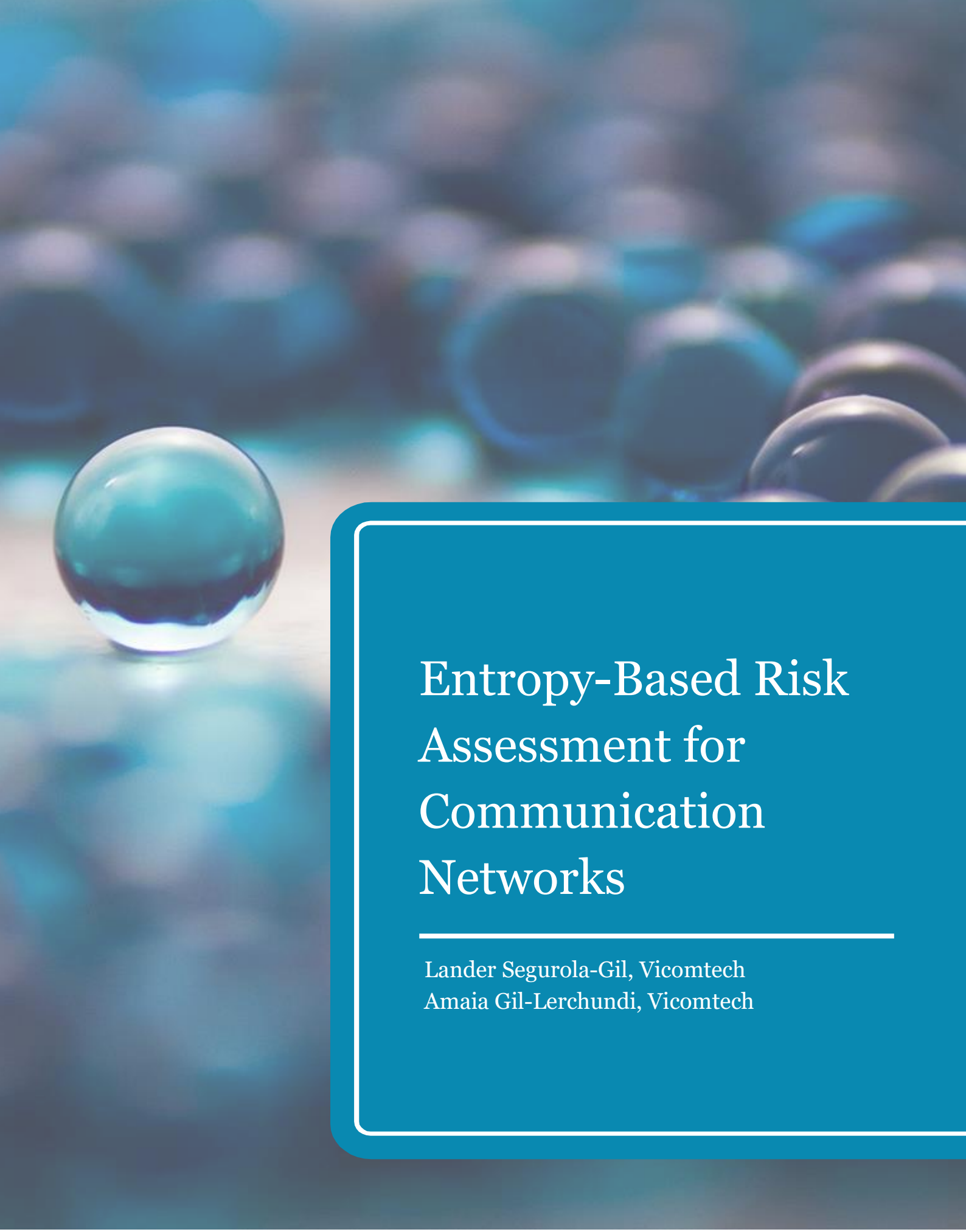

Entropy-Based Risk Assessment for Communication Networks

Lander Segurola-Gil, Vicomtech
Amaia Gil-Lerchundi, Vicomtech

Co-funded by the European Union

This project has received funding from the European Union's Horizon Europe Research and Innovation Programme under the Grant Agreement No. 101073909.

Entropy-Based Risk Assessment for Communication Networks

Lander Segurola-Gil and Amaia Gil-Lerchundi (Vicomtech)

In the midst of the Internet of Things, where interconnected devices have exponentially grown, cybersecurity has gained particular relevance in the modern society. Risk assessments a key concept in cybersecurity. Therefore, this work aims to provide a dynamic measure of uncertainty within a network, leading to optimization of decision-making when facing risky scenarios.

1. Introduction

The exponential increase in Internet connected devices has make cybersecurity a matter of concern [1]. In fact, cybersecurity breaches do not only affect in terms of information leak, but in economical, reputational, psychological, and societal terms too [2]. To cover the needs that cybersecurity problems require, several solutions have been proposed. In recent years, where the Artificial Intelligence has provoked a change of paradigm. In the past, a variety of conventional methods were employed to detect, and counter cyberattacks, yet these approaches proved inadequate against emerging threats. Presently, Machine Learning (ML) techniques have gained prominence across various domains (including cybersecurity), offering enhanced computational efficiency and rapid processing network data. Several ML based cybersecurity solutions have been proposed, as the literature [3] shows, where several ML based techniques are shown, facing different cybersecurity intrusion detection problems. A critical aspect of cybersecurity lies on the associated, informed risk assessment, focusing on quantifying the effects of a certain threat and the potential losses (financial, societal, environmental) related to the threat. For example, [4] provides an overview of risk assessment for Supervisory Control and Data Acquisition (SCADA) and Distributed Control Systems (DCSs).

2. The Current State of Affairs in Network Risk Assessment

Network risk assessment involves evaluating potential risks and vulnerabilities within a network environment. It typically includes asset identification, threat identification, vulnerability assessment, risk analysis, risk prioritization, mitigation strategies and monitoring and review. Overall, network risk assessment helps organizations identify and understand potential risks to their network infrastructure, enabling them to take proactive measures to mitigate these risks and enhance their overall cybersecurity posture.

Even the ML paradigm has found its place in risk assessment. For example, the work presented in [5], provides a fuzzy probability Bayesian network approach for Dynamic Cybersecurity Risk Assessment for Industrial Control Systems (ICS). In [6], the authors present an object typing, data mining and quantitative risk assessment approach for Smart

Cities. For that, they provide a Neural Network (NN) based solution, where the NN outputs a risk measure, given a device within the Smart City and certain characteristics of it. In general, state of the art works focused on risk assessment try to provide a measurable way of indicating a risk in a certain system. Entropy, as a measure of uncertainty, fits good with the problem characteristics. In [7], the authors provide a weighted entropy method to measure risk on cybersecurity systems.

Commonly, these approaches address risk assessment problems component wise, without a wider perspective of the matter.

3. The Role of Entropy-Based Risk Assessment (EBRA)

To the best of our knowledge, our work is the first one combining graph entropy measures together with machine learning anomaly detection algorithms to provide dynamic network topology state knowledge. In this way, our proposal provides a way to predict the uncertainty level of each device within a network, depending on the surrounding connections and device disposition. The key points for this are the intrinsic entropy of nodes induced by the surroundings within a graph and network traffic analysis. The analysis of network traffic provides the information for reconstructing the network topology with more or less accuracy (it depends on if the traffic has been gathered from a mirroring port for example, or, instead, the traffic is gathered from a machine within a LAN, where the reconstruction would be more limited) and characterizing the normal behaviour of device communications. Then the risk is measured in two ways. On one hand, entropy scores are computed for each node considering the normality state for the network connections (for more abnormal connections, more uncertainty in the network, and, thus, more risk). On the other hand, the other measure considered is the location of a node within a network (the more reachable a node is starting from any other node, the more uncertainty it causes, and, again, more risk). This technology fits particularly good in Software Defined Networks (SDR) where network topologies may be frequently changed.

4. The Research and Development Path in ATLANTIS

The work proposes a network traffic analyser that constructs a graph from the network traffic flows which represents the network topology. Once the graph is done, it computes the intrinsic entropy for each node within the graph. For this, several steps are followed. First, the graph is completed and weighted following reachability criteria. In other words, the weight for each edge in the completed graph is determined by minimum path length between two nodes in the original graph. In this way, we determine the distance between two nodes by a reachability criterion, computed on the number of steps to give. As the motivation relies on measuring the uncertainty of a node given its surroundings, we would like to measure how a node is affected by other nodes. Then, it could be inferred that the greater the distance between one node and another (resulting in more steps needed to traverse the path between them), the lesser the impact. Therefore, in a second step, the weights are replaced by their inverse. In this way, a probability of a node being affected considering all the surrounding nodes might be defined. As we constructed a complete graph, all nodes are connected to every other node. Thus, let it be v a node in the completed

graph G, and $w_{u,v}$ the weight of any edge (v,u) in the neighbour $N_G(v) \subseteq G$ of v. We may define a variable X_v by all the edges $(v,u) \in N_G(v)$. Then, $P\big(X_v = (v,u)\big) = \frac{w_{u,v}}{\sum_{w_{u,v}} w_{u,v}}$. Now, the entropy for each v, in the graph G, is defined as

$$H(X_v) = -\sum P\big(X_v = (u,v)\big) \log P\left(X_v = (u,v)\right)$$

This entropy measures the uncertainty of a node within a network.

For example, in the graph in Figure 1, under the defined value, the darkest blue will have the biggest (most "centralized" node) entropy, whereas the white one, will have the lowest (most "isolated" node). This might be interesting to understand and optimize network topologies. However, for static networks, this might not provide more than a first impression of the state of a network. To overcome that, anomaly detectors are integrated into the solution, to provide information of the connection between nodes. Let $M: R^n \rightarrow [0,1]$ be a model which analyses n dimensional points gathered from network traffic metrics and outputs an anomaly score between 0 and 1. Let $x_{u.v}(t)$ be an n dimensional point collected at time t for a given (u,v) connection. Then a temporal dependant weight is defined such as $w_{u,v}(t) = M(x_{u.v}(t))w_{u,v}$. From here, $P(X_v = (v,u), t, M) = \frac{M(x_{u.v}(t))w_{u,v}}{\sum_{w_{u,v}} w_{u,v}}$ and

$$H(X_v, t, M) = -\sum P(X_v = (u,v), t, M) \log P\left(X_v = (u,v), t, M\right).$$

Like this, maximum entropy for a node is achieved when surrounding connection's anomaly score is maximum $\big(M\big(x_{u.v}(t)\big) = 1\big)$ and minimum is achieved in the opposite case $\big(M\big(x_{u.v}(t)\big) = 0\big)$.

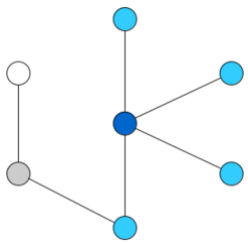

Figure 1. Graph example.

5. The Challenges and Barriers

As described in Section 2, the methodology exposed is particularly well-suited for SDNs, where network topologies may undergo frequent changes. However, this particular method might face network reconstruction problems other scenarios. If network topology is provided (periodically in scenarios where the network changes frequently), the method would not face any barrier. The problem comes when no information is given and the accessibility to network traffic is limited (not in a port mirror for example). In this case, some heuristics may be applied to reconstruct the network, but probably local network may be the only rebuildable network.

A more general problem, but common, is related to data quality. If network data quality does not reflect the normality of the network, then the anomaly detection models will not be able to detect anomalies in a proper way.

6. The Benefits and Impact

The technology provides a way of measuring the uncertainty that a node faces given its surroundings, leading to an increased perception of the network state. Given this information, an operator might be assisted on decision-making, optimizing network topologies into more secure ones (entropy minimization) and deciding whether a node should be isolated or moved at a certain point in time due to entropy maximizing anomalous connections. This is directly related to Dynamic Risk Assessment handling, as the lower the entropy (uncertainty), the lower the risk, improving security and safety of device networks.

7. Future Outlook

The technology is adaptable to any dynamic environment, as is intended to be so. Some updating mechanisms should be integrated to optimize the way the model is updated. However, due to graph problems nature, the scalability when facing huge graphs might be limited. Again, for facing this kind of problems, some heuristics might be adopted. In general, except for the mentioned case, the proposed method adapts to any kind of communication network, independently to the underlying network producing use case.

8. Conclusions

This work provides the technical insight of an entropy-based risk assessment (EBRA) tool, that computes the entropy of a network given its topology and anomaly scores for each connection. This provides a measure of risk with which an operator of a network can optimize its decision-making. The method is scalable, even if some heuristic approach might be adopted in huge network scenarios.

References

[1] S. Ray, Y. Jin in A. Raychowdhury, „he changing computing paradigm with internet of things: A tutorial introduction. I," *EEE Design and Test 33,* p. 76–96, 2016.

[2] I. Agrafiotis, J. Nurse, M. Goldsmith, S. Creese in D. Upton, „A taxonomy of cyber-harms: Defining the impacts of cyber-attacks and understanding how they propagate.," *J. Cybersecur.,* 2018.

[3] P. Dixit, R. Kohli, A. Acevedo-Duque, R. Gonzalez-Diaz in R. Jhaveri, „Comparing and analyzing applications of intelligent techniques in cyberattack detection.," v *Security and Communication Networks*, 2021.

[4] Z. M. King, D. S. Henshel, L. Flora, M. G. Cains, B. Hoffman in C. Sample, „Characterizing and measuring maliciousness for cybersecurity risk assessment.," *Frontiers in psychology,* p. 39, 2018.

[5] Q. Zhang, C. Zhou, Y. C. X. Tian, N., Y. Qin in B. Hu, „A fuzzy probability Bayesian network approach for dynamic cybersecurity risk assessment in industrial control systems.," *IEEE Transactions on Industrial Informatics,* pp. 2497-2506, 2017.

[6] M. Kalinin, V. Krundyshev in P. Zegzhda, „Cybersecurity risk assessment in smart city infrastructures.," *Machines,* 2021.

[7] T. Hamid, D. Al-Jumeily, A. Hussain in J. Mustafina, „Cyber security risk evaluation research based on entropy weight method.," v *International Conference on Developments in eSystems Engineering*, 2016.

Front cover image by SplitShire via Pixabay.
https://pixabay.com/users/splitshire-364019