AI for Security
SaTC 2022 PI meeting breakout group report

Co-leads: Sagar Samtani (Indiana University), Gang Wang (University of Illinois, Urbana-Champaign), and Anita Nikolich (University of Illinois, Urbana-Champaign)
Scribe: Saptarshi Debroy (City University of New York)

1. Problem domain and motivation

In this day and age of unprecedented cyber threats, common cybersecurity tasks such as asset identification, control allocation, vulnerability management, and threat detection and response are facing major challenges of data overload with limited human resources. AI can address such challenges by automating common cybersecurity tasks, e.g., sifting through data more efficiently and effectively than a human and identifying patterns in large datasets missed by manual analysis. There are major gaps in existing efforts on AI for security, including siloed resources and initiatives between academia and industry, lack of publicly accessible and realistic datasets, lack of standardization in model/data/software sharing, poor usability of the tools/systems, challenges to scale up/out, lack of adaptability and transferability across scenarios, and the difficulty to incorporate/integrate with human intelligence/knowledge.

In this breakout session, the group discussed the following thrusts that can help to better determine the roles of AI in cybersecurity and bridge the existing gaps:

- Data/Model Sharing Needs
- Needs of AI Working with Humans
- Needs of AI-Cyber Workforce Development

2. Data/Model Sharing Needs

On the topic of why data sharing is not prevalent in the community, the group agreed that discoverability of useful data is a problem. There is a need for data centralization (maybe enabled by an NSF program for data curation) along with proper tools and resources to ascertain the data quality and maintain useful metadata. The group also identified that encrypted data from different sources along with industry imposed restrictions and industry-standard NDAs and Data Use Agreements (DUAs) are barriers towards wider adoption. The group also discussed ideas to encourage researchers to share useful datasets (e.g., having a dataset-track in top conferences).

On the topic of how to standardize the data sharing process and format, the group argued that increased centralization can improve standardization. The group identified the lack of security related datasets in APT, automotive, and softwarized (i.e., SDN) system domains. The group argued that NSF (in particular the Office of Advanced Cyberinfrastructure (OAC) can take steps towards such standardization and data collections and centralization through grant supplements and dedicated data curation related programs.

There was a discussion on whether it is possible to share large “pre-trained” models for general downstream security applications, where the group’s response was affirmative with the dedicated focus on developing and sharing such models, e.g., HuggingFace. The role of transfer learning/knowledge was emphasized, as well as the need for model validation.

On the topic of running large-scale AI models to secure resource constrained systems (e.g., 5G/6G, IoT devices), the group identified the role of PAWR testbeds such as AERPAW and FABRIC in generating new datasets through community experimentation. The group also identified lack of information on security-related instrumentations on such platforms, which often prevents useful security experimentation.

3. Needs of AI Working with Humans

On the topic of how to make AI-based security tools truly usable, the group emphasized the need to understand analyst-centric, security operation center (SOC)-type workflows and the associated operational concepts/end-goals when developing AI tools. The group recognized the need to propose new sets of usability metrics and develop consolidated toolsets for practitioners.

On the topic of why human analysts trust/distrust AI systems, many in the group argued that in many cases operational environments are different than theoretical. Thus the community needs more resources to train AI systems on operational fronts using relevant real world data. The lack of explainable AI models in security uses in particular, which results in critical decisions, is a barrier towards trust. The group believed the community needs to form consensus on what types of AI models can (are suitable to) solve what types of problems.

When discussing how human analysts can work/learn with AI systems, the group felt that it is a major challenge to establish the explainability of AI models to real users. It is also necessary to go beyond labels when building human-AI collaboration systems. More studies are needed to enable knowledge exchange between human and AI systems and to better understand the context dependency. The group argued that it is important to adopt human-in-the-loop models. The role of ethics in such issues is also identified as an important problem space.

4. Needs of AI-Cyber Workforce Development

While discussing the challenges that many institutions face when teaching AI for cybersecurity, the group identified institutional roadblocks as major non-technical challenges, and discoverability and accessibility of resources, and benchmarked datasets and testbeds as major technical barriers. The group felt that AI for cybersecurity related competitions can be useful.

On the topic of how AI-cyber lab resources are different from cybersecurity-only resources, the group identified larger data/compute requirements, need for more background knowledge, and

lack of faculty expertise as major barriers. The group felt that developing courses/labs with co-instructors having synergistic knowledge and interdisciplinary backgrounds can be useful solutions.

On the question if knowing employer needs for AI for cybersecurity would help with curricular development, the group was of the opinion that such knowledge can help specify real/operational problems to effectively train students and professionals. The need for balancing fundamentals with industry/operational trends was also discussed.

Finally, on the topic of how to promote dataset and resource sharing in the CyberCorps community to help develop AI for cybersecurity curriculum, the group strongly recommended convergence of AI and cyber communities, developing competitions, and new textbooks and playbooks.