Collecting Contextual Information About a DDoS Attack Event Using Google Alerts

Abhishta Abhishta
s.abhishta@utwente.nl

Reinoud Joosten
r.a.m.joosten@utwente.nl

Mattjs Jonker
m.jonker@utwente.nl

Wim Kamerman
info@wimkamerman.com

Lambert J. M. Nieuwenhuis
l.j.m.nieuwenhuis@utwente.nl

* Industrial Engineering and Business Information Systems (IEBIS), Faculty of Behavioural, Management and Social Sciences (BMS), University of Twente, Enschede, The Netherlands.
† Design and Analysis of Communication Systems (DACS), Faculty of Electrical Engineering, Mathematics and Computer Science, University of Twente, Enschede, The Netherlands.

Motivation:
Distributed Denial of Service (DDoS) attacks can lead to massive economic damages to victims. In most cases, the damage caused is dictated by the circumstances surrounding the attack (i.e. context). One of the ways of collecting information on the context of an attack can be by using the online articles written about the attack.

Goals:
- Introduce a dataset collected using Google Alerts that provides contextual information related to DDoS attacks.
- Invite other researchers for collaboration.

Data Collection:
- **Step 1:** Export the emails and store them in a local file storage.
- **Step 2:** Scrape the text from the emails using `mailbox` package (Python) and extract the following features from the alert using regular expressions: 1) Alert Header, 2) Associated Text, 3) Type of Alert (News or Web).
Then we filter the duplicate alerts as the same alert may be reported by both the triggers.
- **Step 3:** Introduce two additional features to the dataset: 1) the language of the alert, 2) the historical alexa rank of the source of the alert.
- **Step 4:** Store all data in a relational database.

Data Sharing & Future Work:
- We are working on a web portal in order to make the dataset public. We will also share all the scripts used for scraping and preparation of the data. In near future we plan to build an algorithm to label and track articles belonging to a single attack event.

Analysis:
- Analyse the metadata of the articles related to four major DDoS attack events within first 20 days following an attack.
- Fig. 1 shows the number of articles related to each of the attacks within 20 days of the attack.
- We use a machine learning algorithm to classify articles reporting a DDoS attack. Fig. 2 clearly shows that we are able to remove all noise from our dataset (there are no attack reporting articles after the attack day).

Observations:
- We observe that we record a relatively large number of articles just after the attack day. This proves that we are able to successfully track articles reporting DDoS attack using our data collection strategy.
- The fact that more articles discussed the attack on Pokemon than attack on OVH shows that the popularity of an attack on web forums is not proportional to the intensity of an attack.