SaTC: CORE: Small: Collaborative: Learning Dynamic and Robust Defenses Against Co-Adaptive Spammers

Submitted by Philip Yu on Mon, 07/06/2020 - 12:20pm

Project Details

Lead PI

Philip Yu

Performance Period

Oct 01, 2019 - Sep 30, 2022

Institution(s)

University of Illinois at Chicago

Sponsor(s)

National Science Foundation

Award Number

1930941

Ranked 562 out of 2290 Group Projects.
103 related hits.

Online reputation systems are ubiquitous for customers to evaluate businesses, products, people, and organizations based on reviews from the crowd. For example, Yelp and TripAdvisor rank restaurants and hotels based on user reviews, and RateMDs allows patients to review doctors and hospitals. These systems can however be leveraged by spammers to mislead and manipulate the inexperienced customers with fake but well-disguised reviews (spams). To comprehensively protect customers and honest businesses, advanced spam detection techniques have been deployed. Nonetheless, intelligent spammers can still probe and then evolve to bypass the deployed detectors. This project investigates dynamic and robust countermeasures to defeat the evolving spammers. This research will allow regulatory agencies to enforce a more fair, transparent, and trustworthy online environment, encourage business owners to offer higher quality products and services rather than fake opinions, and ultimately, allow consumers to increasingly rely on the reputation systems confidently to save money, time and even lives.

The project will investigate the design of adaptive spam detection technologies and systems against intelligent spammers that learn to bypass static detectors. The investigation will follow two principles: (1) the goals and workings of the detectors and spammers can be sensed through their behaviors; (2) both parties should act dynamically to optimally defeat their opponents who co-adapt with the other's behaviors. Based on these principles, the researchers aim to: (i) investigate the footprint of dynamic spamming and formalize the gained insights into evasion models against static detectors; (ii) model the interactions between the evolving spammer and dynamic detections through deep reinforcement learning and Markov games; and (iii) introduce multiple cooperative spammers to inform more complex spammer-detector co-adaptations through multi-agent and hierarchical reinforcement learning. The research aims will be complemented by metrics and evaluations that capture realistic spammer and detector goals and constraints. The project will result in datasets, algorithms, and testbed system for the research community, and gamified educational software and materials to increase awareness of fake contents among a broader population.

Philip Yu

Related Artifacts

No related artifacts found.

CORE Small Collaborative Division of Computer and Network Systems (CNS)