Crowdsourced App Review Manipulation
Shanshan Li, James Caverlee, Wei Niu, and Parisa Kaghazgaran
Computer Science and Engineering, Texas A&M University

Motivation
• Mobile app marketplaces like Google Play and Apple’s App Store serve as the nexus for many of our online experiences. A key factor driving user engagement with apps is user reviews and ratings.
• Previous research has explored collusion in app rating systems, the use of incentivized review marketplaces to attack the trustworthiness of app reviews, and found evidence of app popularity manipulation.
• In contrast, we study a complementary attack vector on the trustworthiness of app reviews: crowdsourcing platforms that allow a single manipulator to recruit a crowd of human review writers to target app reviews.
• Do crowdsourced reviews bypass review filters to actually be posted? Are reviews positive or negative? Do they actually impact the aggregate ratings of apps? Are there correlations among a developer’s apps in terms of being targeted for manipulation? Does the platform’s “related apps” feature expose users to more targeted apps?

Observations
➢ While all apps have a right-skewed rating distribution, the targeted apps are almost all rated 4.5 or 5 stars.

➢ Before being targeted, the median app rating is around 2 stars; after targeting, these same apps jump to 5 stars.

➢ We measure the timestamp standard deviation of reviews on targeted apps before and after they were targeted. The “after” distribution is left-skewed indicating that crowdsourced reviews tend to be posted in bursts.

➢ Self-similarity: The reviewers of targeted apps tend to mimic themselves by using repeated terms and phrases in their reviews.

➢ Apps sharing same developers with targeted apps are more likely to be targeted for promotion.

Overall Approach

Findings
• We find that apps targeted by crowdsourcing platforms are rated significantly higher on average than other apps, indicating that app manipulation is focused on app promotion, rather than in punishing the apps of competitors.
• The reviews themselves arrive in bursts, and have an immediate positive impact on the average ratings of the apps.
• The patterns of linguistic evolution suggest that app reviewers tend to repeat themselves by relying on some standard repeated text.
• Apps by the same developer tend to share a more similar language model: if one app has been targeted, it is likely that many of the other apps from the same developer have also been targeted; in contrast, we find the platform’s “related apps” feature tends to identify apps that have not been targeted.

Future Work
• Expanding our collection of targeted apps to include more App Store apps as well as apps on the Google Play store.
• Exploring machine learning models for identifying hidden targeted apps, reviews, and reviewers.
• Creating evolutionary user models to capture more fine-grained changes in reviewer styles over time toward automatically detecting reviewers who engage in occasional review-for-pay jobs.

In total, we collected 50,461 reviews on 100+ targeted apps and 142,400 reviews on 450+ random apps.