Mass scale manipulation of Twitter Trends discovered

June 2 2021, by Tanya Petersen

Credit: CC0 Public Domain

New EPFL research has found that almost half of local Twitter trending topics in Turkey are fake, a scale of manipulation previously unheard of. It also proves for the first time that many trends are created solely by bots due to a vulnerability in Twitter's Trends algorithm.

Social media has become ubiquitous in our modern, daily lives. It has

changed the way that people interact, connecting us in previously unimaginable ways. Yet, where once our social media networks probably consisted of a small circle of friends most of us are now part of much larger communities that can influence what we read, do, and even think.

One influencing mechanism, for example, is "Twitter Trends." The platform uses an algorithm to determine hashtag-driven topics that become popular at a given point in time, alerting twitter users to the top words, phrases, subjects and popular hashtags globally and locally.

Now, new EPFL research focused on Turkey, from the Distributed Information Systems Laboratory, part of the School of Computer and Communication Sciences has found a vulnerability in the algorithm that decides Twitter Trending Topics: it does not take deletions into account. This allows attackers to push the trends they want to the top of Twitter Trends despite deleting their tweets which contain the candidate trend shortly afterwards.

"We found that attackers employ both fake and compromised accounts, which are the accounts of regular people with stolen credentials, or who installed a malicious app on their phones. Generally, they are not aware that their account is being used as a bot to manipulate trending topics, sometimes they are but don't know what to do about it and in both cases they keep using Twitter," said Tuğrulcan Elmas, one of the authors of the research, accepted by the IEEE European Symposium of Security and Privacy 2021, a top cybersecurity conference.

"We found that 47% of local trends in Turkey and 20% of global trends are fake, created from scratch by bots. Between June 2015 and September 2019, we uncovered 108,000 bot accounts involved, the biggest bot dataset reported in a single paper. Our research is the first to uncover the manipulation of Twitter Trends at this scale," Elmas continued.

The fake trends discovered include phishing apps, gambling promotions, disinformation campaigns, political slogans, hate speech against vulnerable populations and even marriage proposals.

"This manipulation has serious implications because we know that Twitter Trends get attention. Broader media outlets report on trends, which are used as a proxy for what people are talking about, but unfortunately, it's a manipulated proxy, distorting the public view of what conversations are actually going on," said Rebekah Overdorf, another of the paper's authors. "For example, one of the manipulated hashtags that we found that was pushed to Trends artificially was #SuriyelilerDefolsun, translated to 'Syrians get out,' and this was then picked up by several news reports, other social media platforms and in academic papers. In reality, it was completely fabricated," Overdorf continued.

The researchers contacted Twitter twice, with the company acknowledging in both cases the vulnerability in its Trends algorithm. In the first case Twitter declined to make any changes, in the second case the company did not respond to the researchers' follow-up emails. "The problem has not been fixed and we still see obvious spam trends occurring. It's clear that until the vulnerability in the algorithm is corrected adversaries will continue to create fake trends with the same attack methodology," concluded Elmas.

More information: Ephemeral Astroturfing Attacks: The Case of Fake Twitter Trends. arXiv:1910.07783v4 [cs.CR] arxiv.org/abs/1910.07783

Provided by Ecole Polytechnique Federale de Lausanne