Обнаружение атак на киберфизические системы на основе глубокого обучения

Людмила Сухостат
Институт Информационных Технологий НАНА, Баку, Азербайджан
lsuhostat@hotmail.com

***Аннотация*— Частота и серьезность атак на киберфизические системы требуют разработки новых подходов обнаружения и локализации кибератак. В работе предлагается подход, сочетающий в себе преимущества вариационного автоэнкодера и «остаточной» нейронной сети. Предлагаемый подход тестируется на наборе данных системы газопроводов, собранном в Центре защиты критически важных инфраструктур штата Миссисипи (США). Он сравнивается с логистической регрессией и конволюционной нейронной сетью. Сравнительный анализ показывает, что предлагаемый подход может идентифицировать практически все атаки, присутствующие в наборе данных.**

Ключевые слова— информационная безопасность, киберфизические системы, кибератаки, глубокое обучение, вариационный автоэнкодер, «остаточная» нейронная сеть.

I. Введение

Широкое внедрение киберфизических систем (КФС) формирует процесс объединения технологий и знаний, обеспечивая автономию, надежность и систематичность [1]. Быстрый рост приложений КФС приводит к необходимости обеспечения информационной безопасности таких систем.

Следовательно, возникает потребность в оценке влияния кибератак на нормальное функционирование физических процессов. При этом важно не только проанализировать разрушительное воздействие кибератак и их последствия, но и обеспечить доступность информации о киберактивности. Важно заранее разработать алгоритмы обнаружения и контрмеры для всех известных атак, чтобы уменьшить влияние атак в течение ограниченного времени и свести к минимуму возможный причиняемый ущерб системы.

Кибератаки могут быть как преднамеренными, так и непреднамеренного действия, дифференциация которых в больших и сложных КФС бывает порой невозможна. В связи с этим возрастает потребность в разработке новых подходов для обнаружения аномального поведения системы. Эффективная киберфизическая безопасность достигается посредством стратегии защиты, которая направлена на обнаружение вторжений в разных областях КФС.

Глубокое обучение рассматривается как современный подход искусственного интеллекта и включает в себя ряд алгоритмов и решений, которые значительно расширяют сферу действия и эффективность нейронных сетей [2-4]. Наличие большого количества слоев позволяет нейронной сети построить концепцию объекта исследования от простых признаков, постепенно переходящих к более сложным.

В данной статье предлагается подход, сочетающий в себе преимущества вариационного автоэнкодера (variational autoencoder, VAE) и одномерной глубокой «остаточной» нейронной сети (deep residual neural network, ResNet), как эффективный метод обнаружения кибератак в КФС. Предложенная модель была протестирована с использованием набора данных системы газопроводов, собранного в Центре защиты критически важных инфраструктур штата Миссисипи [5]. Предлагаемый подход сравнивается с логистической регрессией [6], конволюционной нейронной сетью [7] до и после применения энкодера из VAE.

II. Вариационный Автоэнкодер

VAE представляет собой разновидность автоэнкодера, которая также состоит из энкодера и декодера. Энкодер оценивает скрытые переменные из переменных наблюдения, а декодер восстанавливает переменные наблюдения из скрытых переменных.

Пусть x - вектор переменных наблюдения, z - скрытый вектор, а ϕ и θ - наборы параметров энкодера и декодера, соответственно. Вектор наблюдения x вводится в энкодер, на выходе которого получают математическое ожидание $\mu_\phi(x)$ и дисперсию $\sigma^2_\phi(x)$, которые используются для сэмплирования (sampling) скрытого вектора z [8].

z сэмплируется после распределения Гаусса с $\mu_\phi(x)$ и $\sigma^2_\phi(x)$ [9, 10]. Декодер $p_\theta(x|z)$ генерирует x из z. Затем непосредственно оцениваются $\mu_\phi(z)$ и $\sigma^2_\phi(z)$ для генерации выходного вектора x.

Градиенты рассчитываются на основе $p_\theta(x)$:

DOI: 10.25045/NCInfoSec.2018.07

$$p_\theta(x) = \int p_\theta(z) p_\theta(x \mid z) dz \,. \quad (1)$$

III. «Остаточная» Нейронная Сеть

С увеличением числа слоев в глубоких нейронных сетях ошибка обучения возрастает, а производительность сети ухудшается. Это указывают на то, что в современных методах обучения нейронной сети с большим числом слоев по-прежнему существует множество проблем. Чтобы решить эти проблемы была разработана ResNet [11]. Она состоит из серии остаточных модулей. Формула расчета для выхода n-го остаточного модуля следующая:

$$x_n = x_{n-1} + F(x_{n-1}; W_n) \,, \quad (2)$$

где $F(x_{n-1}; W_n)$ - остаток в сети, а W_n - параметр сети.

Во время обучения сеть напрямую не вычисляет выход (x_n), а вычисляет остаточное значение между x_n и входом x_{n-1}, обозначаемое как F.

Однако ResNet с большим числом слоев требует больших вычислительных ресурсов [12]. По этой причине в работе с целью уменьшения пространства признаков, содержащихся в наборах данных, предлагается использовать VAE.

IV. Предлагаемый Подход

Предлагаемый подход включает следующие этапы:

1. Получение данных для обучения и оценки эффективности модели.

2. Предварительная обработка данных. Извлечение признаков на основе энкодера из VAE, состоящего из полносвязных слоев (Dense) (Рис. 1)

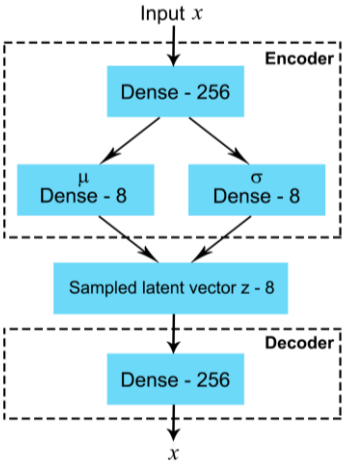


Рис. 1. Структура вариационного автоэнкодера.

3. Построение одномерной ResNet (1DResNet) (Рис. 2) с целью обнаружения кибератак. Сеть состоит из 25 слоев. Она включает конволюционные (convolutional) и пулинг (maxpooling) слои. Для более быстрого обучения были добавлены слои Batch Normalization. a, b и c обозначают число выходных фильтров в конволюционных слоях. В первых трех блоках a=64, b=64 и c=64. A в последних четырех они равны 128, 128 и 512, соответственно.

4. Обучение полученной модели для нахождения оптимального решения.

5. Оценка полученной модели на тестовом наборе данных.

V. Описание Базы Данных

Для проведения экспериментов была рассмотрена база данных системы газопроводов, собранная в Центре защиты критически важных инфраструктур штата Миссисипи США [5]. Лог файл данных сетевого трафика был взят из системы SCADA (Supervisory Control And Data Acquisition) для лабораторного стенда системы газопроводов, включающего как нормальную работу системы, так и реальные кибератаки. Сетевые пакеты имеют временную метку и записываются в лог файл с уникальными признаками, в т.ч. автоматический режим работы системы (вкл/выкл), управление насосом и клапаном соленоида и др.

В наборе данных есть четыре основные категории атак: атаки путем внедрения команд (command injection attacks), атаки путем внедрения данных, отказ в обслуживании (Denial of Service, DoS) и зондирование (reconnaissance). Эти четыре категории дополнительно подразделяются на семь конкретных типов атак: атака внедрения злоумышленного отклика (naïve malicious response injection, NMRI), атака комплексной вредоносной реакции (complex malicious response injection, CMRI), внедрение команд вредоносного состояния (malicious state command injection, MSCI), внедрение команд с вредоносными параметрами (malicious parameter command injection, MPCI), внедрение вредоносного кода (malicious function code command injection, MFCI), DoS и reconnaissance (Recon.). Набор данных содержит 214580 стандартных сетевых пакетов (Normal) и 60048 пакетов с атаками [2].

VI. Метрики Оценки Методов Классификации

Для оценки производительности классификаторов используются следующие метрики: accuracy, recall, precision и F-measure. Для любого алгоритма классификации возможны четыре классификационных случая, и это помогает понять разницу между рассматриваемыми метриками: истинно-положительные результаты (True Positives, TP), ложно-положительные результаты (False Positives, FP), истинно-отрицательные результаты (True Negatives, TN) и ложно-отрицательные результаты (False Negatives, FN).

Рассматриваемые метрики определяются следующим образом:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \,, \quad (3)$$

которая определяет долю правильных результатов, полученных классификатором.

$$precision = \frac{TP}{TP + FP}. \quad (4)$$

Эта метрика показывает, какая доля объектов, выделенных классификатором как положительные, действительно является положительной.

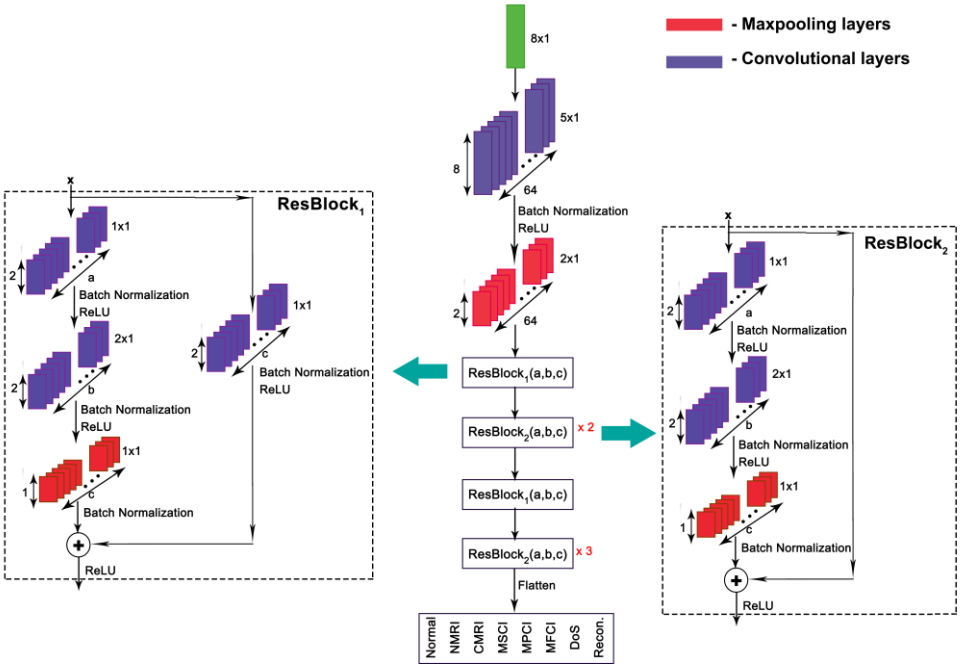


Рис. 2. Структура 1DResNet.

$$recall = \frac{TP}{TP + FN}, \quad (5)$$

которая показывает, какая часть положительных объектов была выделена классификатором.

Следующая метрика сочетает в себе меры точности и полноты:

$$F - measure = \frac{2 \times recall \times precision}{recall + precision}. \quad (6)$$

VII. Результаты Экспериментов

Эксперименты проводились на Intel Xeon (R), CPU X5670 @ 2.93 ГГц×4, 10 Гб RAM. Предлагаемый подход оценивался на языке Python 2.7.13 с использование различных библиотек, включая Tensorflow и Keras.

Тенденция вариационной функции потерь при применении VAE показана на Рис. 3.

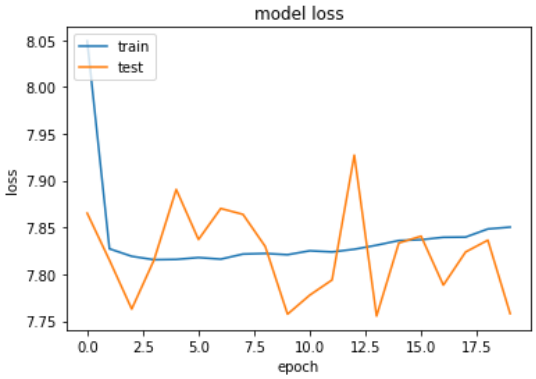
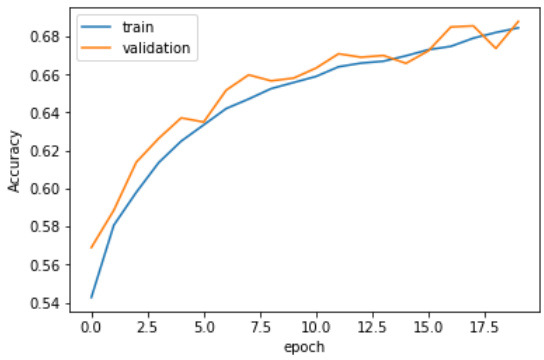


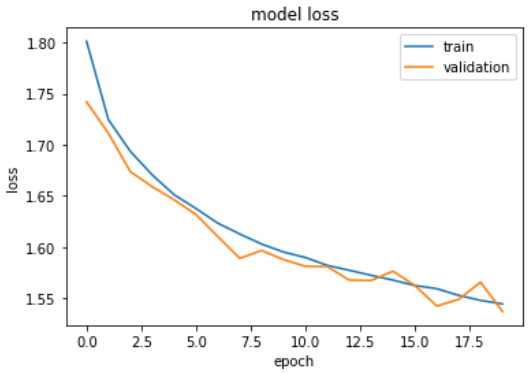
Рис. 3. Вариационная функция потерь.

Результаты сравнения предлагаемого подхода с логистической регрессией (LR) и конволюционной нейронной сетью (CNN) показаны в таблице I.

На Рис. 4 показаны точность и ошибка кросс-энтропии VE+1DResNet в зависимости от номера итерации.



а)



б)

Рис. 4. Точность и функция потерь для VE+1DResNet.

ТАБЛИЦА I. СРАВНЕНИЕ ПРЕДЛОЖЕННОГО ПОДХОДА С ЛОГИСТИЧЕСКОЙ РЕГРЕССИЕЙ И КОНВОЛЮЦИОННОЙ НЕЙРОННОЙ СЕТЬЮ

	LR	**CNN**	**1DResNet**	**VE+ LR**	**VE+CNN**	**VE+1DResNet**
Точность обучения	59.33%	63.29%	70.31%	47.54%	60.09%	**68.77%**
Точность тестирования	59.28%	63.11%	65.32%	47.51%	59.04%	**68.66%**

ТАБЛИЦА II. СРАВНЕНИЕ РЕЗУЛЬТАТОВ КЛАССИФИКАЦИИ

Метод	**Метрика**	**Normal**	**NMRI**	**CMRI**	**MSCI**	**MPCI**	**MFCI**	**DoS**	**Recon.**
LR	*Precision*	58.63%	45.72%	50.07%	49.19%	47.31%	87.13%	50.88%	**100%**
	Recall	34.61%	**90.05%**	12.42%	47.04%	53.06%	**100%**	61.15%	75.52%
	F-measure	43.53%	60.65%	19.91%	48.09%	50.02%	93.12%	55.54%	86.05%
CNN	*Precision*	78.07%	48.92%	45.06%	73.07%	42.13%	89.95%	**78.45%**	**100%**
	Recall	36.92%	9.26%	92.06%	38.76%	**89.63%**	**100%**	52.74%	85.81%
	F-measure	50.13%	15.57%	60.50%	50.66%	57.31%	94.71%	63.07%	92.37%
1DResNet	*Precision*	**93.67%**	49.61%	49.21%	76.44%	70.59%	**100%**	57.26%	66.40%
	Recall	**51.02%**	57.50%	52.44%	59.57%	64.02%	50.19%	88.97%	**98.96%**
	F-measure	**66.06%**	53.27%	50.77%	**66.96%**	67.15%	66.84%	69.68%	79.47%
VE+ LR	*Precision*	46.25%	45.81%	43.67%	39.28%	37.93%	78.05%	35.70%	60.23%
	Recall	31.89%	51.14%	49.86%	57.65%	21.61%	64.29%	40.85%	62.98%
	F-measure	37.75%	48.33%	46.56%	46.72%	27.53%	70.50%	38.10%	61.57%
VE+CNN	*Precision*	90.60%	52.31%	45.12%	44.72%	39.20%	94.17%	58.84%	**100%**
	Recall	31.76%	7.27%	**94.49%**	50.85%	60.30%	**100%**	45.86%	90.48%
	F-measure	47.03%	12.76%	61.08%	47.52%	47.52%	97.00%	51.55%	95.00%
VE+1DResNet	*Precision*	92.20%	**55.29%**	**52.12%**	**78.24%**	**73.44%**	97.60%	74.05%	**100%**
	Recall	50.52%	74.26%	63.09%	**68.28%**	63.67%	**100%**	**88.98%**	**98.96%**
	F-measure	63.58%	**60.67%**	**61.53%**	**66.96%**	**68.20%**	**98.79%**	**69.70%**	**98.21%**

Предлагаемый подход (VE+1DResNet) распознал все 7 классов атак и нормальное состояние системы в отличие от других рассмотренных методов. Он незначительно уступил 1DResNet в классификации согласно трем метрикам для класса Normal (Таблица II).

VIII. ЗАКЛЮЧЕНИЕ

Обнаружение вторжений является одной из серьезных проблем в области обеспечения информационной безопасности КФС. В этом исследовании был предложен подход VE+1DResNet, сочетающий в себе преимущества VAE и одномерной ResNet.

В целом, предложенный подход показал высокую точность обнаружения атак в ходе проведения экспериментов. Подводя итоги, можно сделать вывод, что анализ оценки предложенного подхода, основанного на глубоком обучении, может быть полезен для будущих исследований и идентификации атак.

ЛИТЕРАТУРА

[1] R. Alguliyev, Y. Imamverdiyev, L. Sukhostat, “Cyber-physical systems and their security issues,” Computers in Industry, Vol. 100, pp. 212–223, 2018.

[2] C. Feng, T. Li, D. Chana, “Multi-level anomaly detection in industrial control systems via package signatures and LSTM networks,” Proc. of the 47th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN), pp. 261–272, 2017.

[3] J. Goh, S. Adepu, M. Tan, Z.S. Lee, “Anomaly detection in cyber physical systems using recurrent neural networks,” Proc. of the IEEE 18th International Symposium on High Assurance Systems Engineering (HASE), pp. 140-145, 2017.

[4] J. Shin, Y. Baek, J. Lee, S. Lee, “Cyber-physical attack detection and recovery based on RNN in automotive brake systems,” Preprints, pp. 1-21, 2018.

[5] T. Morris, R. Vaughn, Y.S. Dandass, “A testbed for SCADA control system cybersecurity research and pedagogy,” Proc. of the 7th Annual Workshop on Cyber Security and Information Intelligence Research, pp. 1-4, 2011.

[6] D. Hosmer, S. Lemeshow, “Applied Logistic Regression,” John Wiley & Sons, Inc., 2000. 373 p.

[7] Y. LeCun, Y. Bengio, “Convolutional networks for images, speech, and time series,” The handbook of brain theory and neural networks, pp. 255-257, 1995.

[8] H. Nishizaki, “Data augmentation and feature extraction using variational autoencoder for acoustic modeling,” Proc. of APSIPA Annual Summit and Conference, pp. 1222-1227, 2017.

[9] D.P. Kingma, M. Welling, “Auto-encoding variational bayes,” Proc. of International Conference on Learning Representations (ICLR), pp. 1-14, 2014.

[10] D.P. Kingma, T. Salimans, M. Welling, “Variational dropout and the local reparameterization trick,” Proc. of the 29th Annual Conference on Neural Information Processing Systems (NIPS), pp. 2575–2583, 2015.

[11] K. He, X. Zhang, S. Ren, J. Sun, “Deep residual learning for image recognition,” Proc. of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778, 2016.

[12] C. Szegedy, S. Ioffe, V. Vanhouche, A. Alemi, “Inception-v4, inception-resnet and the impact of residual connections on learning,” Proc. of the AAAI Conference on Artificial Intelligence, 4278–4284, 2017.

DETECTION OF ATTACKS ON CYBER-PHYSICAL SYSTEMS BASED ON DEEP LEARNING

Lyudmila Sukhostat

Institute of Information Technology of ANAS, Baku, Azerbaijan

lsuhostat@hotmail.com

Abstract – The frequency and severity of attacks on cyber-physical systems require the development of new approaches for cyber-attacks detection and localization. The paper proposes an approach that combines the advantages of a variational autoencoder and a residual neural network. The proposed approach is evaluated on a gas pipeline system dataset collected at the Critical Infrastructure Protection Center of the Mississippi State University (USA). It is compared with logistic regression and a convolutional neural network. A comparative analysis shows that the proposed approach can identify almost all attacks in the dataset.

Keywords – information security, cyber-physical systems, cyber-attacks, deep learning, variational autoencoder, residual neural network.