Where are Cybersecurity Boundaries?

(This summary is based on the Panel discussion of SciSec2021, which is edited by Weixia Cai and includes ideas of Yugang Jiang, Feng Liu, Wenlian Lu, Kun Sun, Lingyu Wang, Shouhuai Xu, Moti Yung)

2021-9-6

Scisec'2021 was successfully held totally online on August 13-15, 2021. At the conference, a wonderful panel discussion on the topic of "Where are Cybersecurity Boundaries?" was successfully held, with Professor Shouhuai Xu being the moderator and Professor Yugang Jiang, Professor Lingyu Wang and Professor Moti Yung being the panelists. The panel discussed three specific issues concerning cybersecurity boundary, which helped deepen our understanding of the Science of Cybersecurity.

Before the panel discussion, it's necessary to make two terminologies clarifications to set the stage for the panelists, the audiences, and the readers of this note. The term "cybersecurity" broadly deals with entire network and devices in question (e.g., cyberspace, infrastructure, enterprise networks, cyber-physical-human systems) holistically (rather than specific building-blocks). Moreover, the term cybersecurity accommodates security, agility, resilience, etc. The term "security" specifically deals with building-blocks and components, accommodating standard notions such as C.I.A., authentication, etc. For the purpose of this panel, the term "cybersecurity" (or "cyber security") is much broader than "security".

Science of Cybersecurity is a scientific discipline addresses practical and fundamental problems. Currently, practical cybersecurity problems are better understood (e.g., communication security and private computing via cryptography), while fundamental cybersecurity problems are much less understood, with exceptions of undecidable problems. An undecidable problem is a problem for which it is proved to be impossible to construct an algorithm that always leads to a solution. Take universally capable malware detector as an example, there is no such algorithm is capable for detecting malicious viruses universally and correctly.

The three questions discussed by the Panel are summarized as follows.

Where is the boundary between the cybersecurity problems that can be tackled by the logic and mathematical approach and the cybersecurity problems that cannot?

To formalize things carefully and logically, there are two types of boundaries. One boundary is the understanding of problems. If a cybersecurity problem cannot be understood or modeled mathematically, there will be no solution to it. In the history people encountered many such hard problems, for those hard problems that cannot be fully formalized, people are not hopeless. Such problems can be broken down into sub-problems. A different language, such as graph theory, is then used to get a better solution. It's kind of a gradual process started with what we can handle. And eventually there will be some unified understanding of the problem. The other boundaries is the undecidable problems. In cybersecurity, there are many problems that are less possible to model, examples include intrusion detection, zero-day vulnerabilities discovery, social engineering, phishing email and other human factors. Such problems cannot be tackled fundamentally using a logic and mathematical approach. However, in the practical sense, most problems can be modeled as long as sufficient assumptions are made, and an approximate solution can then be obtained. The following question is how to be sure the assumptions are complete and realistic.

Where is the boundary between the cybersecurity problems that can be tackled by the AI/ML/Data Sciences approach and the cybersecurity problems that cannot?

It is difficult to define a boundary of AI/ML/Data Sciences approaches when trying to increase the security level of a real system. Some specific engineering problems can be solved by AI/ML/Data Sciences approaches as they have good performance and help to understand some relationships at the beginning of a research. These methods are scattered throughout each cybersecurity problem, like rational numbers scattered throughout real numbers, they extend to wherever cybersecurity problems are extended. Nevertheless, blandly using of these algorithms may not be the right answer due to their poorly explanations, and the design of the current artificial neural network is quite different from the neural network in our brain. Furthermore, there are cybersecurity problems cannot be solved by these methods owing to a lack of deep understanding. It does not mean these problems are the boundary of the methods, just like 100-meter athletes continue to break the human perception of physical limits, people's understanding of cybersecurity problems is also constantly breaking through. In view of the technological trends in the last decades, some new fundamental breakthroughs in cybersecurity field are needed to understand the problem more precisely and turn an unsolvable problem into a practical one.

Where is the boundary between what are quantifiable vs. what are not quantifiable in cybersecurity?

Usually, we use metrics to quantify cybersecurity. From this point of view, the quantification of cybersecurity can draw on interdisciplinary help, such as analogies with economics. Microeconomics deals with a firm or a house while macroeconomics manages nationwide or worldwide things, they all concern about what is going to happen when a point changes. Even though there is a gap between economic and cybersecurity, and the tools used in economics cannot be applied directly, it would be a promising way to coherently connect the understanding of these two bodies. In addition, it should be clear that doing quantification is not just about metrics. Security metrics is a mathematical model based on assumptions which can be wrong. For example, when strategies are adopted to make metrics perform better, it may also create a broad bad impact on cybersecurity. The curious but meaningful question is if you can't quantify metrics very well how to quantify the whole system? Therefore, we can never apply metrics too carefully.

These issues are related to each other. In order to quantify some problems, mathematical models must first be developed, which in turn requires a sound understanding of the problem. And the boundaries are much harder if humans are part of the situation, but it is with humans that there is cybersecurity, where uncertainty is inherent and we have to live with approximate solutions. In summary, these are fundamental and intriguing questions, we hope this panel discussion will inspire and further help you in finding opportunities for your future research.

Disclaimer: This summary is not reviewed by the panelists. It is made available on the conference website purely for the purpose of academic knowledge sharing.