Разработка Метода Обнаружения Киберугроз с Использованием Обработки Естественного Языка и Усовершенствованной Модели XLNet

Рамиз Алыгулиев[1], Людмила Сухостат[2]
[1,2]Институт Информационных Технологий, Баку, Азербайджан
[1]*r.aliguliyev@gmail.com*, [2]*lsuhostat@hotmail.com*

***Аннотация*— Рост числа киберугроз в мире требует их своевременного обнаружения с целью повышения кибербезопасности. Для определения наличия киберугроз широко применяются методы обработки естественного языка. Распознавание именованных сущностей позволяет более точно определять различные киберуязвимости. Однако анализ текстовых данных по-прежнему остается сложной задачей. Для решения этих проблем в данной работе предлагается метод обнаружения киберугроз с использованием обработки естественного языка и усовершенствованной модели XLNet. Эксперименты проводятся на двух наборах данных: DNRTI и APTNER. Сравнение с другими известными методами доказывает эффективность данного метода и обеспечивает более надежное и устойчивое обнаружение киберугроз, устраняя ограничения рассматриваемых методов.**

Ключевые слова— интеллектуальный анализ киберугроз; кибербезопасность; обработка естественного языка; глубокое обучение.

I. Введение

В настоящее время различные организации сталкиваются с растущим числом киберугроз (DDoS-атаки, фишинг, программы-вымогатели и т. д.) [1]. Это приводит к необходимости извлечения информации о потенциальных кибератаках и уязвимостях [2].

Информация об угрозах кибербезопасности из различных источников, таких как публикации в социальных сетях, электронная почта, онлайн-форумы, блоги по кибербезопасности и т. д. может быть получена путем интеллектуального анализа киберугроз (Cyber Threat Intelligence, CTI). Эти публикации предоставляют важную информацию, связанную с кибербезопасностью. CTI представляет собой один из ключевых компонентов кибербезопасности, предоставляя информацию о потенциальных и реальных киберугрозах для улучшения механизмов защиты [3].

Инструменты обработки естественного языка (Natural Language Processing, NLP) фокусируются на взаимодействии между компьютерным программным обеспечением и человеческим языком и используются для анализа отчетов CTI. NLP основан на применении моделей глубокого обучения и машинного обучения [4]. Методы NLP нашли широкое применение в различных областях, включая кибербезопасность [5, 6]. Однако существующие методы NLP сталкиваются с определенными проблемами при извлечении информации из отчетов CTI [7]. Поэтому необходимо разработать подход на основе NLP, специфичный для области кибербезопасности, с использованием современных инструментов глубокого и машинного обучения.

Инструменты NLP ориентированы на область кибербезопасности и используют современные инструменты глубокого и машинного обучения. Распознавание именованных сущностей (Named Entity Recognition, NER) — важная задача для извлечения значимой информации из неструктурированного текста, помогающая анализировать киберугрозы и улучшать разведку угроз [3, 8].

Большинство существующих методов могут быть неэффективными, особенно при работе с большими неструктурированными наборами текстовой информации. Разработка новых подходов позволяет принимать своевременные решения и последующие контрмеры для борьбы с существующими и неизвестными киберугрозами.

Основная цель данной работы — разработка эффективной модели извлечения сущностей из текстовой информации, связанной с кибербезопасностью. Предложенный подход основан на XLNet, сверточной нейронной сети и условных случайных полях. Текстовая информация из различных источников киберугроз (таких как отчеты об угрозах, статьи в блогах, оповещения об уязвимостях нулевого дня, новостные статьи и социальные сети) подается на вход модели. Предлагаемый подход позволяет эффективно обнаруживать именованные сущности, связанные с киберугрозами, для улучшения защиты и автоматизации реагирования на кибер-инциденты и угрозы.

Основные результаты данной работы включают следующее:

- Разработана предварительно обученная модель на основе XLNet, адаптированная для области кибербезопасности.
- Эксперименты проводились на больших наборах данных (DNRTI и APTNER).

DOI: 10.25045/forensic.gov.2026.02

- Сравнение с известными моделями доказывает эффективность предложенного подхода к обнаружению именованных сущностей на основе текстовых данных связанных с кибербезопасностью.

II. Предлагаемый Подход

В этом разделе описывается предложенный подход к интеллектуальному анализу данных о киберугрозах. На первом этапе текстовые данные собираются из различных источников, таких как отчеты об угрозах, новостные статьи, СМИ, статьи в блогах онлайн и др. Они могут помочь предотвратить и обнаружить потенциальные угрозы кибербезопасности.

Далее полученная текстовая информация подается на блок предварительной обработки текстовых данных, который включает этапы сегментации предложений, токенизации, удаления избыточных символов, стоп-слов и лемматизации (Рис. 1).

Для аннотирования текстовых сущностей применяется метод разметки «Начало-Внутри-Снаружи» (Beginning-Inside-Outside, BIO) [9]. В результате слово помечается как начало именованной сущности (B) или внутреннее (I). Слова, не упомянутые в текстовом наборе данных, помечаются как O.

На следующем этапе для извлечения именованных сущностей о киберугрозах применяется модель XLNet-CNN-CRF.

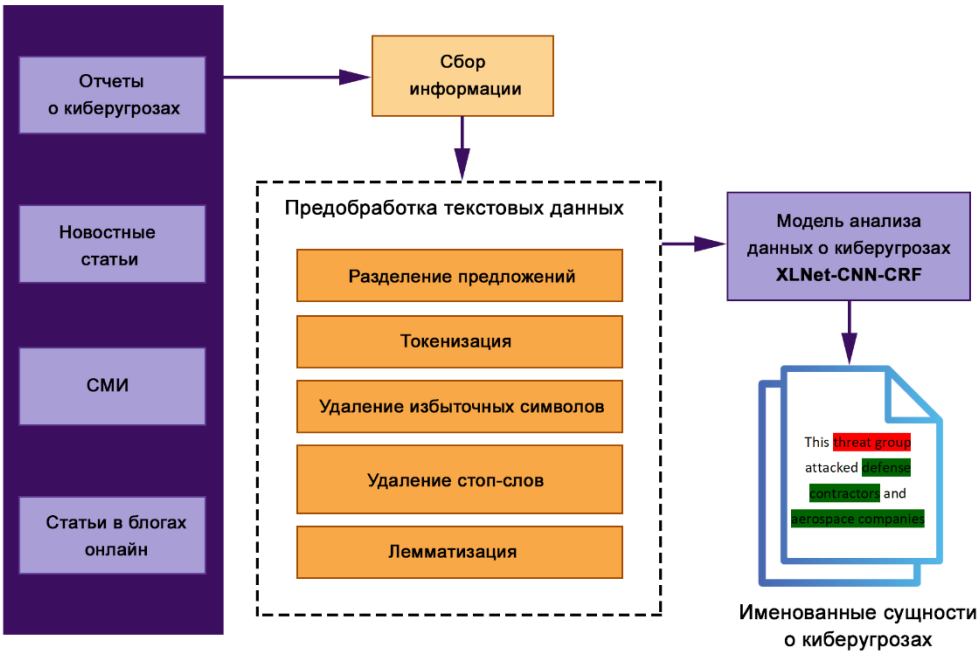

Рисунок 1. Общая схема предлагаемого подхода

Предложенная модель XLNet-CNN-CRF состоит из трех основных блоков. После предварительной обработки вектор слов, содержащий информацию о киберугрозах подается на слой XLNet ($g_1^{(2)}, g_2^{(2)}, g_3^{(2)}, \ldots, g_N^{(2)}$, где N - длина входных данных). Главное преимущество модели XLNet заключается в том, что она модифицирует целевую функцию обучения языковой модели, которая изучает условные распределения для всех перестановок токенов в последовательности [10]:

$$\max_{\theta} \left[E_{z \sim Z_N} \left[\sum_{i=1}^{N} log\, p_\theta \left(x_{z[i]} | x_{z[<i]} \right) \right] \right], \quad (1)$$

где θ - параметры модели, X - текстовая последовательность, Z - множество всех перестановок последовательности длиной T, z - последовательность из Z, $z[i]$ и $z[< i]$ обозначают i и $(i-1)$ элементы перестановки.

Выход слоя XLNet подается на блок CNN (Convolutional Neural Network), который включает четыре одномерных сверточных слоя и четыре слоя максимального пулинга. Выход одномерного сверточного слоя вычисляется следующим образом:

$$y_i = \alpha(w_i y_{i-1} + b_i), \quad (2)$$

где y_i - выход i слоя, y_{i-1} - выход предыдущего слоя, w_i матрицы весов, α - функция активации и b_i - векторы смещения.

Слой CRF (Conditional Random Field) обеспечивает точные и согласованные последовательности меток, эффективно обрабатывая сложность и изменчивость текстовой информации о киберугрозах.

$$s = \underset{s'}{argmax} P(s'|o), \quad (3)$$

где $P(s'|o)$ - условная вероятность последовательности s', а o - эмиссии.

Вероятность конкретной последовательности меток s' рассчитывается следующим образом:

$$P(s|o) = \frac{exp(\phi(o,s))}{\sum_{s'} exp(\phi(o,s'))}, \quad (4)$$

где

$$\phi(o,s) = \sum_{i=1}^{N}(A_{s_{i-1}.s_i} + o_{i,s_i}). \quad (5)$$

Здесь A - матрица переходов.

Гиперпараметры предлагаемой модели были определены эмпирически (Таблица I).

ТАБЛИЦА I. ГИПЕРПАРАМЕТРЫ ПРЕДЛОЖЕННОЙ МОДЕЛИ

Гиперпараметр	**Значение**
Размер батча	32
Оптимизатор	Adam
Число эпох	30

Подводя итоги можно сказать, что предлагаемый метод, основанный на XLNet, CNN и CRF изучает текстовую информацию из различных источников о киберугрозах. Данный подход позволяет предотвращать и обнаруживать потенциальные угрозы кибербезопасности.

III. ЭКСПЕРИМЕНТАЛЬНЫЕ НАБОРЫ ДАННЫХ

Набор данных DNRTI (Dataset for Named Entity Recognition in Threat Intelligence) был собран Wang и соавт. [11] и включает отчеты об угрозах с веб-сайтов различных компаний, занимающихся вопросами безопасности, государственных учреждений и т. д. Он состоит из 175220 токенов, 6587 предложений и 13 типов сущностей, помеченных в соответствии со схемой разметки границ слов BIO: Tool, Time, HackOrg, SecTeam, Purp, Area, Idus, OffAct, SamFile, Org, Way, Features и Exp. Они включают хакерские организации, наступательные действия, цели, группы безопасности, вредоносные файлы и другие.

Набор данных APTNER (Advanced Persistent Threat NER) включает 260134 токена и 21 тип сущностей. Он включает отчеты, полученные от различных компаний, занимающихся сетевой безопасностью [12]. Набор данных включает такие сущности, как URL-адреса, домены, хеш-значения, субъекты угроз, группы безопасности и т. д. Тип сущности был определен в соответствии со стандартом STIX 2.1. Набор данных был аннотирован вручную и проверен экспертами.

IV. РЕЗУЛЬТАТЫ ЭКСПЕРИМЕНТОВ

Эксперименты проводились на Python 3.10.12 с использованием различных библиотек, включая TensorFlow и Keras. Использовалась машина Intel Xeon (R) CPU X5670 @ 2.93GHz * 24 с 24 ГБ оперативной памяти.

Эксперименты проводились на двух наборах данных: DNRTI [11] и APTNER [12]. Набор данных DNRTI включает три набора данных: обучающий (5261 предложение), валидационный (662 предложения) и тестовый (664 предложения). Для проведения экспериментов на наборе данных APTNER было рассмотрено 6645 (обучающий набор данных), 1664 (проверочный набор данных) и 1528 (тестовый набор данных) образцов.

В таблице II показана производительность предложенного подхода для различных именованных сущностей в сфере кибербезопасности на основе метрик precision (P), recall (R) и F-measure (F1) [13]

ТАБЛИЦА II. ОЦЕНКА ПРОИЗВОДИТЕЛЬНОСТИ ПРЕДЛОЖЕННОГО МЕТОДА НА НАБОРЕ ДАННЫХ DNRTI

Тип именованной сущности	**UTERMMF [16]**			**base+BERT+HSA [17]**			**QA+DCR-CharNet+ SecRoBERTa [18]**			**Предлагаемый подход**		
	P (%)	R (%)	F1 (%)	P (%)	R (%)	F1 (%)	P (%)	R (%)	F1 (%)	P (%)	R (%)	F1 (%)
HackOrg	88.86	84.51	86.63	82.92	81.79	82.35	91.33	89.52	90.41	**92.73**	**96.92**	**94.78**
Idus	**93.75**	**93.02**	**93.39**	91.04	94.57	92.78	90.08	95.16	92.55	79.46	85.54	82.39
Tool	89.85	76.60	82.70	78.85	70.51	74.45	82.33	82.89	82.61	**99.82**	**99.19**	**99.50**
Area	88.34	**91.63**	**89.95**	86.26	84.26	85.26	**90.05**	89.16	89.60	89.32	90.56	89.94
Org	80.92	77.37	79.10	72.06	71.53	71.79	**87.72**	**79.37**	**83.33**	75.81	74.60	75.20
Time	91.76	**92.31**	92.04	87.65	88.17	87.91	**94.27**	91.93	**93.08**	85.48	77.94	81.54
Way	95.10	98.98	97.00	84.21	97.96	90.57	86.46	98.81	92.22	**96.09**	**100**	**98.04**
OffAct	92.91	78.67	85.20	85.94	73.33	79.14	92.91	85.51	89.06	**96.45**	**90.33**	**93.29**
SamFile	90.16	**92.34**	**91.24**	**96.38**	85.89	90.83	94.35	85.20	89.54	92.83	84.37	88.40
Features	**96.67**	**100**	**98.31**	91.27	99.14	95.04	91.43	**100**	95.52	90.94	84.27	87.48
SecTeam	**93.42**	**93.42**	**93.42**	88.36	84.87	86.58	91.67	90.41	91.03	74.21	75.00	74.60
Exp	98.51	**100**	**99.25**	98.51	**100**	**99.25**	97.53	**100**	98.75	**100**	97.83	98.90
Purp	89.84	**100**	94.65	85.82	**100**	92.37	86.46	98.81	92.22	**91.20**	**100**	**95.40**

Из таблицы видно, что предложенный подход наиболее точно распознает классы Tool, Exp, Purp, OffAct, HackOrg и Way в соответствии с рассматриваемыми метриками. Это показывает, что, хотя XLNet-CNN-CRF эффективно распознает большинство именованных сущностей, предложенный подход также получил некоторые ложноположительные результаты.

В таблице III представлены экспериментальные результаты для набора данных APTNER. Сравнение с моделями BERT-BiLSTM-CRF [12], Dyn-Att Net [14] и DarkBERT [15] продемонстрировало превосходство предложенной модели XLNet-CNN-CRF по показателям precision, recall и F-measure.

Таким образом, приведенные выше экспериментальные результаты подтверждают эффективность и применимость предложенного подхода XLNet-CNN-CRF для распознавания и классификации именованных сущностей, полученных из источников о киберугрозах.

ТАБЛИЦА III. ОЦЕНКА ПРОИЗВОДИТЕЛЬНОСТИ ПРЕДЛОЖЕННОГО МЕТОДА НА НАБОРЕ ДАННЫХ APTNER

Метрики / Метод	F-measure (%)	Precision (%)	Recall (%)
BERT-BiLSTM-CRF [12]	82.31	82.99	81.64
Dyn-Att Net [14]	90.33	89.79	90.88
DarkBERT [15]	86.33	84.72	88.00
Предложенный подход	**99.39**	**99.75**	**99.05**

ЗАКЛЮЧЕНИЕ

Подходы на основе обработки естественного языка позволяют анализировать обширную текстовую информацию, полученную из отчетов CTI. NLP оперативно идентифицирует все виды угроз, сокращая время реагирования и вероятность ложных срабатываний.

В данной работе предложен подход к распознаванию именованных сущностей кибербезопасности на основе XLNet, CNN и CRF. Эксперименты проводились на больших текстовых наборах данных, связанных с CTI: DNRTI и APTNER. Метод XLNet-CNN-CRF был сравнен с различными методами обработки естественного языка, такими как XLNet, DarkBERT, BiLSTM, TiKG и др. Результаты показали, что предложенный подход может эффективно обнаруживать именованные сущности, связанные с киберугрозами, для повышения уровня защиты и автоматизации реагирования на киберинциденты и киберугрозы. Применение предложенного подхода к автоматическому анализу неструктурированных текстовых данных позволит экспертам по кибербезопасности быть в курсе новых уязвимостей, кибератак и возможных угроз.

БЛАГОДАРНОСТЬ

Данная работа была выполнена при финансовой поддержке Азербайджанского технического университета **Грант № AzTU-DQL-2025-M01/60182631**.

СПИСОК ЛИТЕРАТУРЫ

[1] S. Srivastava, B. Paul, and D. Gupta, "Study of word embeddings for enhanced cyber security named entity recognition," Procedia Computer Science, vol. 218, pp. 449-460, 2023.

[2] M. Arazzi, D. R. Arikkat, S. Nicolazzo, A. Nocera, K. A. Rafidha Rehiman, P. Vinod, and M. Conti, "NLP-based techniques for cyber threat intelligence," 2023, arXiv preprint arXiv:2311.08807

[3] Z. Liu, "A review of advancements and applications of pre-trained language models in cybersecurity," In Proc. of 12th International Symposium on Digital Forensics and Security (ISDFS), 2024, pp. 1-10.

[4] Q. Lyu, M. Apidianaki, and C. Callison-Burch, "Towards faithful model explanation in NLP: A survey," Computational Linguistics, vol. 50, no. 2, pp. 657-723, 2024.

[5] Y. Zhou, Y. Tang, M. Yi, C. Xi, and H. Lu, "CTI View: APT Threat Intelligence Analysis System," Security and Communication Networks, vol. 2022, no. 1, 9875199, 2022. https://doi.org/10.1155/2022/9875199

[6] P. Cetin, and O. O. Tanriover, "Priority rule for resource constrained project planning problem with predetermined work package durations," Journal of the Faculty of Engineering and Architecture of Gazi University, vol. 35, no. 3, pp. 1537-1549, 2020.

[7] C. Hanks, M. Maiden, P. Ranade, T. Finin, and A. Joshi, "Recognizing and Extracting Cybersecurtity-relevant Entities from Text," 2022, arXiv preprint arXiv:2208.01693

[8] Z. Gao, Z. Li, J. Luo, and X. Li, "Short-text aspect-based sentiment analysis based on CNN+BiGRU," Applied Sciences, vol. 12, no. 5, pp. 1-17, 2022.

[9] P. C. Aravind, D. R. Arikkat, A. S. Krishnan, B. Tesneem, A. Sebastian, M. J. Dev, K. R. Aswathy, K. A. Rafidha Rehiman, and P. Vinod, "CyTIE: Cyber Threat Intelligence extraction with named entity recognition," In ASCIS 2023. Advancements in Smart Computing and Information Security. Communications in Computer and Information Science, vol. 2039, S. Rajagopal, K. Popat, D. Meva, and S. Bajeja, Eds., Cham: Springer, 2024, pp. 163-178.

[10] Z. Yang, Z. Dai, Y. Yang, J. Carbonell, R. R. Salakhutdinov, and Q. V. Le, "XLNet: Generalized autoregressive pretraining for language understanding," In Proc. of 33rd International Conference on Neural Information Processing Systems, 2019, pp. 5753 - 5763.

[11] X. Wang, X. Liu, S. Ao, N. Li, Z. Jiang, Z. Xu, Z. Xiong, M. Xiong, and X. Zhang, "DNRTI: A large-scale dataset for named entity recognition in threat intelligence," In Proc. of IEEE 19th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom), 2020, pp. 1842-1848.

[12] X. Wang, S. He, Z. Xiong, X. Wei, Z. Jiang, S. Chen, and J. Jiang, "APTNER: a specific dataset for NER missions in cyber threat intelligence field," In Proc. of IEEE 25th International Conference on Computer Supported Cooperative Work in Design (CSCWD), 2022, pp. 1233–1238.

[13] C. Ferri, J. Hernández-Orallo, and R. Modroiu, "An experimental comparison of performance measures for classification," Pattern Recognition Letters, vol. 30, no. 1, pp. 27-38, 2009.

[14] J. Hou, S. Saad, and N. Omar, "Enhancing traditional Chinese medical named entity recognition with Dyn-Att Net: A dynamic attention approach," PeerJ Computer Science, vol. 10, no. e2022, pp. 1-25, 2024.

[15] T. Paladini, L. Ferro, M. Polino, S. Zanero, and M. Carminati, "You might have known it earlier: Analyzing the role of underground forums in threat intelligence," In Proc. of the 27th International Symposium on Research in Attacks, Intrusions and Defenses, 2024, pp. 368-383.

A Cyberthreat Detection Method Using Natural Language Processing and an Improved XLNet Model

Ramiz Aliguliyev[1], Lyudmila Sukhostat[2]
[1,2]Institute of Information Technology, Baku, Azerbaijan
[1]r.aliguliyev@gmail.com, [2]lsuhostat@hotmail.com

Abstract— The rise of cyberthreats worldwide requires their timely detection to improve cybersecurity. Natural language processing methods are widely used to detect cyberthreats. Named entity recognition allows for more accurate identification of various cyber vulnerabilities. However, analysing text data remains a complex task. To address these issues, this paper proposes a cyberthreat detection method using natural language processing and an improved XLNet model. Experiments are conducted on two datasets: DNRTI and APTNER. Comparisons with other well-known methods demonstrate the effectiveness of this method and contribute to reliable, robust cyberthreat detection, addressing the limitations of the considered methods.

Keywords- cyber threat intelligence; cybersecurity; natural language processing; deep learning.