

Reference ID: 11101310503_White

Reference ID: 11101310503_White

Submission Date and Time: 10/29/2019 10:47:15 AM

This contribution was submitted to the National Science Foundation in response to a Request for Information, <https://www.nsf.gov/pubs/2020/nsf20015/nsf20015.jsp>. Consideration of this contribution in NSF's planning process and any NSF-provided public accessibility of this document does not constitute approval of the content by NSF or the US Government. The opinions and views expressed herein are those of the author(s) and do not necessarily reflect those of the NSF or the US Government. The content of this submission is protected by the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (<https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode>).

Consent Statement: "I hereby agree to give the National Science Foundation (NSF) the right to use this information for the purposes stated above and to display it on a publicly available website, consistent with the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (<https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode>)."

Consent answer: I consent to NSF's use and display of the submitted information.

Author Names & Affiliations

Submitting author: Chelsea White - Georgia Tech

Additional authors: None

Contact Email Address (for NSF use only): (hidden)

Research domain(s), discipline(s)/sub-discipline(s)

stochastic control

Title of Response

Next Generation Analytics

Abstract

As the real time control of systems becomes more dependent on rapidly extracting value from an ever-increasing velocity, volume, and variety of possibly noise-corrupted real-time data for productivity improvement, decision making models that explicitly consider the role of data and when these data

become available will become more useful. This increasing dependency on real-time data for systems control will increase cyber security risk

Question 1 (maximum 400 words) – Data-Intensive Research Question(s) and Challenge(s). Describe current or emerging data-intensive/data-driven S&E research challenge(s), providing context in terms of recent research activities and standing questions in the field. NSF is particularly interested in cross-disciplinary challenges that will drive requirements for cross-disciplinary and disciplinary-agnostic data-related CI.

Of particular concern is the adversary who is capable of:

- Sensing possibly noise-corrupted data relevant for a situation assessment
- Based on these data, developing a situation assessment
- Based on this assessment, determining an action that is intended to help achieve the attacker's objectives.

We call such an adversary intelligent and adaptive, and we assume that the agent defending the system, the defender, is also intelligent and adaptive.

Question 2 (maximum 600 words) – Data-Oriented CI Needed to Address the Research Question(s) and Challenge(s). Considering the end-to-end scientific data-to-discovery (workflow) challenges, describe any limitations or absence of existing data-related CI capabilities and services, and/or specific technical and capacity advancements needed in data-related and other CI (e.g., advanced computing, data services, software infrastructure, applications, networking, cybersecurity) that must be addressed to accomplish the research question(s) and challenge(s) identified in Question 1. If possible, please also consider the required end-to-end structural, functional and performance characteristics for such CI services and capabilities. For instance, how can they respond to high levels of data heterogeneity, data integration and interoperability? To what degree can/should they be cross-disciplinary and domain-agnostic? What is required to promote ease of data discovery, publishing and access and delivery?

A mathematical model of an agent (either adversary or defender) with these characteristics requires the following functionality:

- A model of a data sensing system and a communications channel that can transmit possibly noise-corrupted data to the agent
- An inference function that can determine a situation assessment based on these data
- An action selection function that can determine an action intended to achieve the agent's objectives, based on the situation assessment.

The partially observed Markov game (POMG) is a mathematical model that assumes this functionality for both the adversary and the defender. At each of a possibly countable number of decision epochs, each agent takes newly arriving data, uses these data and the agent's knowledge of the dynamics of the system to update a situation assessment, and then selects an action intended to help achieve the agent's objectives. We remark that in general, the agents do not share the same data sensing system, communications channel, or inference function, and share the same objectives only if the game is cooperative (and if so, the two agents would no longer be referred to as an adversary and a defender). However, the decisions of each agent can affect the dynamics of the other agent's state. Computational approaches for the POMG have recently been developed. One of these approaches

develops an optimal defender policy, assuming the defender knows the adversary's policy. This assumption significantly improves the tractability of the POMG. A next step in moving toward a more realistic yet tractable scenario is to assume that the adversary will use one of a finite set of attack policies and then develop a defender policy that is reasonably effective against any attack policy in this set. We call such a policy a robust policy. Recent research focuses on developing such robust defender policies. In an effort to further improve the realism of the defender's assumptions regarding the actions an adversary might take, it would be prudent to introduce genetic algorithms and AI-based self-training ideas with the intent to model:

- The capacity to create plausible and potentially effective attacker policies not considered (or imagined) by the defender in the original set of attacker policies (using the genetic algorithm)
- The capacity of the adversary to anticipate and adjust to possible robust defender policies and consequently the capacity of the defender to adjust to an adaptable adversary in order to produce more effective robust defender policies (using self-training algorithms similar in spirit to ideas behind AlphaGo Zero and similar self-training systems).

We think that such an integrated approach represents the next generation of analytics tools to support decision-makers in sequential alternative selection tasks subject to uncertainty and risk from an intelligent and adaptive adversary.

Question 3 (maximum 300 words) – Other considerations. Please discuss any other relevant aspects, such as organization, processes, learning and workforce development, access and sustainability, that need to be addressed; or any other issues more generally that NSF should consider.

-- End Submission --