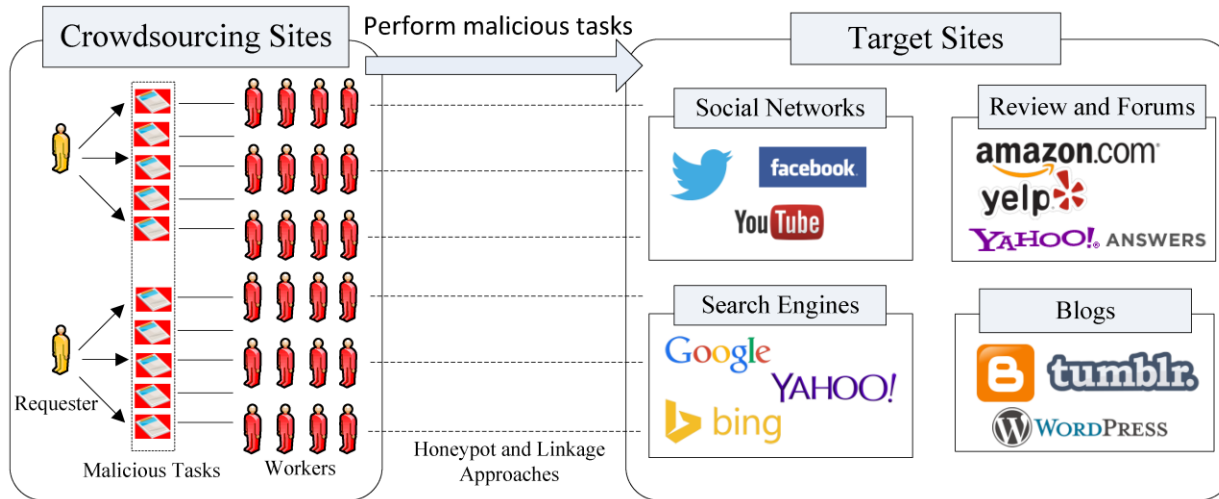


Tracking, Revealing and Detecting Crowdsourced Manipulation



Scientific Impact:

- Early detection of malicious tasks has the potential to prevent propagation of malicious tasks to popular online target sites
- Our detection methods will identify manipulated/malicious content in the target sites and help build secure online systems

Challenge:

- Widespread manipulated content over online social networks (e.g., hate speech, fake news, fake reviews)
- Tracking, revealing, detecting these manipulated content is critical to improve information trust and quality online

Solution:

- Build a task blacklist service, which automatically identifies malicious tasks
- Develop deep learning based manipulated content detection methods under both few-shot learning and supervised learning settings

Broader Impact and Broader Participation:

- The malicious task and manipulated content detection framework will enable crowdsourcing service providers and target sites service providers to detect crowdsourced manipulation with protecting information quality and trust
- So far, 6 graduate and 20 undergraduate students (5 female) have had research experience through this project
- The outcome of the project was integrated to 5 new and existing courses. We published over 30 research papers, and released code and collected data

Award Number: 1755536

Institution: Worcester Polytechnic Institute

PI: Kyumin Lee