Computer Communications 103 (2017) 18–28

Contents lists available at ScienceDirect

Computer Communications

journal homepage: www.elsevier.com/locate/comcom

Review

Detection of DDoS attacks and flash events using information theory metrics–An empirical investigation

Sunny Behal[a,*], Krishan Kumar[b]

[a] Research Scholar, I.K.G. Punjab Technical University, Kapurthala, Punjab, India
[b] Department of CSE, Shaheed Bhagat Singh State Technical Campus, Punjab, India

ARTICLE INFO

Article history:
Received 1 August 2016
Revised 8 February 2017
Accepted 10 February 2017
Available online 12 February 2017

Keywords:
Network security
DDoS attacks
Flash events
Entropy
Information distance

ABSTRACT

A Distributed Denial of Service (DDoS) attack is an austere menace to extensively used Internet-based services. The in-time detection of DDoS attacks poses a tough challenge to network security. Revealing a low-rate DDoS (LR-DDoS) attack is comparatively more difficult in modern high speed networks, since it can easily conceal itself due to its similarity with legitimate traffic, and so eluding current anomaly based detection methods. This paper investigates the aptness and impetus of the information theory-based generalized entropy (GE) and generalized information distance (GID) metrics in detecting different types of DDoS attacks. The results of GE and GID metrics are compared with Shannon entropy and other popular information divergence measures. In addition, the feasibility of using these metrics in discriminating a high-rate DDoS (HR-DDoS) attack from a similar looking legitimate flash event (FE) is also verified. We used real and synthetically generated datasets to elucidate the efficiency and effectiveness of the proposed detection scheme in detecting different types of DDoS attacks and FEs. The results clearly show that the GE and GID metrics perform well in comparison with other metrics and have reduced false positive rate (FPR).

1. Introduction

DDoS attacks have been in existence for many years. Legitimate users are deprived of using web-based services due to such attacks. Typically, a DDoS attack is launched in a coordinated manner by compromising hundreds of computer systems available freely on the Internet [1]. DDoS attacks deny the target service by sending the redundant stream of packets to a victim rendering it unavailable to legitimate clients. Usually prominent websites are the prime victims of such attacks. Recently Twitter, Spotify, and Amazon suffered interruptions in their services for almost two hours on Oct 21, 2016 because of DDoS attacks. Such interruptions in the services lead to huge financial losses. The revenue loss due to DDoS attacks has touched to $209 million in the first quarter of 2016, compared to $24 million for all of 2015 [2]. As per the worldwide infrastructure security report (WISR) [3], the volume of DDoS attack traffic has increased to around 600 Gbps in 2016.

Primarily, there are two types of DDoS attack detection methods in existence (a) the signature-based detection methods which works on the basis of already stored attack signatures that match a known pattern with the pattern of incoming packets, and (b) the anomaly-based detection methods which compare the pre-built network behavior model with the incoming network behavior in real-time. Anomaly-based detection has some inherent limitations. Firstly, sophisticated attackers can monitor the network traffic to train their detection systems. Secondly, the difficulty in setting up an optimal threshold leads to an increase in false positive rate. Thirdly, it is very difficult to extract both qualitatively and precisely appropriate features of legitimate and anomalous network behavior. On the other hand, signature based detection methods require updated signatures for their efficient working [4]. Based on the traffic rate, the DDoS attacks can be categorized into (a) high-rate DDoS (HR-DDoS) attacks, when the traffic rate is very different from the legitimate traffic, and (b) low-rate DDoS (LR-DDoS) attack, when traffic rate is similar or less than the legitimate traffic [5]. However, it is comparatively easy to detect HR-DDoS attacks as their traffic profile significantly deviates from the legitimate traffic profile [6]. As per [7], sophisticated attackers have shifted their focus to carrying out more subtle and stealthy DDoS attacks that are more difficult to detect and can easily evade the traditional anomaly based detection deployments.

Apart from the detection of DDoS attacks, there is a another kind of network traffic which is gaining popularity among security researchers, and which also causes a denial of service to legitimate users of a web service, that is, a flash event (FE). An FE is similar

[*] Corresponding Author:
E-mail address: sunnybehal.sbs@gmail.com (S. Behal).

http://dx.doi.org/10.1016/j.comcom.2017.02.003
0140-3664/© 2017 Elsevier B.V. All rights reserved.

to a HR-DDoS attack wherein thousands of legitimate users try to access a particular computing resource such as a website simultaneously [8]. This sudden surge in legitimate traffic is mainly due to some breaking news happening around the world like the publishing of an Olympic schedule or the launching of a new product by companies like Apple, Samsung, etc. It causes untimely delivery of responses from a web service, and thus requires immediate action. A recent example of such an event occurred against the Australian census website on August 21, 2016. Millions of users simultaneously accessed the census website to fill their personal details. The lack of sufficient resources on the web server caused the website to crash down. It is interesting to note that the worldwide DDoS attack capturing agency Arbor networks say it wasn't a DDoS attack, but more likely an FE whereas the census officials pretended that it was a series of DDoS attacks [9]. Such situations highlight the severity of the problem. Both HR-DDoS attacks and FEs share many common characteristics like a change in the rate of traffic volume, delay in responses from the webserver, etc. but still there a few parametric differences between them. The request rate per source IP is smaller in FEs than in HR-DDoS attacks. The similarity of network flows, less throughput, and more duration of continuous traffic per source IP are some of the key rationales that can differentiate HR-DDoS attacks from FEs [8].

Both DDoS attacks and FEs cause a significant deviation in the packet header features of the network traffic. The information theory-based detection metrics such as entropy or information divergence can quickly capture such variations in the network traffic behavior. There are many key advantages of using information theory-based solutions as compared to the other methods. They require fewer packet header features to characterize the different types of network traffic. They usually have small time, space, and computational complexity as only packet header information is used for calculation. They have fewer storage requirements, so there is no need to accumulate huge network traces [6,10]. For analyzing one traffic sample in a time interval T with a total of n samples per time window, the time complexity of information theory-based detection metrics is linear i.e. O(T_n). This means that even if we perform multi-variate analysis i.e. analyze multiple packet header features simultaneously, this will not affect the overall time complexity of the information theory-based solutions.

Several research efforts have been conducted in isolation to detect DDoS attacks and FEs as mentioned in section 2.2 but none of the researchers have attempted to devise a common methodology to detect them collectively. In this paper, we have extended the idea of [10] to use GE and GID metrics to detect different types of DDoS attacks and FEs collectively. The major contributions of this paper are:

- It investigates and highlights the preeminence of GE and GID metrics in the detection of DDoS attacks.
- It proposes the use of GE and GID metrics to discriminate HR-DDoS attacks from FEs.
- The GID metric is shown to compare favorably with other popular information divergence measures.
- The proposed detection methodology is generalized and hence can detect future attacks and FE events.

The rest of the paper is organized as follows. Section 2 describes the background of information theory metrics and related work, Section 3 focuses on the experimental setup, Section 4 describes the methodology used, Section 5 summarizes the results obtained and the concluding remarks are given in Section 6 along with scope for future work.

2. Background and related work

Information theory-based detection metrics are extensively used in the anomaly based DDoS attack detection systems. Shannon entropy and Kullback–Leibler divergence (also known as relative entropy) are the two most fundamental detection metrics in information theory.

2.1. Background of GE and GID metrics

Claude Shannon in 1948 defined entropy to measure the uncertainty, disorder or randomness in a physical system. It can also represent the amount of information gained by the observations of disordered systems. Formerly, Shannon entropy (Sh_E) [11] is given by:

$$H(x) = -\sum_{i=1}^{n} p_i log_2 p_i \tag{1}$$

where p_i is the probability of the occurrence of an event x. Subsequently, Alfred Renyi gave the more general definition of Sh_E called a generalized information entropy (GE) of order α (also called α-Entropy or Renyi's α Entropy) [12] and is defined as follows:

$$H_\alpha(x) = \frac{1}{1-\alpha} log_2 \left(\sum_{i=1}^{n} p_i^\alpha \right) \tag{2}$$

where p_i are the probabilities of the events $\{x_1, x_2, \dots \dots x_n\}$, $p_i \geq 0$

The GE metric has the capability to highlight the different contributions of the tail and the main proportion of the probability distributions. The GE metric measures these contributions by making use of the powers of α parameter. For $\alpha \geq 0$, GE metric is more sensitive to the frequent occurring events whereas for $\alpha < 0$, GE metric is more sensitive to the less frequent events. By changing the value of α-order, the different types of entropies can be derived. For example, when $\alpha = 0$, the maximum value of information entropy is reached also known as Hartley entropy. It is defined as:

$$H_0(x) = log_2 n \tag{3}$$

When $\alpha \longrightarrow 1$, the Shannon entropy is derived as follows:

$$H_1(x) = -\sum_{i=1}^{n} p_i log_2 p_i \tag{4}$$

If $\alpha = 2$, the collision entropy or Renyi's quadratic entropy is derived. This type of entropy is very popular and has found its applications in physics, signal processing and economics. When $\alpha \longrightarrow \infty$, minimum information entropy H_∞ (x) is reached.

There are a plethora of divergence metrics that can be used to quantify the difference between a set of probability distributions. For any two discrete probability distributions $P = (p_1, p_2, \dots \dots ,p_n)$ and $Q = (q_1, q_2, \dots \dots ,q_n)$ with $\sum_{i=1}^{n} p_i = \sum_{i=1}^{n} q_i = 1$, i = 1,2,... .,n, the information divergence is given as:

$$D_\alpha(P \parallel Q) = \frac{1}{1-\alpha} log_2 \left(\sum_{i=1}^{n} p_i^\alpha q_i^{1-\alpha} \right), \alpha \geq 0. \tag{5}$$

Based on the different order of α, the following useful formulas can be derived:

$$D_0(P \parallel Q) = log_2 \left(\sum_{i=1}^{n} q_i \right), \alpha = 0. \tag{6}$$

$$D_1(P \parallel Q) = \sum_{i=1}^{n} p_i log_2 (\frac{p_i}{q_i}), \alpha \to 1. \tag{7}$$

which is the Kullbeck–Leibler divergence [13].