Structural and spectral analysis of dynamic graphs for attack detection

Majed Jaber[1,2], Nicolas Boutry[2], Pierre Parrend[1,2]
[1] ICube - Laboratoire, des sciences de l'ingénieur, de l'informatique et de l'imagerie, UMR 7357, Université de Strasbourg, CNRS, 67000, Strasbourg, France
[2] Laboratoire de Recherche, de L'EPITA (LRE), 14-16 rue Voltaire, 94270 Le Kremlin-Bicêtre, France

May 5, 2023

Abstract

At this time, cyberattacks represent a constant threat. Many approaches exist for detecting suspicious behaviors, but very few of them seem to benefit from the huge potential of mathematical approaches like spectral graph analysis, known to be able to extract topological features of a graph using its Laplacian spectrum. For this reason, we consider our network as a dynamic graph composed of nodes (representing the devices) and of edges (representing the requests), and we compute its Laplacian spectrum across time. An important change of topology inducing an important change in the spectrum, this spectrum seems to be the key to detect threats. Dynamic spectrum-based metrics have been developed for this aim.

Keywords

Cybersecurity, anomaly detection, spectral graph analysis, Laplacian spectrum, graph topology.

1 Introduction

The frequency of cyberattacks has been on the rise over the years, with an increasing number of threat actors from nation-states and hacker groups that are getting involved in such activities. The shortage of skilled personal needed to counter these threats grew at the same time. Skill shortage lead to an increase need for automation of cybersecurity analysis, and thus more expressive and powerful models for the detection of cyberattacks. To efficiently ease cybersecurity risks, we need advanced solutions that allow us to relate and analyze connections on a practical scale. Defenders usually depend on lists : alerts and logs from software tools, and thus supporting heterogeneous data sources and formats.

Attackers can find a weakness in the network and exploit it to gain access to more devices on the network. Graph data representation and graph analysis models grow as a promising approach to support analysis, detection and reaction capability, providing a high level of transparency with respect to the origin of alerts, and of explainability to help the security analysis react to identified malicious actions. Graphs nowadays are concepts that have been widely used in various applications especially in cybersecurity domain [1–3] Using graphs, we can cover cybersecurity patterns and detect anomalies and threats on networks [4, 5]. Usually, we model a network as a graph $G = (V, E)$ whose nodes V are the devices and whose edges E are the relations between nodes.

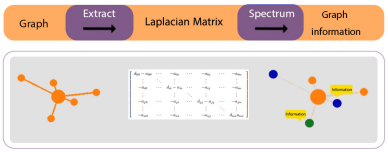

FIGURE 1 – Information extraction using spectral graph analysis

The proposed approach in this paper involves extracting the spectrum of the Laplacian of dynamic graphs and analyzing its evolution for the purpose of detecting cyberattacks (refer to Fig. 1). The goal is to detect topological patterns using spectral graph analysis, to allow us to identify cyberattacks types.

2 State-of-the-Art

Let us present a brief overview of the current state-of-the-Art in matter of anomaly detection and cyberattacks. For *anomaly detection* [6], such approaches are *statistical ones* [7] and ML-based ones [8]. The shortage of labeled data in network security poses a challenge in training classifiers effectively, and the limited existing labeled data may not be applicable to other contexts as noted by [9]. However, graph-based machine learning techniques are expected to have a considerable impact on the development of next-generation cybersecurity systems. One such technique is walk-based sampling, which involves sampling graph-structured data by traversing through the graph using walks. This approach has been investigated in previous studies [10,11] that introduce the *DeepWalk* and *node2vec* methods, respectively. Deep learning has gained significant attention in the field of graph data, with Graph Convolutional Networks [12, 13] (GCN) being the best choice for graph

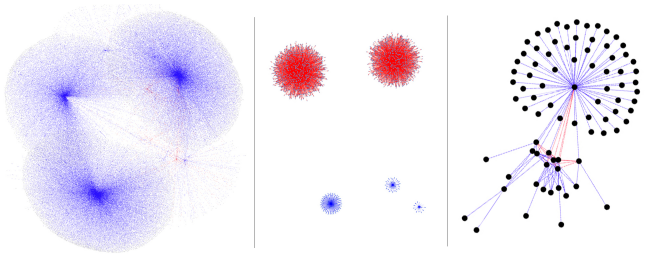

FIGURE 2 – IP-IP graph representation for Ton-IoT, IoT Healthcare Security, and Bot-IoT data sets

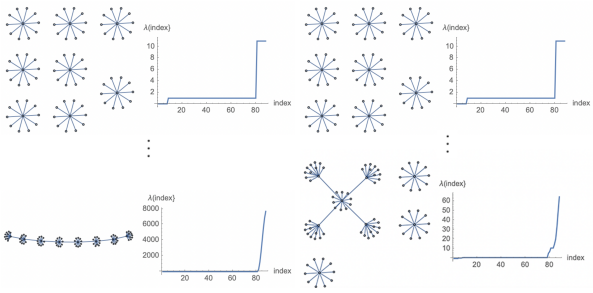

FIGURE 3 – Two different scenarios. On the left side, Scenario 1 : many interconnections depict that a threat is occurring in the network. On the right side, Scenario 2 : a « normal » scenario is depicted with no excessive connections or huge packets.

data learning tasks. Another approach is Graph Attention Networks [14], which utilize the self-attention mechanism to encode hidden representations of each node and attend over its neighbors.

Spectral graph analysis is a useful tool for extracting topological properties from graphs [15]. The Laplacian spectrum can provide features such as the number of connected components, the bipartiteness [16], and the robustness [17] to edge rewiring. Spectral analysis has been applied to cybersecurity, such as in change detection in TCP packet transport [18], in forensic evidence analysis [19], in complexity reduction of graphs [20], in various attacks identification [21], in clustering evolving graphs [22], and in anomaly detection. Techniques like power spectral density [23], diffusion and spectral methods, dictionary learning, and hypothesis testing have been used in these approaches.

3 Proposed metrics

In order to effectively evaluate changes in the spectra resulting from different datasets Fig. 2, it is necessary to consider a range of metrics. By observing the spectrum at various timestamps and tracking the graph's evolution, it becomes possible to identify different factors that can impact the spectrum. Examples of such factors include flooding of packets, node connectivity, and degree of nodes. To facilitate the evaluation of such changes, we propose four metrics.

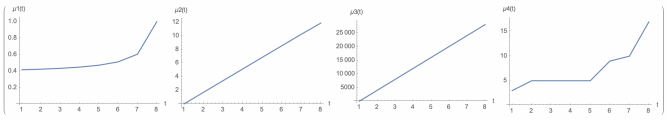

FIGURE 4 – From left to right, the four dynamic metrics μ_1 to μ_4 in Scenario 1.

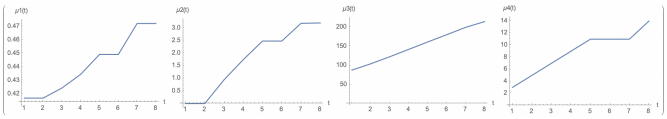

FIGURE 5 – From left to right, the four dynamic metrics μ_1 to μ_4 in Scenario 2.

The first metric, μ_1, called the *connectedness*, is based on the number of connected components in the graph. This metric is useful in evaluating the overall connectivity of the network, and can provide insights into potential areas of weakness or vulnerability.

The second metric, μ_2, called the *flood value*, takes into account both interconnections and connecting edge weights. This metric is designed to quantify the flooding events on the network which can have a significant impact on network performance and security.

The third metric, μ_3, called the *wiringness*, is primarily influenced by the degrees of the nodes of the graph. Nodes with a high degree of connectivity indicate potential areas of interest for further analysis.

Finally, the fourth metric, μ_4, called the *asymmetry*, measures the cardinality of identical patterns in the network. The variation of the number of patterns in the network can be a sign of potential threat.

Thanks to these metrics, it becomes possible to gain valuable insights into potential security threats or other issues impacting network performance.

4 Experiments and observations

Consider a graph G that represents a network of interconnected devices, where the nodes V represent devices and edges E represent the connections between these devices, with weights W corresponding to the amount of data transmitted. In Scenario 1 (see Figure 3), we propose to start with a graph consisting of $\mathcal{N} = 8$ separate connected components and gradually connect the central nodes, as if a *threat* is occurring. It has been observed that most of the traffic information is stored in the first and last $\mathcal{N}$ values of the spectrum. In Scenario 2 (see Figure 3), we start with $\mathcal{N} = 8$ disconnected star graphs and gradually establish connections between non-central client nodes and central server nodes or between two central server nodes. The corresponding edge in the adjacency matrix is assigned a weight of 10 to represent the connection. This scenario is considered *normal*. Let us recall that the bigger the metric, the bigger the risk of a threat. If we observe the dynamic metrics (see Figures 4, 5) , we understand that, compared to a normal case, the threatening case is easily detected.

5 Conclusion and future works

As we have observed through many experiments (not depicted in this paper due to a lack of space), strong changes in the topology of the network due to threats imply strong variations in our metrics ; threats can then be detected more easily and in a more explainable way. The next phase of our research involves applying these metrics to real-world datasets. However, working with real-world datasets presents numerous challenges, particularly when attempting to detect changes in the behavior of large graphs. To address these challenges, we plan to apply explainable machine learning algorithms to detect threats, allowing us to provide detailed warnings to network administrators when malicious activity is detected.

Références

[1] V. Vlachos, Y. C. Stamatiou, P. Tzamalis, S. Nikoletseas, and K. Chantzi, "A social network analysis tool for uncovering cybersecurity threats," in *6th International Symposium for ICS & SCADA Cyber Security Research 2019 6*, 2019, pp. 97–106.

[2] F. Böhm, F. Menges, and G. Pernul, "Graph-based visual analytics for cyber threat intelligence," *Cybersecurity*, vol. 1, no. 1, pp. 1–19, 2018.

[3] H. Karimipour and H. Leung, "Relaxation-based anomaly detection in cyber-physical systems using ensemble kalman filter," *IET Cyber-Physical Systems : Theory & Applications*, vol. 5, no. 1, pp. 49–58, 2020.

[4] L. Akoglu, H. Tong, and D. Koutra, "Graph based anomaly detection and description : a survey," *Data mining and knowledge discovery*, vol. 29, no. 3, pp. 626–688, 2015.

[5] D. Sensarma and S. S. Sarma, "A survey on different graph based anomaly detection techniques," *Indian J Sci Technol*, vol. 8, no. 31, pp. 1–7, 2015.

[6] N. M. Adams and N. A. Heard, *Dynamic networks and cyber-security*. World Scientific, 2016, vol. 1.

[7] M. Ahmed, A. N. Mahmood, and M. R. Islam, "A survey of anomaly detection techniques in financial domain," *Future Generation Computer Systems*, vol. 55, pp. 278–288, 2016.

[8] D. K. Bhattacharyya and J. K. Kalita, *Network anomaly detection : A machine learning perspective*. Chapman and Hall/CRC, 2019.

[9] B. Bowman and H. H. Huang, "Towards next-generation cybersecurity with graph ai," *ACM SIGOPS Operating Systems Review*, vol. 55, no. 1, pp. 61–67, 2021.

[10] B. Perozzi, R. Al-Rfou, and S. Skiena, "Deepwalk : Online learning of social representations," in *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2014, pp. 701–710.

[11] A. Grover and J. Leskovec, "node2vec : Scalable feature learning for networks," in *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*, 2016, pp. 855–864.

[12] C. Yin, Y. Zhu, J. Fei, and X. He, "A deep learning approach for intrusion detection using recurrent neural networks," *Ieee Access*, vol. 5, pp. 21 954–21 961, 2017.

[13] Y. Mirsky, T. Doitshman, Y. Elovici, and A. Shabtai, "Kitsune : an ensemble of autoencoders for online network intrusion detection," *arXiv preprint arXiv :1802.09089*, 2018.

[14] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Lio, and Y. Bengio, "Graph attention networks," *arXiv preprint arXiv :1710.10903*, 2017.

[15] F. R. Chung, *Spectral graph theory*. American Mathematical Soc., 1997, vol. 92.

[16] F. Bauer and J. Jost, "Bipartite and neighborhood graphs and the spectrum of the normalized graph laplacian," *arXiv preprint arXiv :0910.3118*, 2009.

[17] S. de Lange, M. de Reus, and M. Van Den Heuvel, "The laplacian spectrum of neural networks," *Frontiers in computational neuroscience*, vol. 7, p. 189, 2014.

[18] C.-M. Cheng, H. Kung, and K.-S. Tan, "Use of spectral analysis in defense against dos attacks," in *Global Telecommunications Conference, 2002. GLOBECOM'02. IEEE*, vol. 3. IEEE, 2002, pp. 2143–2148.

[19] W. Wang and T. E. Daniels, "Diffusion and graph spectral methods for network forensic analysis," in *Proceedings of the 2006 workshop on New security paradigms*, 2006, pp. 99–106.

[20] P.-Y. Chen, S. Choudhury, and A. O. Hero, "Multicentrality graph spectral decompositions and their application to cyber intrusion detection," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2016, pp. 4553–4557.

[21] X. Ying, X. Wu, and D. Barbara, "Spectrum based fraud detection in social networks," in *Proceedings of the 17th ACM conference on Computer and communications security*, 2010, pp. 747–749.

[22] C. C. Bilgin and B. Yener, "Dynamic network evolution : Models, clustering, anomaly detection," *IEEE Networks*, vol. 1, 2006.

[23] Y. Chen and K. Hwang, "Collaborative detection and filtering of shrew ddos attacks using spectral analysis," *Journal of Parallel and Distributed Computing*, vol. 66, no. 9, pp. 1137–1151, 2006.