Game Theory and Distributed Systems Security
NSF SaTC PI Meeting Breakout 2022

Saurabh Bagchi[a] Kevin Chan[b] Mustafa Abdallah[a] Xing Gao[c]

[a]Purdue University, [b]Army Research Lab, [c]University of Delaware

This serves as the report of the breakout session at the NSF SaTC PI meeting that happened in Arlington, VA. The first day's session had enthusiastic participation from 30 members while the second day had about 12 members. We then reported out the result of our discussions to the combined attendees of the entire SaTC PI meeting.

We started off preparing for the breakout session by creating a pre-session survey. The objective of the survey was to gauge the mood of the potential attendees. We asked questions on what they believed to be the right foundations to build on and what are the important short-term and long-term research problems in this topic, to be chosen from among a pre-populated list. Each item was to be marked as "Important", "Neutral", or "Not Important". The respondent could also include free-form responses. The survey was distributed to a subset of the attendees, considering those who the organizers knew in some capacity. The hyperlink to the survey is https://bit.ly/satc22presurvey and readers are encouraged to take this survey.

1. *Foundations:* The most important foundations were considered to be "Intrusion root cause analysis and response", followed by (equally) "Intrusion detection in distributed applications" and "Control & data plane attacks".

2. *Short-term research problems:* The most important one (from among 5 choices) was "Adapt some theory (game/convergence/percolation/...) into practice of security", followed by "Adaptive techniques for dynamic environments".

3. *Long-term research problems:* The most important one (from among 3 choices) was "Can we secure heterogeneous systems, with some nodes possibly resource constrained?", followed by "Can we build secure distributed applications with partially trusted data sources?". The third one was "Can we integrate game theory and machine learning to secure distributed systems?"

1 Problem Context

There has been significant work in understanding vulnerabilities in large-scale distributed systems and putting technological patches to address specific classes of vulnerabilities. However, the works often lack understanding of the impact of cascading attacks or mitigation on the resilience of the overall system. Due to the large legacy nature of many distributed infrastructures and budgetary constraints, a complete re-architecting and strengthening of the system is often not possible. Rather, rational decisions need to be made to strengthen parts of the system, taking into account the risks and the interdependencies among the assets.

While static game theory has been extensively studied for several decades, the large-scale distributed systems present critical challenges that preclude the direct application of existing theory. Specifically, there is a need for new techniques to account for both the interdependencies and the dynamical nature of the subsystems. Furthermore, some of these dynamical subsystems may be complex in their own right (e.g., a perception system that employs multi-modal sensors) and may only be represented by simulation models.

This problem context led us to three overarching questions that formed a starting point for our discussion.

1. Can the security community extend traditional game theory to develop tractable analysis and design techniques that can be applied to security of large-scale interdependent distributed systems?

2. Can the community learn from behavioral economics where human biases are taken into account in decision making?

3. Can that be incorporated into traditional game theory to understand the effect of biases on security decision making and possible mitigation actions?

This session has the goal of understanding the foundations that we can build upon and the open challenges that the community can get behind, both in the short term, defined by us as 2-3 years, and in the long term, defined by us as 5-10 years.

2 Foundations: Build on Them

We have significant foundations on the topics of distributed systems security and game theoretic security that we can build upon. These have led to beneficial outcomes to society that the technical community should capitalize on and amplify through ongoing work. Here we survey the notable foundations categorizing them into two categories.

2.1 Game-theoretic Security

There have been notable successes in developing and applying game theory for security of interdependent systems. This has been used in the context of proactive or reactive, fixed or adaptive schemes. The most commonplace game-theoretic model for security is that of two-player games, where a single attacker attempts to compromise a system controlled by a single defender [4].

One promising line of work has been in **Reinforcement Learning**. Examples are RL under malicious falsification of cost signals that has been used to mislead agent policy [8]. Another has been RL and Infinite-horizon Semi-Markov Decision Process (SMDP) to characterize a stochastic transition and sojourn time of attackers in the honeynet [7].

Game theory for security has been found to be tractable when considering sequential attacks, through **Stackelberg security games**. Such games may incorporate real-time observations and consideration of non-myopic players [15]. In reality, many such games may be partially observable [5, 25] as the actions of a player may not be visible to other players (e.g., an attacker may conceal her steps).

Game theory for security of distributed systems. Game theoretic models have also been used in [13, 21, 22] to study the Distributed Denial of Service attacks. Game theoretic models have also been used to study critical infrastructure security [11], censorship-resilient proxy distribution [14], and protecting networks from cascade attacks [10].

2.2 Distributed Systems Security

One way of organizing the foundations that have been developed here is through each step of the workflow for distributed systems security, namely, detection, diagnosis, and containment and response.

Detection is a mature area of work and relevant to our breakout, there is influential work on collaborative intrusion detection using multiple sensors placed in a distributed system. This line of work has contributed algorithms to determine where to place the sensors and how to integrate outputs from multiple sensors to come up with an integrated decision on the detection of an attack. A survey work on this topic is [17]. This area saw some of the early applications of ML to security.

Diagnosis has contributed algorithms to identify the root cause of the attack. This was initially substantially rule-based of the form if metric A $>$ threshold $\tau 1$ and B $<$ threshold $\tau 2$, then A is the root cause [9]. Later, foundational work was done on this topic on using ML, such as causal theory [18]. Diagnosis has used to create isolation zones when an attack is spreading through a distributed system. A key problem that has been solved is, when interactions between elements of the distributed system are changing dynamically or when connections are changing dynamically, how can the security algorithm still achieve diagnosis.

Containment and Response has had notable success in Moving Target Defense, which seeks to change some parameters of the defended system such as IP address to thwart adversary [3]. Game analysis for critical infrastructure protection has been quite successful where the protection encompasses containment and response [2, 7]. A commonly used model for driving the algorithms is Attack graphs which can seamlessly represent interdependent systems [1].

3 Challenges Ahead

Here we summarize the technical challenges on the topic of this breakout session that the technical community can get behind. These are both technically substantive challenges to be solved as well as of importance that they will decisively improve the security landscape of distributed systems that our society relies on.

We structure our discussion into challenges that are on the **analytical directions**, **systems directions**, and those that involve combination of the two called, **integration directions**. The orthogonal dimension on which we structure these challenges is the time horizon for us to solve them, with short term indicating 2-3 years and long term indicating 5-10 years.

3.1 Analytical Directions

1. **Personalized learning:** (Short time horizon) Different actors (say different defenders and adversaries) learn differently and at different rates. The learning happens for human actors as well as for machines (in an ML context). Another form of heterogeneity that comes in is asymmetric capabilities among the various players, e.g., some defenders have access to assets where there is trusted hardware environment, like ARM's TrustZone or Intel's SGX.

2. **Incorporating biases and incomplete information:** (Short time horizon) For the learning, one has to incorporate incomplete information sharing among the actors. For the human learning, one also has to incorporate cognitive biases among the human players, such as, the predisposition to overweight low probability events and to underestimate high probability events. The personalization of the learning must also be able to accommodate partial cooperation (among defenders) or partial collusion (among adversaries), in addition to the typical formulation of complete cooperation or collusion.

3. **Scalability and tractability:** (Long time horizon) A well-known challenge with the application of game theoretic formulation to security of distributed systems is the scalability and the tractability of the solution. Scalability implies scaling to the large number of actors or large amounts of data or large volume of interactions among the actors. Tractability implies being able to handle realistic attack models or realistic workloads incident on the protected system. To ease this challenge, we need to develop rigorous approximation of the game-theoretic formulation. This should allow one to produce bounds for best-case/worst-case outcomes. As an example, one can use scalable techniques from epidemic theory to analyze effect of cascading attacks while accommodating the case of a large numbers of players.

4. **Incorporating stochastic behavior in game theoretic formulation:** (Long time horizon) The game theoretic formulations are often rigidly deterministic in nature, e.g., a specific deterministic action is coded in for a particular state. The open question is can machine learning be integrated with game theory and incorporate stochastic behavior. This is important as failures and attacks are inherently stochastic in nature.

3.2 Systems Directions

1. **Resource-aware defenses:** (Short time horizon) Different nodes have different capabilities and available resources and the system should be able to calibrate the defense mechanism using node-specific attributes. Some of these node-specific attributes will be static and unchanging, such as, the intrinsic hardware capability of the node, while some attributes will be dynamic, such as, the current battery level on the node. The cost of an attack may also be varying, e.g., the cost to corrupt data may be higher if there is some security protection being overlaid on the data.

2. **Security guarantees as a dynamic function:** (Short time horizon) The security guarantees may, under certain situations, be a function of the current system state. As an example, under this regime, the guarantees could be a function of the number and the capability of attackers and defenders rather than an absolute. Thus, the security guarantees that the system can provide are a dynamic property varying with the system state. For example, hardware degrades and the software ecosystem changes over time. The guarantees could also be a function of the level of collusion among attackers, e.g., non Byzantine or Byzantine attackers.

3. **Designing for security in the tradeoff space:** (Long time horizon) A radical design principle would be to design for security in the tradeoff space between security and (performance and resource usage). For example, the security design may use hardware-level virtualization rather than (software) containers, the former providing greater protection against side channel attacks. If one can design specialized functions (specialized to the resource available at a

node say), this has the added benefit of reducing the attack surface and making debugging easier. The security guarantees must be clearly delineated as a function of the performance and the resource usage, so that the end user can understand the guarantees she is getting or in case of automatic composition with other software packages, it becomes clear what security guarantees are in effect in the composed system.

3.3 Integration Directions

Here we lay out the research challenges that will need integration of advancements on the analytical side and on the systems side.

1. **Security of distributed systems in CPS domain:** (Short time horizon) To secure Cyber-Physical Systems (CPS), which are a prototypical distributed system, there are several unique aspects that we need to consider. These are typical interdependent systems, often with multiple stakeholders as owners. The nodes are embedded in physical environment and are subject to environmental effects, which contribute to the difficulty of securing them. For example, it is often difficult to tell apart a node malfunction due to environmental effects from a node compromise. Further, some parts of the system are opaque to defenders, as they are developed by an external party. Consequently, security mechanisms that rely on observability, even fine-grained observability, of the events happening in the software stack are out of bounds.

2. **Continuous verification:** (Short time horizon) This topic needs to answer the question: are our models and practical software instantiations generating useful results even under attacks and perturbations? This should be done on a continual basis rather than in batch mode as is typically done today, when verification is used at all. The continuous verification should happen as the system is processing inputs and generating outputs during its operation. This could use *sparse* human feedback online, i.e., without putting undue cognitive burden on the user. Existing methods for incremental verification/testing would be useful for this challenge [6, 12, 23] as would be recent progress on verification of highly non-linear ML models [16, 19, 20].

3. **Secure distributed applications with partially trusted data sources** (Long time horizon) The overarching question that we need to answer is can we build secure distributed applications when the input data is only partially trusted. This is particularly important for the significant class of systems that are stochastic and data dependent in nature. The data may be streaming, rather than at rest, adding to the challenge of verifying the data. The nodes that comprise the distributed system are heterogeneous (as argued in an earlier item) in terms of resources (including secure hardware), but also in terms of access to trustworthy data sources. Finally, the trust in data is a dynamic property, increasing say when there has been successful validation of data and decreasing when there is a detected attack.

4. **Integration of game theory and machine learning** (Long time horizon) The large unanswered question here is can we, in a principled manner, integrate game theory and machine learning to secure distributed systems [24]. In such integration, we have to be cognizant that there can be multiple players (tens to hundreds) in terms of attackers and defenders.The interactions and actions may evolve over time, which necessitates learning, rather than static

spaces for actions and rewards, as is typical today. There may also be partial information sharing among defenders, asymmetric information between attackers and defenders, and cognitive biases among players. A subset of these factors may be relevant in a specific application context, but the framework and the algorithms should be able to encompass them.

5. **Integrated evaluation environments:** (Long time horizon) The availability of an integrated evaluation environment, synonymous with testbed, would be important for the community to evaluate if we are making progress toward the hard security goals. Such testbeds across application domains will have two parts, one would provide application generic functionality and the other will be specific to the application domain. The desiderata for testbeds will be that they will allow for injection of various types of attacks, creation of different kinds of players (with the heterogeneous characteristics mentioned earlier in subsection 3.1), and measurement of a variety of metrics of interest. Such testbeds would be important, in addition to their value to demonstrate research artifacts, for educating policy makers. Such education may help inform what kinds and degrees of security investments should be made to different parts of an interconnected system to reach a desired level of security. Such evaluation environments should be configurable in several dimensions, including minimally the following.

   - Evaluate different action spaces and mechanism designs
   - Evaluate different red/blue team configurations, including partial information sharing
   - Evaluate different capabilities of attackers/defenders

4 Takeaways

The technical themes of security of distributed systems and game theory applied to security have a lot to contribute to each other. We trust that the two vibrant communities will continue the process of working together. In fact, efforts such as this should accelerate that so we can reach faster to the large societal benefits that can accrue from secure distributed systems. Many of the critical infrastructure systems that we rely on as well as personal computing systems are structured as distributed systems. Take for example, from the large end of the spectrum, transportation, power grid, and financial infrastructures, and to the personal end of the spectrum, the plethora of connected cyber physical devices that are meant to make our work lives more productive and safer and our personal lives more enjoyable and more resource efficient. As the attack surfaces for such systems become larger and the sophistication and the incentives for attacks against such systems increase, it is time to bring in rigorous reasoning to secure such systems. At the same time, the rigorous analytical foundations need to be made scalable and tractable to apply to real-world applications under realistic resource constraints and timing constraints. In this article we have laid out a set of foundations that will serve as useful starting points for our journey. We have then described a set of open research challenges that the community would hopefully take up. These are structured as advancements needed in analytical directions, systems directions, and then integration of these two directions. To guide us toward actions, we have structured the open challenges as short term time horizon, defined as 2-3 years, and long term time horizon, defined as 5-10 years. We look forward to an ongoing conversation and strides toward our goal of practical and secure distributed systems.

References

[1] Mustafa Abdallah, Parinaz Naghizadeh, Ashish R Hota, Timothy Cason, Saurabh Bagchi, and Shreyas Sundaram. Behavioral and game-theoretic security investments in interdependent systems modeled by attack graphs. *IEEE Transactions on Control of Network Systems*, 2020.

[2] Juntao Chen, Corinne Touati, and Quanyan Zhu. A dynamic game approach to strategic design of secure and resilient infrastructure network. *IEEE Transactions on Information Forensics and Security*, 15:462–474, 2019.

[3] Jin-Hee Cho, Dilli P Sharma, Hooman Alavizadeh, Seunghyun Yoon, Noam Ben-Asher, Terrence J Moore, Dong Seong Kim, Hyuk Lim, and Frederica F Nelson. Toward proactive, adaptive defense: A survey on moving target defense. *IEEE Communications Surveys & Tutorials*, 22(1):709–745, 2020.

[4] Cuong T Do, Nguyen H Tran, Choongseon Hong, Charles A Kamhoua, Kevin A Kwiat, Erik Blasch, Shaolei Ren, Niki Pissinou, and Sundaraja Sitharama Iyengar. Game theory for cyber security and privacy. *ACM Computing Surveys (CSUR)*, 50(2):1–37, 2017.

[5] Christopher N Gutierrez, Mohammed H Almeshekah, Saurabh Bagchi, and Eugene H Spafford. A hypergame analysis for ersatzpasswords. In *IFIP International Conference on ICT Systems Security and Privacy Protection*, pages 47–61. Springer, 2018.

[6] Travis Hance, Marijn Heule, Ruben Martins, and Bryan Parno. Finding invariants of distributed systems: It's a small (enough) world after all. In *18th USENIX Symposium on Networked Systems Design and Implementation (NSDI 21)*, pages 115–131, 2021.

[7] Linan Huang and Quanyan Zhu. Adaptive honeypot engagement through reinforcement learning of semi-markov decision processes. In *International conference on decision and game theory for security*, pages 196–216. Springer, 2019.

[8] Yunhan Huang and Quanyan Zhu. Deceptive reinforcement learning under adversarial manipulations on cost signals. In *International Conference on Decision and Game Theory for Security*, pages 217–237. Springer, 2019.

[9] Gunjan Khanna, Mike Yu Cheng, Padma Varadharajan, Saurabh Bagchi, Miguel P Correia, and Paulo J Veríssimo. Automated rule-based diagnosis through a distributed monitor system. *IEEE Transactions on Dependable and Secure Computing*, 4(4):266–279, 2007.

[10] Richard J La. Interdependent security with strategic agents and cascades of infection. *IEEE/ACM Transactions on Networking*, 24(3):1378–1391, 2015.

[11] Aron Laszka, Mark Felegyhazi, and Levente Buttyan. A survey of interdependent information security games. *ACM Computing Surveys (CSUR)*, 47(2):1–38, 2014.

[12] Haojun Ma, Aman Goel, Jean-Baptiste Jeannin, Manos Kapritsos, Baris Kasikci, and Karem A Sakallah. I4: incremental inference of inductive invariants for verification of distributed protocols. In *Proceedings of the 27th ACM Symposium on Operating Systems Principles*, pages 370–384, 2019.

[13] Michael Mirkin, Yan Ji, Jonathan Pang, Ariah Klages-Mundt, Ittay Eyal, and Ari Juels. Bdos: Blockchain denial-of-service. In *Proceedings of the 2020 ACM SIGSAC conference on Computer and Communications Security*, pages 601–619, 2020.

[14] Milad Nasr, Sadegh Farhang, Amir Houmansadr, and Jens Grossklags. Enemy at the gateways: Censorship-resilient proxy distribution using game theory. In *NDSS*, 2019.

[15] Thanh H Nguyen, Amulya Yadav, Branislav Bosansky, and Yu Liang. Tackling sequential attacks in security games. In *International Conference on Decision and Game Theory for Security*, pages 331–351. Springer, 2019.

[16] Kexin Pei, Yinzhi Cao, Junfeng Yang, and Suman Jana. Deepxplore: Automated whitebox testing of deep learning systems. In *proceedings of the 26th Symposium on Operating Systems Principles*, pages 1–18, 2017.

[17] Emmanouil Vasilomanolakis, Shankar Karuppayah, Max Mühlhäuser, and Mathias Fischer. Taxonomy and survey of collaborative intrusion detection. *ACM Computing Surveys (CSUR)*, 47(4):1–33, 2015.

[18] Haopei Wang, Guangliang Yang, Phakpoom Chinprutthiwong, Lei Xu, Yangyong Zhang, and Guofei Gu. Towards fine-grained network security forensics and diagnosis in the sdn era. In *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security*, pages 3–16, 2018.

[19] Shiqi Wang, Kexin Pei, Justin Whitehouse, Junfeng Yang, and Suman Jana. Efficient formal safety analysis of neural networks. *Advances in Neural Information Processing Systems*, 31, 2018.

[20] Shiqi Wang, Kexin Pei, Justin Whitehouse, Junfeng Yang, and Suman Jana. Formal security analysis of neural networks using symbolic intervals. In *27th USENIX Security Symposium (USENIX Security 18)*, pages 1599–1614, 2018.

[21] Zhiheng Xu and Jun Zhuang. A study on a sequential one-defender-n-attacker game. *Risk Analysis*, 39(6):1414–1432, 2019.

[22] G. Yan, R. Lee, A. Kent, and D. Wolpert. Towards a bayesian network game framework for evaluating ddos attacks and defense. In *Proceedings of the 2012 ACM conference on Computer and communications security (CCS)*, pages 553–566, 2012.

[23] Jianan Yao, Runzhou Tao, Ronghui Gu, Jason Nieh, Suman Jana, and Gabriel Ryan. {DistAI}:{Data-Driven} automated invariant learning for distributed protocols. In *15th USENIX Symposium on Operating Systems Design and Implementation (OSDI 21)*, pages 405–421, 2021.

[24] Mu Zhu, Ahmed H Anwar, Zelin Wan, Jin-Hee Cho, Charles A Kamhoua, and Munindar P Singh. A survey of defensive deception: Approaches using game theory and machine learning. *IEEE Communications Surveys & Tutorials*, 23(4):2460–2493, 2021.

[25] Saman A Zonouz, Himanshu Khurana, William H Sanders, and Timothy M Yardley. Rre: A game-theoretic intrusion response and recovery engine. *IEEE Transactions on Parallel and Distributed Systems*, 25(2):395–406, 2013.