From threat to boon: understanding and controlling strategic information transmission in cyber-socio-physical systems

Cedric Langbort, UIUC, *langbort@illinois.edu*

Challenge:

As cyber-socio-physical and infrastructure systems are increasingly relying on data and integrating an ever growing range of disparate, sometimes unconventional, and possibly untrusted data sources, there is a growing need to consider the problem of estimation in the presence of strategic and/or self-interested sensors. This class "strategic information transmission" (SIT) problems differs from classical fault-tolerant estimation since the sensors are not merely failing or dysfunctioning, but are actively trying to mislead the estimator for their own benefit.

Scientific Impact:

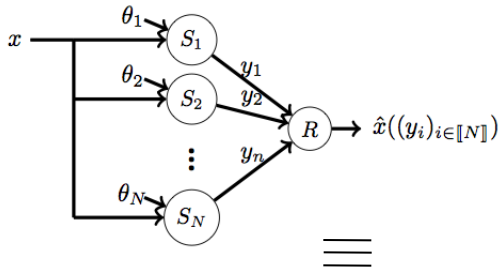
Such strategic behavior can happen through at least two mechanisms – sensor hijacking and willful misreporting– both of which have been observed in practice, e.g., in the contexts of causative attacks on Machine Learning, data injection attacks on CPS control systems, and online social engineering/participatory sensing.

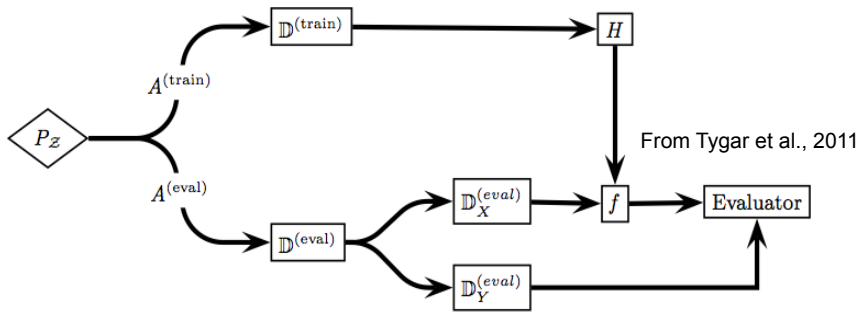
Interestingly, the presence of strategic sensors may be beneficial to a system, depending on how sensors expect others to act. Our work will help define and straddle this boundary between damageable and helpful SIT.

Solution:

We build on and extend existing models of so-called "cheap talk" and "(bayesian) persuasion" from the Information Economics literature , where strategic information transmission has been considered before under widely different assumptions.

The novelty, pertinence, and intellectual merit of this project, lie in (1) its formulation of new models that more closely account for the specificity of strategic information transmission in the three applications of interest than existing frameworks, (2) its combined use of information theoretic and game theoretic tools to analyze these models and, (3) the use of behavioral economics experiments to help characterize, and straddle, the boundary between detrimental and beneficial strategic information transmission in practice.





From Tygar et al., 2011

Causative attacks on ML can be captured as "cheap talk" problems – Illustrating the power Of the proposed modeling framework...

Broader Impact:

Beyond the three application domains mentioned above, we expect the tools and approach resulting from this project to be widely applicable, and to provide additional options for the design of trustworthy and secure data-driven cyber-sociophysical systems. Applications to Machine Learning will be pursued in collaboration with the group of Dr. Ben Rubinstein at U Melbourne in Australia.

Award Number: 1619339; 10/01/2016--09/30/2019