AI for Security
SaTC 2019 PI meeting breakout group report

Co-leads: Kamalika Chaudhuri (UCSD), Hao Chen (UC Davis), Xinyu Xing (Penn State)

1. Problem/Domain Summary

Detecting and defending against security threats are important and integral parts of the SaTC agenda. Unfortunately manual detection is extremely time-consuming, and currently, existing methods require a great deal of human effort. The hope is to integrate current methods with AI/machine learning to ensure the less human intervention is required in the process.

In this breakout session, the group summarizes the following topics

- Using AI to scale the cybersecurity workforce (e.g., finding software bugs efficiently and effectively)
- Using AI to build proactive security defense (e.g., building alpha0 type of learning method from rules)

2. Key Research Challenges

The group agreed that a major research challenge in this area is the lack of well-annotated and curated datasets. This is particularly hard as data often have partial and incomplete labels and annotating it properly requires a great deal of manual effort and expertise.

A second major challenge is that current AI/ML methods lack good quality explanations for the predictions that they make. This makes it hard for a user to understand predictions and use them in decisions.

A final difficulty in current methods is also mostly data-driven and it is hard to inject domain knowledge into them.

3. Potential Approaches

The group agreed that we need more and stronger community efforts for building high-quality and representative datasets for this area. There is also room for closer collaboration between domain experts and technical experts. This can be facilitated by closer collaboration between security and AI communities, as well as collaboration with HCI/data visualization experts. Finally, there is a great need for better AI explainability solutions that might be tailored towards security applications; this can be done by building stronger foundations and methods that can answer different kinds of “why” questions on decisions made by black-box classifiers.

4. Long-Term (> 10 years) Significance

The discussion group believes this domain/problem will remain relevant in 10 years. This is because we have already witnessed AI has been largely used for classification and received great success in some cybersecurity tasks. With AI being used in more security tasks (e.g., software fuzzing and disassembling), new AI technologies need to be designed and developed. In addition, we believe AI will become a target of attacks. Motivated by this, we envision that new defense methods for AI are also needed in the future.