Interactions between artificial intelligence and cybersecurity to protect future networks

Gregory Blanc, Yang Liu, Rongxing Lu, Takeshi Takahashi, Zonghua Zhang

Annals of Telecommunications (2022) 77:727–729
https://doi.org/10.1007/s12243-022-00935-6

EDITORIAL

Interactions between artificial intelligence and cybersecurity to protect future networks

Gregory Blanc[1] · Yang Liu[2] · Rongxing Lu[3] · Takeshi Takahashi[4] · Zonghua Zhang[5]

Published online: 26 November 2022

Artificial intelligence (AI), as a field, is one of the most broad and dynamic, attracting more and more students, practitioners, and researchers around the world. In particular, natural language processing, one of its many avatars, is in the news every week. Recently, various programs have been released enabling anyone to generate artistic or realistic renditions of text prompts.

The general public can therefore experience firsthand what AI is.

On the other hand, these AIs can be diverted to fool or scam other human beings, such as face generation programs, raising ethical issues. This has prompted AI researchers to take matters into their hands to improve the fairness and safety of machine and deep-learning algorithms. Indeed, as AI is expected to penetrate all sectors of society, from healthcare to agriculture or transport, lawmakers become wary of its influence on society, and try to legislate on its usage, preventing its ability to automate all complex decision-making processes.

This is why, human-in-the-loop approaches are mostly developed, in particular for safety-critical applications.

Cybersecurity is another such critical sector that may benefit from AI approaches for tasks as diverse as malware analysis, intrusion detection, alert correlation, and threat intelligence, with a goal of enhancing the cybersecurity situation of next-generation networks.

Their efficiency will rely on AI to discover patterns and make predictions in complex, distributed, and heterogeneous network environments. For more than three decades of interactions between AI and cybersecurity, we have witnessed the tremendous capabilities offered by AI in analyzing threats, risks, and attacks in various networked systems, enabling comprehensive and in-depth defense.

There is also an emerging trend on utilizing AI to assess and improve cybersecurity measures, e.g., by generating evaluation datasets, which cybersecurity is direly lacking.

On the other hand, AI platforms, algorithms, and systems are attracting significant attention from the cybersecurity community because of their increasing development, deployment, and application in our ICT infrastructure and services.

Recently, adversarial machine learning approaches have also prompted considerations for more critical thinking with respect to the adoption of AI-based systems, as next-generation network infrastructures that depend on them could then be crippled by AI-related incidents.

This Special Issue is focused on the intersection between AI and cybersecurity for the benefit of next-generation networks, with an objective of bringing together the engineers, researchers, and practitioners from both communities, in industry and academy, as well as stakeholders in next-generation networks, to explore together the emerging issues, topics, and solutions.

We solicited submissions on topics of interest as diverse as the application of AI to the cybersecurity of next-generation networks, including but not limited to intrusion detection (anomaly-based, malware, botnet), root cause analysis, security information and event management, network forensics, countermeasure selection, and synthetic data generation, as well as critical analyses of the application of AI or machine learning to cybersecurity.

We also received submissions with respect to the security of AI-based next-generation networks, with topics spanning from the security and privacy of the AI algorithms to evasion and deception, to robustness against adversarial examples, or trust in AI platforms.

Not only was the review process in a world stricken by a pandemic heavily slowed down, but we had to constantly

✉ Gregory Blanc
gregory.blanc@telecom-sudparis.eu

[1] Institut Mines-Télécom/Télécom SudParis, Institut Polytechnique de Paris, Palaiseau, France

[2] Nanyang Technological University, Singapore, Singapore

[3] University of New Brunswick, Fredericton, Canada

[4] Cybersecurity Laboratory, National Institute of Information and Communications Technology, Koganei, Japan

[5] Huawei France Research Center, Paris, France

reach out to both communities to solicit quality contributions. It was a feat in itself, and we wish to thank all the reviewers that contributed to this Special Issue, as well as the editorial team at the *Annals of Telecommunications* journal.

But first and foremost, we wish to extend our deepest gratitude to the authors of both accepted and more unfortunate submissions who did put their trust in our guest editorial team to review and manage their submissions in such uncertain conditions.

This Special Issue, although modest in length, gathers articles offering insights and perspectives into a diverse set of topics from detection to incident response, applied to domains as varied as aeronautics, IT networks or computer processors, and exploiting models trained on a broad range of algorithms. Classification is a supervised learning task that is quite represented in the field of intrusion detection. As such, we can trace first occurrences of machine learning approaches to intrusion detection in the late 1990s or early 2000s. Considering the wealth of contributions in event detection, Mushtaq et al. (2022) set out to model hardware events representative of legitimate and malicious programs' access to cache memory. Their solution, the Kingsguard, a detection-based mitigation approach, is relying on an intensive profiling phase of hardware performance counters (HPCs).

When monitoring HPCs under realistic conditions, increased interferences with caches make it hard to separate attack and non-attack scenarios using a simple threshold, even more so when subject to multiple cache side-channel attacks (CSCAs). As protection against such attacks require to monitor the whole computing stack, it incurs some large overhead that could heavily impact the computing performances. To prevent that, Mushtaq et al. (2022) design their system to provide needs-based protection. Several clustering approaches are compared on stringent criteria including classification accuracy and minimum overhead, among others.

In domains as uncertain as aeronautical radar communications, it is difficult to characterize and classify all suspicious patterns. Departing from supervised learning, anomaly-based detection approaches offer the ability to detect unknown events, possibly indicative of attacks, without the need to actually learn all attack patterns, hopefully preventing 0-day cases. In their work, Riberolles et al. (2022) endeavor to propose an anomaly detector able to indicate which communications may have been hijacked by a spoofing attacker, in a man-in-the-middle (MiTM) setting. Convinced by their survey of the state of the art to use time series analysis, they settle for a long short-term memory (LSTM) anomaly detection approach which has proven effective when using the prediction error to measure the anomaly severity. Such prediction error is obtained using an autoencoder, an unsupervised method solely trained on normal data learning its relevant characteristics.

At inference time, the autoencoder detects deviations from the learned behavior with high accuracy. Their resort to a deep-learning approach is motivated by the extended feature space induced by the use of sliding windows to obtain even richer time series. The sliding windows actually help magnify the anomalies, yielding satisfactory detection results in the face of half a dozen attack patterns performed against very different aircraft radar traces, demonstrating its stability, even across other industrial control system sectors.

Following the renewal of interest in leveraging machine learning for intrusion detection, it is natural to question the effectiveness of such approaches, in particular its robustness against adversarial examples. Adversarial examples are usually malicious samples whose features are slightly perturbated to induce misclassification, i.e., generating false negatives. Adversarial machine learning employs these examples to improve the robustness of intrusion detectors. However, Merzouk et al. (2022) demonstrate several issues about the practicality of adversarial examples generated for network intrusion detection systems (NIDS). They show that perturbations modify the features of malicious samples in ways that do not conform with network-related data constraints. Merzouk et al. (2022) study the impact of state-of-the-art adversarial example generation methods on the performance of a multi-layer perceptron (MLP) on three selected datasets. These methods can often be divided into the ones that slightly modify most features and the ones that heavily modify a few features, the latter usually faring worse with respect to network compliance. Their work proposes a few criteria to assess the quality of the adversarial examples generated by the selected methods, both at syntactic and semantic levels.

On the other end of the cybersecurity defense chain, incident response usually relies on threat intelligence to provide the most appropriate remediation. In particular, when treating an alert, the amount of actionable information, i.e., information that can be exploited to take actions, is critical in that it should be sufficient to take decisions, and not too noisy to prevent excessive processing overhead. Threat intelligence information is often available in a text format. It is thus natural to conceive that its processing can be advantageously automated using machine learning, in particular NLP. Osada et al. (2022) actually take one step back as security operators do not perfectly agree on how to label the information, making a keyword-based search engine unlikely to yield relevant results. Their approach uses topic models and outlier detection with the goal of improving the number and quality of data-mined labels from large corpora of threat intelligence documents. They propose a 6-step methodology able to extract more precisely multi-labels to be assigned to the summary documents presented to human operators. They produce a realistic evaluation of their methodology against a series of malware campaigns. The resulting topics are useful for prompting the next search query in in-depth investigations.

As AI is expected to be used in interaction with human beings as is the case for the above-mentioned incident response example, the decision-making may still be the responsibility of the human in the loop. At least, these decisions should be sufficiently explicit so that a human operator can take the decision or understand why it has been taken. With the advent of black-box deep learning for cybersecurity, concerns were raised on the ability to explain why certain predictions were made. It is even more serious when other aspects of the models are questioned such as their security, their fairness, or their privacy. All these aspects are indeed essential to critical applications, such as cybersecurity.

This Special Issue thus concludes on a survey of the interactions between explainable AI (XAI) and cybersecurity. The literature review by Charmet et al. (2022) offers a dual perspective, and so does this Special Issue: on one hand are surveyed contributions of XAI to classification tasks in cybersecurity, and its added value in terms of trust, performance, or robustness; on the other hand, the security of XAI methods is scrutinized as explainable methods have also been demonstrated to be vulnerable to attacks against their fairness, integrity, confidentiality, or robustness. The review attempts to introduce the readers to the recent field of XAI and its main achievements in cybersecurity while discussing its known shortcomings.

In conclusion, we invite the readers, be them AI or cybersecurity researchers, to reinforce their interactions for the sake of a safer Internet and hope that the selected articles of this Special Issue will prompt new discussions and proposals to tackle the mentioned open issues.