Spectral Analysis for Attack Detection

Majed Jaber, Nicolas Boutry, Pierre Parrend

Spectral Analysis for Attack Detection

1st Majed Jaber
ICube - Laboratoire
des sciences de l'ingénieur,
de l'informatique et de l'imagerie
UMR 7357, Université de Strasbourg
CNRS, 67000, Strasbourg, France
Laboratoire de Recherche
de L'EPITA (LRE), 14-16 rue Voltaire
94270 Le Kremlin-Bicêtre, France.
majed.jaber@epita.fr

2rd Nicolas Boutry
Laboratoire de Recherche
de L'EPITA (LRE), 14-16 rue Voltaire
94270 Le Kremlin-Bicêtre, France.
nicolas.boutry@epita.fr

3nd Pierre Parrend
ICube - Laboratoire
des sciences de l'ingénieur,
de l'informatique et de l'imagerie
UMR 7357, Université de Strasbourg
CNRS, 67000, Strasbourg, France
Laboratoire de Recherche
de L'EPITA (LRE), 14-16 rue Voltaire
94270 Le Kremlin-Bicêtre, France.
pierre.parrend@epita.fr

***Abstract*— Medical IoT networks face significant cyber threats that compromise system accessibility, patient confidentiality, and data communication security. Our study introduces a novel detection method using spectral graph analysis. This mathematical technique, based on the Laplacian matrix's spectral properties, provides insights into network topology changes by analyzing networks as dynamic graphs over time. This method enables us to track spectral variations over time, enabling the early detection of cybersecurity threats. The spectral analysis shows the detection of the attacks over the Bot-IoT and Ton-IoT datasets, that consist of both benign and simulated malicious network traffic.**

***Index Terms*—Cybersecurity, Spectrum, Spectral graph analysis, Laplacian Matrix, XGBoost.**

I. Introduction

This paper proposes a spectral time windowing approach called *SpectraTW* for attack detection and evaluates it in the context of medical IoT network traffic. By modeling network interactions through a graph model, it enables the identification of unusual behaviours. The preliminary results of this research assess the effectiveness of graph spectral analysis by showing its impact on the performance of machine learning (ML) algorithms such as XGBoost. This evaluation offers initial insights into its applicability in real-world scenarios.

II. State-of-the-Art

This section discusses the advancement in cybersecurity anomaly detection, transitioning from traditional methods to advanced machine learning and deep learning. In real-time detection, systems like ReTiNA [1] stand out in statistical methods, whereas the CAMLPAD [2] framework is designed for real-time cybersecurity data collection and uses machine learning to detect anomalies and perform scoring. Network security faces challenges with limited labeled data for training, leading to the adoption of graph-based ML. Techniques like walk-based sampling [3] transform unstructured graph data into structured forms. Time series graph learning [4] are utilized for identifying complex network patterns. Additionally, deep learning significantly influences graph data analysis, like GCNs [5] becoming increasingly important. Spectral graph analysis accurately identifies complex threats by dynamically monitoring networks and using spectral metrics for network behavior analysis.

III. Spectral Metrics

This section explains the metrics we introduced for detecting attacks over the network. These metrics are derived from the eigenvalues (Λ_t) of the Laplacian matrix. We denote by $\Lambda_t[i]$ the i^{th} eigenvalue, $i \in [1,n]$, sorted in increasing order, and by $\mathscr{Z}(t)$ the multiplicity of zero in Λ_t. **Connectedness**, equal to $\exp(1/\mathscr{Z}(t)-1)$, measures interconnectivity in the network. Let $\mathscr{N}$ be the number of network devices (e.g. switches and servers). **Flooding**, equal to $(\frac{1}{\mathscr{N}}\sum_{i=\mathscr{Z}(t)+1}^{\mathscr{Z}(t)+\mathscr{N}}\Lambda_t[i]) - 1$ and **Wiriness**, equal to $\frac{1}{\mathscr{N}}\sum_{i=n-\mathscr{N}+1}^{n}\Lambda_t[i]$, analyze edge weights and connections, focusing on the central and highest eigenvalues respectively. **Asymmetry**, equal to $Card\{i \geq 2;\ \Lambda_t[i]-\Lambda_t[i-1] > 10^{-12}\}$, tracks spectrum evolution.

IV. Experiments and observations

Table I
Feature selection for each approach

	COD	CTS	CTW	SpectraTW
Packet Features	x	x	x	x
Rate Features	x	x	x	x
Byte Features	x	x	x	x
Other Features	x	x		
Connectedness				x
Flooding				x
Wiriness				x
Asymmetry				x

Our study utilizes the Botnet [6] and TonIoT [7] datasets. We have opted for **XGBoost** as the classification tool [8]. We investigate various anomaly detection strategies with the features shown in Table I, beginning with **Classification on Original Data-logs (COD)**, which involves the examination of raw data. Followed by **Classification on Time-series (CTS)**, which concentrates on transforming data into time-series format. Additionally, we explore **Classification on Time-window**

Table II
METRICS EVALUATION OVER TON-IOT DATASET

Ton-IoT Dataset		COD	CTS	CTW	SpectraTW
DDoS	F1 Score	0.9764	0.6175	0.9943	**1**
	Balanced Acc	0.9833	0.7628	0.9971	**1**
	MCC	0.9752	0.6256	0.9943	**1**
DoS	F1 Score	0.9837	0.9054	0.9957	**1**
	Balanced Acc	0.9902	0.9403	0.9964	**1**
	MCC	**1**	0.8816	**1**	**1**
Scanning	F1 Score	0.9890	0.3363	0.9938	**1**
	Balanced Acc	0.9959	0.6102	0.9995	**1**
	MCC	0.9884	0.3631	0.9933	**1**
Ransomware	F1 Score	0.8290	0.2978	0.9243	**0.9949**
	Balanced Acc	0.9131	0.5973	0.9526	**0.9949**
	MCC	0.8193	0.3450	0.9240	**0.9948**
SQL Injection	F1 Score	0.9735	0.8445	0.8428	**0.9972**
	Balanced Acc	0.9808	0.9040	0.8732	**0.9999**
	MCC	0.9721	0.8433	0.8480	**0.9972**
Password	F1 Score	0.9808	0.7304	0.9148	**0.9939**
	Balanced Acc	0.9882	0.9522	0.9748	**0.9966**
	MCC	0.9798	0.7363	0.9125	**0.9937**
XSS	F1 Score	0.8710	0.6731	**1**	**1**
	Balanced Acc	0.9509	0.8043	**1**	**1**
	MCC	0.8645	0.6753	**1**	**1**
Backdoor	F1 Score	0.9985	0.8589	0.9928	**0.9995**
	Balanced Acc	0.9989	0.8928	0.9967	**0.9995**
	MCC	0.9984	0.8563	0.9923	**0.9995**
MitM	F1 Score	0.7239	0.4542	0.7640	**1**
	Balanced Acc	0.8733	0.6747	0.8148	**1**
	MCC	0.7235	0.4742	0.7818	**1**

(CTW), utilizing a sliding window approach for more intricate data scrutiny. Finally, **Classification on Spectral-Time-Window (SpectraTW)** is employed for a comprehensive spectral analysis of the network's structure. All strategies entails the dataset features, whereas SpectraTW possesses our spectral metrics features in addition. The analysis of the datasets shown in Table II and III, with the XGBoost classifier reveals significant insights. The SpectraTW approach consistently outperforms other methods, particularly in categories like OS-Fingerprint and Keylogging. It achieves high performance, indicating its ability in feature selection and combination.

Table III
METRICS EVALUATION OVER BOTNET-IOT DATASET

Botnet-IoT Dataset		COD	CTS	CTW	SpectraTW
DDoS	F1 Score	**1**	0.8750	**1**	**1**
	Balanced Acc	**1**	0.8888	**1**	**1**
	MCC	**1**	0.8816	**1**	**1**
ScanService	F1 Score	0.8937	0.9966	0.9990	**0.9994**
	Balanced Acc	0.9760	0.9133	0.9834	**0.9942**
	MCC	0.8835	0.8965	0.9797	**0.9885**
OS Fingerprint	F1 Score	0.2617	0.8235	**0.9953**	**0.9953**
	Balanced Acc	0.5804	0.8798	**0.9953**	**0.9953**
	MCC	0.3198	0.8240	**0.9952**	**0.9952**
Keylogging	F1 Score	0.5333	0.6666	**1**	**1**
	Balanced Acc	0.7856	0.7998	**1**	**1**
	MCC	0.5344	0.6703	**1**	**1**

It results in more accurate anomaly detection and correlation between predictions and actual classifications. In contrast, other approaches show varied effectiveness, with SpectraTW emerging as the most efficient method for enhancing classifier performance in complex cyber threat scenarios.

V. ATTACKS BEHAVIORS

In the datasets results shown in Tables II and III, attack distributions fall into categories of exploitation, disruption, scanning, theft, and reconnaissance. This section provides a description of port scanning, OS fingerprinting, and keylogging attacks and explains the role of spectral analysis in identifying these attacks. Port scanning is used to identify open network ports, while OS fingerprinting seeks to gather data on the operating system and its communication protocols. These types of attacks are marked by persistent port scanning activities, noticeable by increased connection attempts and a surge in the quantity and frequency of data packets sent to the target. SpectraTW uses spectral metrics features, relying on communication data that affects its spectral characteristics and on packets count, rate, and bytes per session, over single or multiple predetermined time frames. SpectraTW effectively handles several time windows, corresponding to various states and ports within the same time-frame. They analyze multiple graphs, created based on data like packets, rates, and bytes, which is essential for detecting theft attacks.

VI. ATTACK MODELS

Attackers could evade spectral detection by adopting strategies that disrupt the typical patterns. These strategies might include varying the rate and distribution of scanning activities to avoid creating consistent patterns in the graph topology. The attacker could also vary the attack intensity over time, making the eigenvalue patterns less pronounced and more challenging to detect.

VII. CONCLUSION AND FUTURE WORK

Our proposed approach highlights spectral analysis as an effective tool for detecting cyber threats in medical IoT, and points out XGBoost's successful use in anomaly detection, with SpectraTW proving efficient in most attack scenarios. It achieves 100% detection rates for certain threats, such as MitM attacks, indicating significant improvement. For more complex attacks like Ransomware, it attains over 99% effectiveness. Future work will focus on advanced, sequential attack strategies, which are challenging to detect using conventional methods.

REFERENCES

[1] J. Noble and N. Adams, "Real-time dynamic network anomaly detection," *IEEE Intelligent Systems*, 2018.

[2] A. Hariharan, A. Gupta, and T. Pal, "CAMLPAD: Cybersecurity autonomous machine learning platform for anomaly detection," in *FICC*, 2020.

[3] B. Perozzi, R. Al-Rfou, and S. Skiena, "Deepwalk: Online learning of social representations," in *SIGKDD*, 2014.

[4] A. Pareja, G. Domeniconi, J. Chen, T. Ma, T. Suzumura, H. Kanezashi, T. Kaler, T. Schardl, and C. Leiserson, "Evolving graph convolutional networks for dynamic graphs."

[5] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," 2016.

[6] N. Koroniotis, N. Moustafa, E. Sitnikova, and B. Turnbull, "Towards the development of realistic botnet dataset in the internet of things for network forensic analytics: Bot-IoT dataset," *Future Generation Computer Systems*.

[7] T. M. Booij, I. Chiscop, E. Meeuwissen, N. Moustafa, and F. T. Den Hartog, "ToN_IoT: The role of heterogeneity and the need for standardization of features and attack types in IoT network intrusion data sets," 2021.

[8] L. Grinsztajn, E. Oyallon, and G. Varoquaux, "Why do tree-based models still outperform deep learning on typical tabular data?" *Advances in Neural Information Processing Systems*, vol. 35, pp. 507–520, 2022.