RAISING AN ARMY: ATTACKING CROWD SYSTEMS

Walter Lasecki (wlasecki@cs.rochester.edu), University of Rochester
Ece Kamar (eckamar@microsoft.com), Microsoft Research
Jaime Teevan (teevan@microsoft.com), Microsoft Research

Crowd-powered systems have become a popular way to augment the capabilities of automated systems in real-world settings. Many crowd-powered systems rely on human workers to process potentially sensitive data or make important decisions. This puts these systems at risk of unintentionally releasing sensitive data or having their outcomes maliciously manipulated. While almost all crowd-powered approaches account for errors made by individual workers, few factor in active attacks on the system.

Some crowd systems require sharing potentially sensitive information with unknown people. For example, a photograph of a medication taken by a blind person for assistance reading the label may include personally identifying information about the user. Even if most workers are trustworthy, all it takes is one bad worker to steal that information. Little is known about how systems can prevent unintentional extraction of data when using human intelligence as a computational resource.

The success of crowd-powered systems also means that the decisions made based on the input of crowd workers are becoming increasingly critical. For example, comScore is company that provides digital analytics to some of the world's largest enterprises, agencies, and publishers. Many significant business decisions are made based on comScore data, some of which are created using Mechanical Turk. As another example, Planet Hunters (planethunters.org) uses crowd input to determine where a new planet is most likely to be found, and then uses this information to dedicate scarce telescope resources. With increased reliance on crowdsourcing to make real-world decisions, the potential for external manipulation could become a costly threat. Significant resources could be devoted to attacking crowd systems much in the way they are currently devoted to influencing search engine rankings. The search engine optimization market is estimated at $20 to $30 billion dollars in the United States alone, and poses a real challenge for search engines. As crowd systems become ubiquitous, they will likewise become targets for new types of malicious manipulative attacks, which perhaps even use the crowd itself.

We have analyzed different forms of threats from individuals and groups of workers extracting information from crowd-powered systems or manipulating these systems' outcomes. Via a set of studies performed on Amazon's Mechanical Turk platform and involving over a thousand unique workers, we demonstrated the viability of these threats. Current systems appear vulnerable to coordinated attacks on a task based on the requests of another task and that a significant portion of Mechanical Turk workers are willing to contribute to an attack. The experiments also show that workers' contribution to manipulation and extraction attacks depend on the malicious nature of these attacks. We propose several possible approaches to mitigating these threats, including leveraging workers who are willing to go above and beyond to help, automatically flagging sensitive content, and using workflows that conceal information from each individual, while still allowing the group to complete a task. Our goal is to enable the crowd to continue to play an important part in automated systems, even as the data they use and the decisions they support become increasingly important.