстанавливать именно нормальные паттерны. При подаче на вход аномального образца — трафика, соответствующего атаке или нештатному поведению, — модель будет вынуждена кодировать незнакомую структуру в пространство, «настроенное» под норму, и восстановление окажется неточным. Ошибка реконструкции $r(x) = \|x - \hat{x}\|^2$ служит скалярной мерой аномальности: если $r(x) > \tau$, где τ — калиброванный порог, наблюдение классифицируется как аномальное [5, 7].

Достоинство этого подхода применительно к управляющему трафику систем оповещения состоит в том, что он не требует разметки аномалий в обучающей выборке — задача практически неразрешимая для редких атак на

критическую инфраструктуру. Достаточно иметь репрезентативный набор образцов нормального трафика, собранных в штатном режиме эксплуатации [4, 6].

Архитектура свёрточного автоэнкодера

Для анализа временны́х рядов сетевых метрик разработан свёрточный автоэнкодер (Convolutional Autoencoder, CAE). Выбор свёрточной архитектуры обусловлен её способностью выявлять локальные паттерны в последовательностях независимо от их позиции, что особенно важно для трафика с регулярной структурой, где аномалия может быть локализована в любом временно́м окне наблюдения [7, 8].

Входом модели служит матрица формы (W, F), где W = 30 — ширина скользящего окна в шагах по 30 секунд (15 минут наблюдения), F = 6 — число признаков: задержка, потери пакетов, джиттер, интенсивность управляющего трафика (пакетов в секунду), число активных соединений, доля нестандартных по размеру пакетов. Последние три признака несут информацию именно о характере трафика, а не только о качестве канала, что расширяет возможности детектора [3, 8].

Энкодер состоит из трёх одномерных свёрточных блоков. Первый блок: Conv1D(filters=32, kernel=5, padding=same) → BatchNorm → ReLU → MaxPool(2). Второй блок: Conv1D(filters=64, kernel=3, padding=same) → BatchNorm → ReLU → MaxPool(2). Третий блок: Conv1D(filters=128, kernel=3, padding=same) → BatchNorm → ReLU. После третьего блока применяется операция Flatten, и полносвязный слой Dense(16) формирует латентное представление размерностью 16. Декодер зеркально воспроизводит структуру энкодера с заменой MaxPool на операцию UpSampling и Conv1DTranspose. Функция потерь — MSE по всем элементам входной матрицы. Оптимизатор — Adam, скорость обучения 5×10^{-4} [7, 9].

Формирование обучающей выборки и выбор порога

Обучающая выборка формируется исключительно из образцов нормального трафика, собранных за период штатной эксплуатации системы управления. Для синтетического эксперимента нормальный трафик моделируется как стационарный процесс с характеристиками, типичными для периодического опроса SNMP-агентов и обмена телеметрией по gRPC: регулярные пакеты фиксированного размера с незначительными гауссовыми флуктуациями метрик. Размер обучающей выборки — 50 000 скользящих окон, что соответствует примерно 17 суткам непрерывной записи при шаге 30 секунд [4, 6].

Выбор порога τ является критически важным этапом: заниженный порог ведёт к недопустимому числу ложных тревог в промышленной эксплуатации, завышенный — к пропуску реальных атак. В данной работе порог калибруется на отложенной валидационной выборке нормального трафика по правилу трёх сигм: $\tau = \mu_r + 3\sigma_r$, где μ_r и σ_r — выборочное среднее и стандартное отклонение ошибки реконструкции на нормальном трафике. Этот выбор обеспечивает теоретический уровень ложноположительных срабатываний порядка 0.3% при нормальном распределении ошибки реконструкции, что на практике является приемлемым для систем безопасности [5, 10].

Базовые методы и сценарии аномалий

Для сравнения реализованы два широко применяемых метода обнаружения аномалий без учителя. Isolation Forest строит ансамбль случайных деревьев изоляции: аномальные наблюдения изолируются на меньшей глубине дерева, чем нормальные, поскольку занимают редкие области пространства признаков. Метод эффективен для многомерных данных и устойчив к шуму, однако не учитывает временну́ю структуру последовательностей [9]. One-Class SVM строит гиперсферу минимального объёма в пространстве признаков, охватывающую нормальные наблюдения; точки вне сферы классифицируются

как аномалии. Метод чувствителен к выбору ядра и плохо масштабируется на большие выборки [10].

Для тестирования разработаны пять сценариев аномалий, отражающих реалистичные угрозы управляющему трафику. Первый сценарий — «отравление телеметрии»: агент мониторинга начинает систематически занижать значения задержки на 40%, имитируя атаку на целостность данных. Второй сценарий — «сканирование портов»: резкий рост числа соединений с нетипичных адресов при нормальных значениях QoS-метрик. Третий сценарий — «медленная атака»: постепенное увеличение интенсивности управляющих запросов на 5% каждые десять шагов — сценарий, специально разработанный для обхода пороговых детекторов. Четвёртый сценарий — «replay-атака»: циклическое воспроизведение ранее записанного нормального трафика со смещённой временно́й меткой. Пятый сценарий — «инъекция команды»: единичный нетипичный пакет управления среди нормального трафика [3, 6].

Результаты сравнительного эксперимента

В таблице 1 приведены значения основных метрик качества обнаружения для каждого из пяти сценариев аномалий. Precision характеризует долю истинных аномалий среди всех помеченных детектором, Recall — долю реальных аномалий, обнаруженных детектором, F1-score — их гармоническое среднее. Detection Lag — среднее запаздывание между началом аномалии и первым алертом в шагах.

Таблица 1. Сравнение методов обнаружения аномалий по сценариям

Сценарий / Метод	Метод	Precision	Recall	F1	Lag (шагов)
Отравление телеметрии	Isolation Forest	0.71	0.64	0.67	4.2
	One-Class SVM	0.68	0.59	0.63	5.1
	CAE (предлаг.)	**0.89**	**0.84**	**0.86**	**2.1**
Сканирование портов	Isolation Forest	0.83	0.77	0.80	2.8

Сценарий / Метод	Метод	Precision	Recall	F1	Lag (шагов)
	One-Class SVM	0.79	0.71	0.75	3.4
	CAE (предлаг.)	**0.92**	**0.91**	**0.91**	**1.3**
Медленная атака	Isolation Forest	0.54	0.48	0.51	9.7
	One-Class SVM	0.51	0.43	0.47	11.2
	CAE (предлаг.)	**0.78**	**0.72**	**0.75**	**5.4**
Replay-атака	Isolation Forest	0.61	0.53	0.57	6.3
	One-Class SVM	0.58	0.49	0.53	7.8
	CAE (предлаг.)	**0.84**	**0.79**	**0.81**	**3.2**
Инъекция команды	Isolation Forest	0.74	0.55	0.63	3.1
	One-Class SVM	0.70	0.51	0.59	3.9
	CAE (предлаг.)	**0.88**	**0.83**	**0.85**	**1.7**

Анализ данных таблицы 1 позволяет сделать ряд наблюдений. Во-первых, свёрточный автоэнкодер устойчиво превосходит оба базовых метода по всем метрикам во всех сценариях. Разрыв особенно заметен для сценариев «отравление телеметрии» и «replay-атака» — именно там, где аномалия выражается в тонком нарушении временно́й структуры трафика, которую Isolation Forest и One-Class SVM не моделируют в явном виде, обрабатывая каждое наблюдение независимо [5, 9].

Во-вторых, медленная атака ожидаемо оказывается наиболее трудным сценарием для всех трёх методов: постепенное изменение, не выходящее за пределы нормального диапазона на каждом отдельном шаге, требует накопления достаточной статистики для надёжного обнаружения. Тем не менее автоэнкодер и здесь обеспечивает Detection Lag 5.4 шага — при периоде 30 секунд это 2.7 минуты, что оставляет оператору время для реагирования. Isolation Forest в аналогичном сценарии запаздывает почти вдвое — на 9.7 шага [6, 9].

В-третьих, сканирование портов обнаруживается всеми тремя методами с наилучшими показателями F1 — это объясняется тем, что данный тип аномалии порождает резкий, легко различимый выброс в признаке «число активных соединений», хорошо захватываемый любым методом на основе статистики признакового пространства. Однако и здесь автоэнкодер быстрее реагирует: Detection Lag 1.3 шага против 2.8 у Isolation Forest [9, 10].

Практические аспекты развёртывания

Время инференса обученного автоэнкодера на одном окне наблюдения составляет в среднем 1.8 мс на CPU, что вполне вписывается в требования к оперативности обнаружения при периоде опроса 30 секунд. Модель компактна: суммарное число параметров — около 95 тысяч, размер сериализованной модели — порядка 730 КБ. Это допускает её размещение непосредственно на узле управления без использования специализированного аппаратного обеспечения [7, 8].

Отдельного внимания заслуживает вопрос концепции реагирования на алерты. Учитывая ненулевой уровень ложноположительных срабатываний, прямая автоматическая блокировка по алерту детектора нецелесообразна — она может парализовать управляющую плоскость системы оповещения в ложно интерпретированный критический момент. Более обоснованной является двухуровневая реакция: при единичном алерте система переходит в режим повышенного журналирования и активирует дополнительную верификацию команд управления; при серии алертов на протяжении K последовательных окон инициируется переход в детерминированный резервный режим с уведомлением оператора [3, 4].

Заключение

В рамках проведённого исследования разработан и экспериментально верифицирован метод обнаружения аномалий в трафике систем управления каналами связи на основе свёрточного автоэнкодера. Фундаментальное

преимущество предложенного подхода перед сигнатурными методами заключается в его способности выявлять неизвестные атаки и тонкие манипуляции, не описанные ни в каких правилах, — за счёт моделирования нормального поведения системы и детектирования отклонений от него по ошибке реконструкции.

Сравнительный эксперимент на пяти сценариях аномалий показал, что свёрточный автоэнкодер стабильно превосходит методы Isolation Forest и One-Class SVM по метрикам F1 и Detection Lag. Наиболее существенный выигрыш достигается для аномалий, выражающихся в нарушении временно́й структуры трафика — сценариев «отравление телеметрии» и «replay-атака», — где преимущество свёрточной архитектуры, явно моделирующей последовательную природу данных, проявляется в наибольшей мере.

Практическая применимость метода подтверждается компактностью модели и низкими вычислительными требованиями к инференсу. Вместе с тем результаты указывают на ограничение предложенного подхода: медленные атаки, намеренно оставающиеся в границах нормального диапазона на каждом отдельном шаге, обнаруживаются с заметным запаздыванием. Перспективным направлением развития является интеграция автоэнкодера с методами анализа долгосрочных трендов — в частности, совместное применение с LSTM-предиктором, отслеживающим дрейф распределения метрик на горизонте нескольких часов. Такая гибридная архитектура способна обеспечить как оперативное обнаружение резких аномалий, так и своевременное выявление медленных целенаправленных воздействий на управляющую логику системы оповещения.

Литература

1. О безопасности критической информационной инфраструктуры Российской Федерации : Федеральный закон от 26.07.2017 № 187-ФЗ // Собрание законодательства РФ. – 2017. – № 31 (ч. I). – Ст. 4736.

2. Об утверждении Положения о системах оповещения населения : постановление Правительства РФ от 01.03.2019 № 1074 // Собрание законодательства РФ. – 2019. – № 10. – Ст. 1028.
3. Шарахутдинов Р. Б., Черкесов В. В. Анализ современных технологий и технических средств оповещения и информирования населения о чрезвычайных ситуациях // Пожарная и техносферная безопасность: проблемы и пути совершенствования. – 2020. – № 2 (6). – С. 485–489.
4. Ахмедова Х. М. Организационно-правовое регулирование информирования и оповещения населения при угрозе и возникновении чрезвычайных ситуаций // Молодой учёный. – 2022. – № 38 (433). – С. 87–89.
5. Goodfellow I., Bengio Y., Courville A. Deep Learning. – Cambridge: MIT Press, 2016. – 800 p.
6. Chandola V., Banerjee A., Kumar V. Anomaly Detection: A Survey // ACM Computing Surveys. – 2009. – Vol. 41, No. 3. – P. 1–58.
7. Paszke A. et al. PyTorch: An Imperative Style, High-Performance Deep Learning Library // Advances in Neural Information Processing Systems 32. – 2019. – P. 8026–8037.
8. LeCun Y., Bengio Y., Hinton G. Deep Learning // Nature. – 2015. – Vol. 521. – P. 436–444.
9. Liu F. T., Ting K. M., Zhou Z.-H. Isolation Forest // 2008 Eighth IEEE International Conference on Data Mining. – 2008. – P. 413–422.
10. Schölkopf B. et al. Estimating the Support of a High-Dimensional Distribution // Neural Computation. – 2001. – Vol. 13, No. 7. – P. 1443–1471.