A model to determine the impact of DDoS attacks using Twitter data

October 3 2019, by Ingrid Fadelli

An example of tweets that inspired the research study. Credit: Zhang et al.

Distributed denial of service (DDoS) attacks, which are designed to prevent legitimate users from accessing specific network systems, have become increasingly common over the past decade or so. These attacks make services such as Facebook, Reddit and online banking sites extremely slow or impossible to use by exhausting network or server resources (e.g., bandwidth, CPU and memory).

Researchers worldwide have been trying to develop techniques to prevent DDoS attacks or rapidly intervene in order to reduce their negative effects. An important step in counteracting such attacks is the prompt collection of feedback from users to determine their impact and come up with targeted solutions.

With this in mind, a team of researchers at the University of Maryland have developed a machine-learning model that could help to determine the scale of impact of DoS attacks as they are happening based on tweets posted by users. Their study, recently pre-published on arXiv, was funded by a UMBC-USNA Cyber Innovation Grant.

"The research was based on the observation that when there are difficulties in accessing network services, customers sometimes share that information the social networks," Dr. Tim Oates, one of the researchers who carried out the study, told TechXplore. "Our main objective was to develop a system that tracks network denial-of-service (DoS) attacks by analyzing their ripple effects through social media posts."

To begin with, Dr. Oates and his colleagues collected a curated set of tweets about DoS attacks based on a historical timeline of attacks that occurred in the past. Looking at these tweets, in which users described the problems they were experiencing during an attack, the researchers were able to identify 'language patterns' (i.e., relevant keywords). They then trained a decision-tree classifier to detect DDoS attacks based on these keywords.

"We hypothesized that impacted customers use similar language on social media to describe problems during a DDoS attack such as the system or product being slow or crawling," Chi Zhang, another researcher involved in the study told TechXplore. "Thus, when new tweets are collected (historically or in real-time), the model first finds

out the topics (a set of keywords that broadly define an area of discussion) of the tweets collected in that time window."

Subsequently, the classifier developed by Dr. Oates, Zhang and their colleagues ranks the tweets based on how much the keywords differed from language patterns observed in user posts during past DDoS attacks. Finally, the model uses the number of detected DDos-related tweets to compute the scale of impact of an attack.

When the researchers evaluated their model, they found that it achieved similar results to supervised state-of-the-art approaches to determine the scale of DDoS attacks. A great advantage of their classifier, however, is that it is weakly supervised, thus it requires very little human labeling of training data.

"We were able to develop a weakly supervised model for new event detection that performs nearly as well as supervised models," Zhang said. "Its weakly supervised nature means that only a small amount of human labeled data is needed, thus it saves a lot of resources in terms of human labor, as asking people to annotate potentially thousands of Tweets is typically quite expensive."

In the future, their weakly supervised model could help to determine the scale of DDoS attacks rapidly and more effectively, solely based on Twitter data. It could also be adapted and applied to other tasks that might benefit from the analysis of user tweets in real-time.

In their next studies, the researchers plan to develop their model further in order to analyze tweets written in other languages. Eventually, they would also like to change its classification layer to test its performance in determining the scale of impact of other types of events, such as disease outbreaks (e.g., Ebola).

"We realized that people have many ways to describe problems on Twitter," Ashwinkumar Ganesan, another researcher who carried out the study, told TechXplore. "Hence, there is a need to build a larger cache of tweets and better models that handle this variation in language. In addition, attacks are not restricted to targets in the English speaking world, so designing the system so it can be scaled to other languages is very important too."

More information: Determining the scale of impact from denial-of-service attacks in real time using Twitter. arXiv:1909.05890 [cs.SI]. arxiv.org/abs/1909.05890