Poster: Collecting Contextual Information About a DDoS Attack Event Using Google Alerts

Abhishta Abhishta* University of Twente
Reinoud Joosten University of Twente
Mattijs Jonker University of Twente
Wim Kamerman University of Twente
Lambert J.M. Nieuwenhuis University of Twente

***Abstract*—Distributed Denial of Service (DDoS) attacks may 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. In this poster, we introduce a dataset collected using *Google Alerts* that provides contextual information related DDoS attacks. The goal of the poster is to invite other researchers for collaboration.**

***Index Terms*—Data Mining, Contextual Information, DDoS Attacks, Social Impact, Google Alerts.**

I. Introduction

Distributed Denial of Service (DDoS) attacks lead to unavailability of network based resources that can cause economic damages for user(s) of attacked infrastructure. Several reports by DDoS protection companies estimate the yearly costs of this unavailability in millions of dollars [1]. These costs are usually computed using a simple linear metric or a victim survey which estimates the total damages based on subjective measures such as lost revenue, brand damage and operational cost etc. [2]. However, recent studies indicate that the average damages might not be as high as claimed by these reports. Florencio and Herley [3] find evidence that most cybercrime surveys are dominated by a minority of responses in the upper tail which leads to over estimation of losses. With relation to brand damage, several studies have shown that DDoS attacks have very short lived impact on stock prices. The stock prices tend to recover within 5-10 days of an attack. Studies have also shown that a number of network based businesses are resilient to relatively short DDoS attacks [4, 5, 6]. This is either because they are designed to be resilient or because the economic returns to the users of the service are not affected by the downtime.

The variation in the reported impact of DDoS attacks can be due to fact that these estimates do not take into account the complete *context* of an attack. *Context* is defined as *the circumstances that form the setting for an event*. As DDoS attacks only affect the availability of a network infrastructure unlike any other cyber attacks (e.g. attacks that target the confidentiality and integrity), the circumstances surrounding the attack event may dictate the consequences. In order to gather information on the *context* of an attack, we would need to interview the victims. Journalists working in the technology sector also perform such interviews. With advent of online media outlets most news is available on the web and can be used to gather contextual information on DDoS attacks. Services like *Google Alerts* can be used to collect such news articles. In this poster, we introduce a dataset collected with the help of *Google Alerts* that can provide *contextual* information on a publicly reported DDoS attack.

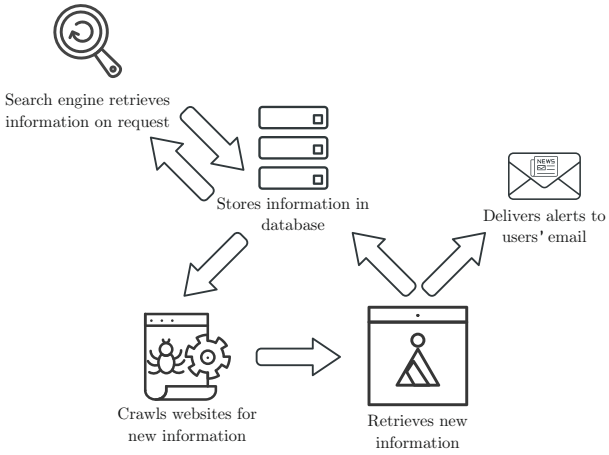


Fig. 1: Generation of 'Google Alerts'.

II. Google Alerts

In 2003, Google started a service named *Google Alerts* for helping users to keep themselves up to date on topics of their choice. The web crawlers of Google continuously search the world wide web for new content. When the crawler finds a new web page or change in content of an old web page, it stores the information in a database for quick response. Google Alerts is a content change detection and notification service. The alerts are delivered via emails to the user when it finds new results, such as web pages, newspaper articles, blogs, or scientific research that match the user's search term(s). If a user has registered alerts related to a topic (trigger word or phrase), then it is possible for Google to notify a user about the new content. Fig. 1 shows the high level working of a search engine and the process of generation of 'Google Alerts'.

III. Methodology

A. Data Collection

Using the *Google Alerts* service we collect articles on DDoS attacks. We do this by subscribing alerts on two trigger words:

*For correspondence please contact s.abhishta@utwente.nl

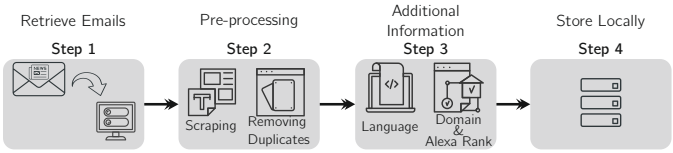


Fig. 2: Data collection and processing steps.

	#Articles Tagged as			
Year	News	Web	# Domains	# Languages
2015	1427	3653	2467	37
2016	4458	9387	4889	42
2017	5805	9658	5692	44
2018	5230	7005	5071	45

TABLE I: Characteristics of the dataset.

1) 'denial of service' 2) 'ddos'. We start collecting this data since 20th August 2015.

B. Pre-processing and Analysis

Fig. 2 gives an overview of the collection of data. To collect the data we take the following steps:

- **Step 1**: Firstly, we download the alert emails from the server and store them in a local file storage for further processing.
- **Step 2**: In the pre-processing step, we scrape the text from the emails using the *mailbox* package† in Python and extract the following features for each article in an alert using *regular expressions*:
  - Article Header
  - Associated Text
  - Type of Article (News or Web)‡

  Then we filter the duplicate articles, as the same article may be reported by both the triggers and proceed to step 3.
- **Step 3:** In this step, we introduce two additional features to the dataset based on the language§ of the article and the historical alexa rank of the source (domain) of the article.
- **Step 4:** Finally, we store all data in a relational database.

Table I shows the characteristics of the data collected between 20th of August 2015 and 31st of December 2018.

IV. Case Study: Tracking Articles on DDoS Attack Events

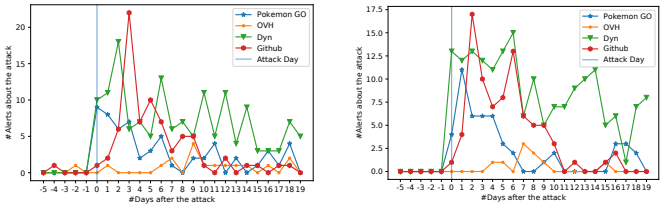
As a case study, we analyse the metadata of the articles related to four major DDoS attack events within first 20 days of an attack. Using regular expression based word search we calculate the number of articles on attacks on *Pokemon Go*, *OVH*, *Dyn* and *Github*. Fig. 3a shows the number of articles related to each of the attacks within 20 days of the attack.

We observe that we record a relatively large number of articles just after the attack day. This proves that we are able

†https://docs.python.org/2/library/mailbox.html

‡Google tags the articles depending on the source of information.

§https://pypi.org/project/googletrans/



(a) Before **ML** based filter. (b) After **ML** based filter.

Fig. 3: #Articles on DDoS Attack Events.

to successfully track articles reporting DDoS attack using our data collection strategy. To provide further proof that these alerts are reporting a DDoS attack we use a machine learning classification algorithm to extract attack reporting articles. With the help of Fig. 3b we can clearly see that we are able to remove all noise from our dataset as there are no attack reporting articles before the attack day. The fact that more articles discussed the attack on *Pokemon Go* than attack on *OVH* shows that the popularity of an attack on web forums is not proportional to the intensity of an attack.

V. Concluding Remarks and Future Works

In this poster, we present a dataset that provides *contextual* information related to DDoS attacks. With the help of a simple case study we prove that the dataset is useful in tracking articles related to DDoS attacks. Currently, we update our dataset on a weekly basis and plan to openly publish the data on a website soon.

The goal of this poster is to invite other researchers for collaboration and to inform them about the dataset. Furthermore, we have already started using this data in several of our projects where contextual information about a DDoS attack event could improve our analysis.

References

[1] *Trends in the Cost of Web Application & Denial of Service Attacks*. URL: https://content.akamai.com/us-en-pg10029-ponemon-cost-of-ddos-web-app-report.html.

[2] *Calculate the Cost of DDoS Attacks*. URL: https://www.akamai.com/uk/en/products/security/calculate-the-cost-of-ddos-attacks.jsp.

[3] Dinei Florencio and Cormac Herley. *Sex, Lies and Cyber-crime Surveys*. June 2011.

[4] Abhishta Abhishta, Reinoud Joosten, Sergey Dragomiretskiy, and Lambert J.M. Nieuwenhuis. "Impact of Successful DDoS Attacks on a Major Crypto-currency Exchange". In: *2019 27th Euromicro International Conference on Parallel, Distributed and Network-based Processing (PDP)*. IEEE, Feb. 2019, pp. 379–384. ISBN: 978-1-7281-1644-0.

[5] Abhishta Abhishta, Roland van Rijswijk-Deij, and Lambert J.M. Nieuwenhuis. "Measuring the Impact of a Successful DDoS Attack on the Customer Behaviour of Managed DNS Service Providers". In: *Computer communication review* 48.5 (Oct. 2018), pp. 70–76.

[6] Giovane Moura, John Heidemann, Moritz Müller, Ricardo de O Schmidt, and Marco Davids. "When the Dike Breaks: Dissecting DNS Defenses During DDoS". In: *Proceedings of the Internet Measurement Conference 2018*. ACM. 2018, pp. 8–21.