Experiments of DDoS detect using Triangle Expectation with MapReduce

Jin-Hyun Yoon, Ho-Seok Kang and Sung-Ryul Kim

1 Division of Internet and Media, Konkuk University, Seoul, Republic of Korea
h2jhyoon@gmail.com, hsriver@gmail.com, kimsr@konkuk.ac.kr

Abstract. With the growing use of the Internet, Internet security has become important. Increasing damage has been done by Distributed Denial-of-Service attacks (DDoS). In this paper, a technique called "triangle expectation" is used, which works to find the sources of attack so that they can be identified and blocked. To analyze a large amount of collected network connection data, a sampling technique has been used and the proposed technique is verified by experiments.

Keywords: Triangle Expectation, Triangle Counting, Hadoop, Sampling

1 Introduction

With increasing use of the Internet, Internet attacks are on the rise. Distributed Denial-of-Service (DDoS) in particular is increasing more [3]. There are four main ways to protect against DDoS attacks: attack prevention, attack detection, attack source identification, and attack reaction. The protection approach proposed in this paper belongs to the attack source identification category.

Most DDoS attacks work by the following three components: the first attacker; the zombie PC; and the victim. Assume that a graph is constructed so that the nodes represent various computers on the Internet and the edges represent having histories of connections between the computers. If there are three nodes, which are connected with two edges, and another edge is needed in order to complete a triangle, then the situation is called a triangle expectation. Triangle expectation is the technique used to find this expected edge. The nodes of the triangle expectations found can be considered the attacker, the zombie PC, and the victim. For performance reasons, we used the sampling method of DOULION [1], which is implemented with MapReduce. We verify that the triangle expectation technique can identify attack patterns and also that the sampling strategy does not reduce the accuracy of triangle expectation.

2 Using Triangle Expectation defense mechanisms

In DDoS attacks zombie PCs connect to the server or router to cause an overload or to starve the server of resources and make service impossible. Illustrated in Fig.1 (a), the
hacker uses the Command and Control (C&C) server [5] to send attacking command to zombie PCs, uses them to attack the victim. Thus there has to be many triangle expectation present between the C&C Server and the victim. Our system works by collecting network connection information from many routers for some time period. We can see that when an attack actually occurs, a large number of triangle expectations are generated between the C&C server and the victim.

![Fig. 1. (a) Botnet Attack (b) Triangle Expectation of relationship](image)

After the triangle expectations are computed, the zombie PCs are easily identified and blocked. Further, the C&C server is also identified so that it may be confiscated for further investigation.

For performance reasons, we have used the sampling method of DOULION. DOULION chooses each edge with probability $p$ and the total amount of computation reduced by a factor of $1/p^3$. We will later verify that the reduced computation does not result in the reduction of correctness of our results.

3 Experiment Result

The experiment was conducted in two ways to verify the triangle expectation technique and also to verify that the sampling technique is justified. We combined random connection patterns (RCP) and simulated DDoS connection patterns (SDCP) as one dataset. In RCP the nodes connect to generally popular sites (approximately 120,000 edges). SDCP are created using DDoS attacks patterns (2,000 edges). We combined the two datasets together. The experiments are performed using Map-Reduce (in Fig2) [4]. The actual algorithm is very similar to that of DOULION and is omitted in this version.

![Fig. 2. Map-Reduce](image)

The result is shown in Fig. 3(a). In the figure, the x-axis is the number of triangle expectations that are shared by one edge, and the y-axis is the number of occurrences of such edges. We can see that when the number of triangle expectation increases, the
number of edges decreases naturally. When the value of the triangle expectation is between 213 and 1,000, there is no edge shared by that many triangle expectations but at 1,000 we have found two edges that are shared by that many triangle expectation, which is an indication that the two nodes connected by any of the edges are the C&C server and the victim. Fig. 3 (b) just focuses on the important part of the graph in (a).

**Fig. 3.** Number of edge occurred Triangle Expectation (a) Total View (b) Subset of (a) between 88 and 1,000 (c) Sampling view

Fig. 3(c) shows the result after the sampling technique is applied. We can clearly see that the edges connecting the C&C server and the victim can still be easily identified. A linear-time computation, which takes much less time than the triangle expectation, will also find the zombie PCs easily.

## 4 Conclusion

In this paper, we proposed a new method of DDOS detection using triangle expectation method. When a DDoS attack occurs, data are collected from many routers, and many triangle expectations are found for many sets of nodes, in order to find the attacker and the attack path. To analyze a large amount of collected data, sampling is used, reducing the amount of calculations needed and the time taken. The results are verified by experiments.

**Acknowledgments.** This research was supported by the MKE(Ministry of Knowledge Economy), Korea, under the ITRC(Information Technology Research Center) support program supervised by the NIPA(National IT Industry Promotion Agency) (NIPA-2012-H0301-12-3006)

## 5 References