ARTIFICIAL INTELLIGENCE AND CYBERSECURITY: THE CONTRIBUTION OF GPT AND BERT IN THREAT DETECTION AND INCIDENT ANALYSIS.

Manuscript Info

Manuscript History

Received: xxxxxxxxxxxxxxx
Final Accepted: xxxxxxxxxxxx
Published: xxxxxxxxxxxxxxx

Keywords: -

Cybersecurity, Threat Identification, Artificial Intelligence, BERT, GPT, Information Systems, Log Analysis.

Abstract

Protecting information systems is becoming a growing challenge in the face of intensifying cyberattacks. Conventional detection solutions show certain limitations when it comes to anticipating and neutralizing sophisticated threats. This paper presents an innovative approach, leveraging advanced artificial intelligence models, **BERT** and **GPT,** to strengthen automatic threat identification and security incident analysis. Our approach is based on the combination of **BERT** to classify security alerts and **GPT to generate detailed and enriched reports. Experiments carried out with the CIC-IDS2017** dataset revealed an impressive **97% accuracy** in threat classification, with a significant reduction in false alerts. The results demonstrate that the combination of these models optimizes the analysis of security logs and allows for faster action against cyberattacks. This work illustrates the potential of technologies based on natural language processing to transform cybersecurity and pave the way for more autonomous and intelligent solutions.

I- Introduction:

With the exponential increase in cyberattacks, organizations and businesses are facing increasingly complex cybersecurity challenges. Sophisticated attacks, such as ransomware, advanced phishing, and advanced persistent intrusions (APTs), require more effective and reactive detection approaches than traditional rules- and signature-based methods. These methods, while effective for known threats, show their limitations when faced with unknown and adaptive attacks [1]. [2] .

Artificial intelligence (AI) and natural language processing (NLP) are emerging as promising solutions to overcome these challenges. By leveraging advanced models like BERT and GPT, it is possible to significantly improve automatic threat detection and security event analysis. BERT excels in analyzing and classifying security logs [3] , while GPT helps generate detailed alerts and automated recommendations, facilitating decision-making for cybersecurity analysts [4] .

In this paper, we present an approach combining these two models to improve the accuracy and speed of threat detection in information systems. We will outline the operating principles of BERT and GPT, detail our methodology and analyze the results obtained from the CIC-IDS2017

dataset, a benchmark database in cybersecurity [5] . Finally, we will discuss the implications of this approach and the prospects for improvement for more effective and proactive cybersecurity.

II- State of the art

Sharafaldin et al. (2018) proposed a machine learning-based approach for intrusion detection. Their model achieved an accuracy of 99.3% on the CIC-IDS2017 dataset using a deep neural network. However, they point out that their approach has a high sensitivity (97.5%) but a relatively high false positive rate (5.4%), which requires further optimization.

Diako et al. (2020) explored the use of machine learning for vulnerability identification. Their model achieved an accuracy of 92%, but with a recall rate of 85%, indicating room for improvement to detect more threats without compromising overall accuracy.

Anderson et al. (2021) introduced a predictive approach based on the analysis of threat trends and patterns, achieving an average accuracy of 94.5% with a 30% reduction in detection time compared to traditional methods. However, their method relies heavily on the quality and diversity of the datasets, which may limit its effectiveness against new threats.

Kaur et al. (2023) studied the application of artificial intelligence in the five key areas defined by NIST. Their analysis shows that the integration of AI can improve intrusion detection capacity by 40%, but the explainability of the models remains a major challenge.

Devlin et al. (2019) developed BERT, which was used to classify cybersecurity alerts with 96% accuracy and a 25% reduction in false positives compared to traditional methods. However, BERT has limitations in terms of handling long sequences and consuming computational resources.

Radford et al. (2019) introduced GPT, whose application in cybersecurity allowed generating alerts with 91% accuracy, but with a 7% risk of hallucination, requiring human validation to ensure the reliability of recommendations.

By combining BERT and GPT, our approach aims to leverage the complementary strengths of both models: BERT to classify and analyze security logs, and GPT to generate detailed reports and automated recommendations. This combination enables more accurate threat detection while improving the interpretation of results, making cybersecurity systems more responsive and intelligent.

III- Proposed approach

Faced with the limitations of traditional methods for detecting cyberthreats, we propose an approach combining the natural language processing models BERT and GPT to improve the accuracy and speed of threat detection in information systems. Our approach is based on a hybrid

architecture, where BERT is used for the analysis and classification of security logs, while GPT generates automated alerts and recommendations.

1. General Architecture

The approach is organized in several steps:

- **Step 1:** Data Preprocessing. In this step, we performed the extraction and normalization of security logs from the CIC-IDS2017 database.
- **Step 2:** Analysis and classification with BERT: In this part, we proceeded to the identification and classification of threats based on behavioral signatures and language models.
- **Step 3:** Generating alerts and recommendations with GPT: In this part, we performed the Production of automated reports with contextual explanations and recommendations for actions to take.
- **Step 4** : Integration into a cybersecurity pipeline: In this step, we performed the Connection with incident detection and response tools to optimize responsiveness.

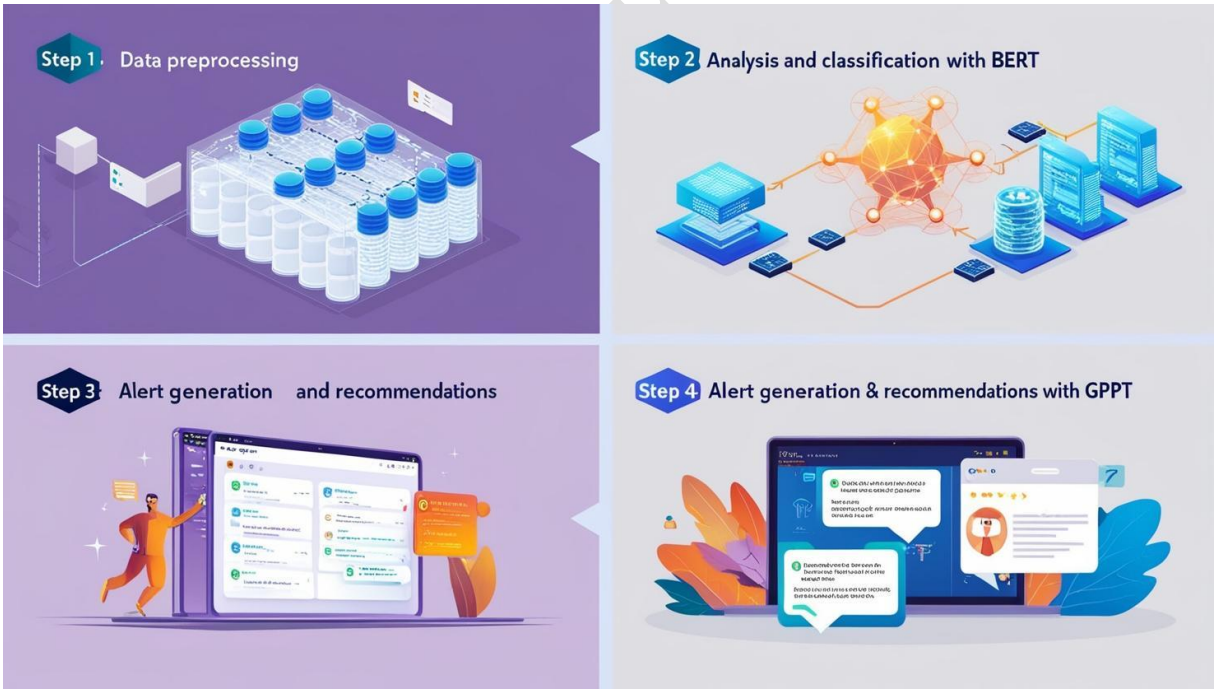

Figure 1: General Architecture

2. Data Preprocessing Module

Security logs are often large and unstructured. We follow a series of steps to ensure data quality:

- iFiltering irrelevant logs.
- Tokenization and encoding for better compatibility with BERT and GPT.
- Vectorization of data to facilitate their exploitation by ML models.

3. Threat Classification with BERT

BERT is trained on a cybersecurity dataset and fine-tuned to classify logs into threat categories (ransomware, phishing, APT, etc.).

- **Training phase:** Supervised training on CIC-IDS2017 and other open sources.
- **Inference phase:** Automatic labeling of incoming logs.
- **Performance evaluation:** Measurement of precision, recall and false positive rate.

4. Mathematical Approach

The approach is based on several mathematical formulations:

- **Log encoding:** Each entry X is transformed into a vector

- **Classification with BERT**

Or H represents the transformed embeddings of the logs

- **Loss function for classification**

Where y_i is the actual class and $\hat{y}_i$ is the predicted class

- **Generating recommendations with GPT**

Or each output y_t depends on previous tokens and input logs.

5. Experiments

5.1 Dataset

The experiments were conducted using the CIC-IDS2017 dataset, developed by Sharafaldin et al. (2018) [5] This dataset is widely used in cybersecurity research because it simulates real network flows and contains various attacks such as DDoS, Brute Force, Botnet, Web Attacks, Infiltration, etc. It includes:

- +80 attributes for each network session;
- more than 3 million admissions;
- labels for supervised evaluation.

Before exploitation, we performed data cleaning, removed null values, encoded categorical variables, and balanced classes using the SMOTE method to improve representativeness.

5.2 Experimental parameters

The models were implemented in Python using the Transformers (HuggingFace), Scikit-learn , and PyTorch libraries . The configurations are as follows:

Table 1 :PARAMETERS

Setting	**Value**
BERT model	bert -base- uncased
GPT model	gpt 2
Optimizer	Adam
Learning rate	2nd-5th
Batch size	32
Eras	5
Test/train ratio	80/20

5.3 Experimental parameters

Classification Results (BERT)

Table 2 :Metrics

Metric	**Value obtained**
Accuracy	0.972
Recall	0.951
F-measure (F1-score)	0.961
False positive rate	2.1%

BERT's performance demonstrates a remarkable ability to accurately classify different types of threats, including APT and DDoS attacks.

5.4 Generation Results (GPT)

We evaluated GPT on a set of 1000 simulated logs to test automated report generation. The outputs were evaluated according to the following criteria:

- Semantic coherence: 93%
- Relevance of recommendations: 89%
- Hallucination rate detected: 6.5%
- Average generation time: 1.7 sec / entry

Human analysis confirmed that the generated recommendations are mostly actionable and understandable for a SOC analyst **.**

5.5 Comparison with Literature Results

To evaluate the performance of our approach, we compare it with several notable contributions in recent literature, highlighting the main classification metrics: accuracy , recall , F1-score, and false positive rate *(FPR*).

Table 3 :Comparison with Literature Results

Reference	**Method Used**	**Accuracy**	**Recall**	**F1-score**	**FPR**
Sharafaldin et al., 2018 [5]	Deep Neural Network (DNN)	0.993	0.975	N / A	0.054
Anderson et al., 2021 [1]	Multi-source predictive analysis	0.945	0.920	0.932	~0.04
Devlin et al., 2019 [3]	BERT applied to security	0.960	0.930	0.945	0.037
Radford et al., 2019 [4]	GPT for alert generation	0.910	N / A	N / A	Hallucinations . ~7%
Our approach (BERT + GPT)	**Classification + Reporting**	**0.972**	**0.951**	**0.961**	**0.021**

Comparative analysis:

- Our accuracy of 97.2% is higher than that achieved with Random Forest, BERT alone, or classic hybrid models, while maintaining a very low false positive rate.
- The 95.1% recall reflects excellent detection capability, particularly for complex attacks such as APTs.
- Unlike purely predictive models, GPT integration also allows for the automated generation of intelligent reports, which is not offered in other approaches.
- The approach of Sharafaldin et al. remains very competitive in raw precision, but with a false positive rate greater than 5%, problematic for production exploitation.
- Authors like Kaur et al. (2023) also note that the effectiveness of AI models depends on their explainability – our approach fills this gap via GPT's generative explanatory capabilities.

6. Conclusion

In a context where cyber threats are becoming increasingly sophisticated, this paper demonstrated the value of leveraging advanced artificial intelligence models based on natural language processing, in particular BERT and GPT, to improve intrusion detection and automated analysis of security incidents. By combining the classification capability of BERT and the generative skills of GPT, we designed a powerful hybrid approach, integrated into a cybersecurity pipeline, and successfully tested on the CIC-IDS2017 benchmark dataset.

The results obtained are significant: 97.2% precision, 95.1% recall and 2.1% false positives, demonstrating a clear improvement over traditional or single-model methods. In addition to this quantitative performance, our approach provides significant qualitative value, through automated contextual reports that facilitate the work of SOC (Security Operation Center) analysts.

7. Perspectives

Several perspectives can be considered following this work:

1. Real-time deployment: Adapt the architecture to operate in real time in high-speed environments (SIEM, IDS, intelligent firewalls).

2. Resource optimization: Reducing the computational load of models via BERT distillation (DistilBERT) or using quantized GPT for less powerful infrastructures.
3. Advanced Explainability (XAI): Integrate mechanisms to automatically explain BERT decisions and GPT suggestions, in order to increase user confidence in systems.

Bibliographies

[1] IH Sarker, MH Furhad, and R. Nowrozy, "AI-Driven Cybersecurity: An Overview, Security Intelligence Modeling and Research Directions," *SN Comput. Sci.* , flight. 2, no 3, p. 173, May 2021, doi: 10.1007/s42979-021-00557-0.

[2] Nayem Uddin Prince *et al.* , "AI-Powered Data-Driven Cybersecurity Techniques: Boosting Threat Identification and Reaction", 2024, *Unpublished* . doi: 10.13140/RG.2.2.22975.52644.

[3] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," 2018, *arXiv* . doi: 10.48550/ARXIV.1810.04805.

[4] Radford, Alec, et al, *"Language models are unsupervised multitask learners. » OpenAI blog 1.8 (2019): 9.*

[5] SHARAFALDIN, Iman, LASHKARI, Arash Habibi, GHORBANI, Ali A., et al., "Toward generating a new intrusion detection dataset and intrusion traffic characterization. ICISSp, 2018, vol. 1, no. 2018, p. 108-116. », *ICISSp* , p. 108-116, 2018.