TViS: A Light-weight Traffic Visualization System for DDoS Detection

Abhishek Kalwar
Dept. of Comp. Sc. & Engg.
Assam Kaziranga University
Jorhat 785006, India
akabhishak554@outlook.com

Monwar H. Bhuyan*
Laboratory for Cyber Resilience, NAIST
Nara 630 0192, Japan
Dept. of Computing Science
Umeå University, Sweden
monowar@cs.umu.se

Dhruba K. Bhattacharyya
Dept. of Comp. Sc. & Engg.
Tezpur University
Assam 784028, India
dkb@tezu.ernet.in

Youki Kadobayashi
Laboratory for Cyber Resilience, NAIST
Nara 630 0192, Japan
youki-k@is.naist.jp

Erik Elmroth
Department of Computing Science
Umeå University, Sweden
elmroth@cs.umu.se

Jugal K. Kalita
Department of Computer Science
University of Colorado, USA
jkalita@uccs.edu

Abstract—With rapid growth of network size and complexity, network defenders are facing more challenges in protecting networked computers and other devices from acute attacks. Traffic visualization is an essential element in an anomaly detection system for visual observations and detection of distributed DoS attacks. This paper presents an interactive visualization system called TViS, proposed to detect both low-rate and high-rate DDoS attacks using Heron’s triangle-area mapping. TViS allows network defenders to identify and investigate anomalies in internal and external network traffic at both online and offline modes. We model the network traffic as an undirected graph and compute triangle-area map based on incidences at each vertex for each 5 seconds time window. The system triggers an alarm if the system finds an area of the mapped triangle beyond the dynamic threshold. TViS performs well for both low-rate and high-rate DDoS detection in comparison to its competitors.

Index Terms—DDoS attack; visualization; network traffic; online and offline; triangle-area;

I. INTRODUCTION

Network systems are becoming more complex rapidly with the proliferation of connected devices in terms of size, topology, and speed. [1]. Simultaneously, the number of network attacks against each host has increased exponentially. These attacks often conceal the vast amount of legitimate and seemingly random traffic. A Denial-of-Service (DoS) attack attempts to make machines or network resources unavailable to its intended users either temporarily or indefinitely, interrupting or suspending services of a host connected to the internet.

Moreover, Distributed DoS (DDoS) attacks are a combination of DoS attacks where attacks are generated by a large number of hosts. These hosts might be amplifiers or reflectors, or even might be zombies. They usually send the traffic to the target or victim host through the reflectors. Early DDoS attacks in 2000 targeted at well-known websites such as CNN, Amazon, and Yahoo, stopped normal services of these victims for hours [2], [3]. A new form of Mirai botnet based threats hide in the Tor network that attempts to compromise legitimate users.

Most existing network defense techniques and tools still heavily rely on security analysts (SA). The security analyst performs manual analysis to detect and trigger actions against network attacks. Network traffic visualization has become more critical in recent years to speed up the attack detection process through visual analytics. Malicious activities, such as DDoS attacks are relatively easy to implement and somewhat hard to prevent. However, detection of DDoS attack requires the processing of vast amounts meta-data (i.e., packet or NetFlow) within short time-scales (e.g., in milliseconds). DDoS attacks can be a low-rate and high-rate attack based on attack-rate dynamics. A low-rate DDoS attacker attempts to bypass the security system by sending attack packets to the victim at a sufficiently low rate to elude detection [4]. In a high-rate DDoS attack, the attacker sends a burst of attack packets to the victim within a short interval of time to overwhelm the bandwidth or resources. The task of analyzing both header and payload of network traffic is an NP-complete problem. So, we mostly focus on packet header information and visualization in terms of different parameters to support low-rate and high-rate DDoS attack detection.

Several works have been reported to detect attacks in large-volume alerts, produced by a detection tool employing visualization methods. DDoSViewer [5] is a visual interactive system used for detecting DDoS attacks. DDoSViewer designs for detecting DDoS attacks through the analysis of visual patterns. The Spinning Cube [6] maps SIP, DIP and Dport to the axes in a 3D plot. The amount of network activity is visualized interactively in the plot using colour, displaying certain attacks (e.g., port scans) very clearly. Zhou et al. [7] presented a low-rate DDoS detection scheme developed based on the distribution of packet size. They estimated the packet size distribution distance between legitimate and low-

*Dr. Bhuyan is on lien from the Department of Computer Science and Engineering, Assam Kaziranga University, Jorhat, India.

In this paper, we present TVis, a visualization system to detect both low-rate and high-rate DDoS attacks based on time-periodic sampled traffic. We model network traffic as an undirected graph and compute the area of a triangle formed based on incidence on each vertex for a time window of 5 seconds. We identify the three consecutive maximal dense vertices to form a triangle and estimate the area of the triangle. The TVis system has the following steps: (a) examine all packets at the monitoring endpoint; (b) use memory efficient data structures; (c) generate statistical summaries that can be retained for further analysis; (d) generate an undirected graph to visualize the network; and (e) perform triangle-area mapping of dense incident vertex. The main contributions of this work are as follows:

- We introduce a light-weight traffic visualization system to detect both low-rate and high-rate DDoS attacks using Heron’s triangle-area map estimation.
- TVis is cost-effective as it visualizes network traffic. It can perform visualization in both online and offline modes.
- TVis has been validated using testbed and benchmark datasets. In both cases, it performs well compared to its competitors.

The rest of the paper is organized as follows. Section II introduces the proposed visualization system describing the framework and the model. Section III presents performance evaluation using testbed and benchmark datasets, and finally concludes with Section IV.

II. TVIS: THE PROPOSED SYSTEM

This section starts by describing the proposed system architecture followed by the algorithm. It includes the strategy for visualization and detection of both low-rate and high-rate DDoS attacks.

A. TVIS: A Framework

We model the proposed system as an undirected graph $H = \{h_1, h_2, \cdots, h_n\}$, with each host as a vertex, and a number of incidences on each vertex $I = \{i_1, i_2, \cdots, i_n\}$ to form a triangle and estimate the area for finding the infected period. To get the end-point traffic, we configure our network to redirect all traffic to a particular port. So, TVis can monitor each traffic instance and visualize them to detect both low-rate and high-rate DDoS attacks. The framework of the proposed system is given in Figure 1. TVis uses the jNetPcap [9] library for capturing and preprocessing traffic. After capturing network traffic, it filters out the IP packets for subsequent analysis. It uses developed subroutines to extract various relevant features from IP packets and finally constructs a 5 min traffic feature sample for offline analysis. Due to the light-weight nature of our system, TVis can visualize the traffic fast to support attack detection.

We employ Heron’s triangle-area map computation to estimate the infected period based on the incidences at a host (i.e., vertex). The concept of triangle-area map computation is shown in Figure 2. Let the triangle-area map be $A = \{a_1, a_2, \cdots, a_n\}$, incidences per host be $H_I = \{h_{i_1}, h_{i_2}, \cdots, h_{i_n}\}$, and time periods be $T = \{t_1, t_2, \cdots, t_n\}$ used for representation of the proposed system. Hence, the area of a triangle can be defined as:

$$\Delta a_1 = \sqrt{s_k(s_k - h_{i_1})(s_k - h_{i_2})(s_k - h_{i_3})}$$

where $s_k$ is the semi-perimeter with $s_k = (h_{i_1} + h_{i_2} + h_{i_3})/2$, and $\Delta a_1$ is the area of an infected period.

B. TVIS: Algorithm

Algorithm 1 shows the significant steps in the design of the TVis system for network traffic visualization and analysis for the detection of both low-rate and high-rate DDoS attacks. It

---

1http://mathworld.wolfram.com/HeronsFormula.html
has two main modules: online() and offline(). In the online() mode, it captures, pre-processes and splits traffic instances, and visualizes them for attack detection. TVis maps the source node to the destination node based on connection information. But in the offline() mode, it just visualizes the already stored preprocessed traffic instances for detection. TVis works in a similar fashion in the offline mode except that it captures and prepossesses packets from a file. So, the cost of computation in online() mode is more than offline mode to provide a near real-time performance. This algorithm offers two categories of graphs: a sparse graph and a dense graph that represent legitimate and attack traffic, respectively.

Further, TVis accumulates three consecutive vertices with maximal incidences and compute the area of the triangle using Heron’s formula based a time window. If the area is beyond than dynamic thresholds $\delta_{A_i}$ and $\delta_{A_h}$ for low-rate and high-rate attacks, and the probability of packet loss is high, then it generates an alarm. Low area of the triangle indicates high-rate attacks and vice versa.

### III. PERFORMANCE EVALUATION

In this section, we describe the datasets used for performance analysis of TVis and report experimental results in detail.

#### A. Datasets

We use three different real-world datasets: (i) MIT Lincoln Laboratory [10], (ii) CAIDA DDoS 2007 [11], and (iii) Assam Kaziranga University (AKU) network dataset. The MIT Lincoln Laboratory dataset is real-time and contains pure normal data. The CAIDA DDoS 2007 dataset contains one hour of anonymized traffic traces from a DDoS attack launched on August 4, 2007. This dataset includes mainly two types of attacks: consumption of computing resources and the consumption of network bandwidth. We also used TVis system to collect, monitor, and visualize live network traffic in the Assam Kaziranga University campus. The AKU dataset is composed of three categories of traffic, viz., normal traffic, low-rate attack traffic, and high-rate attack traffic. The network is comprised of about 500 hosts (i.e., laptop, desktop, server), 6 L3 switches, and 25 wireless routers inside the University campus. We configure the network to monitor all network traffic at an endpoint. We mount four different attacks, viz., syn flood, smurf, ping flood, and fraggle in distributed mode towards multiple targets. We attempt to detect both low-rate and high-rate DDoS attacks within a short time interval. So, TVis can increase the network uptime rather just detection of DDoS attacks.

#### B. Results

We evaluate the TVis system in both online and offline modes by considering multiple attack scenarios. In the offline mode, we evaluate the TVis system using the CAIDA DDoS 2007 dataset. Figure 3 shows the visualized network traffic in offline mode. We can see in Figure 3 how TVis identifies the presence of attack traffic. The TVis system immediately sends

### Algorithm 1 TVis (online, offline)

**Input:** mode

**Output:** The visualized graph with triangle area

```
1: if mode ≠ online then
2:   call online()
3: else
4:   call offline()
5: end if
6: function ONLINE( )
7:   Initialize storePacket[60], device, T[3]  # It is an array of linked list to store packets as they arrive, with 60 arrays each storing 5 sec data, total 5mins
8:   Find all the network devices connected to the machine
9:   device = get the choice of device from the list
10:  Open the device for capturing in promiscuous mode
11:  for i ← 0 to 59 do
12:      storePacket[i] = CAPTURE( device)
13:  end for
14:  function NODEVISUALGRAPH(storePacket)
15:  Initialize graph  # an undirected graph where vertices are host on the network and edges represent communication among them, vertex  # linked list of devices, i.e., IP addresses, edges  # a linked list each having value $(v_i, v_j)$ where $v_i, v_j \in \text{vertex}$
16:  for $i ← 0$, to size of storePacket do
17:     for all packet in storePacket[i] do
18:        if packet is an IP packet then
19:           if packet.sourceIP not in vertex then
20:              add packet.sourceIP to vertex
21:           end if
22:         end if
23:         if packet.destinationIP not in vertex then
24:            add packet.destinationIP to vertex
25:         end if
26:         if (packet.sourceIP, packet.destinationIP) not in edges then
27:            add $(packet.sourceIP, packet.destinationIP)$
28:         end if
29:      end for
30:  end function
31: function OFFLINE( )
32:   Initialize storePacket[60]  # It is an array of linked list to store packets as they arrive, time = 5000 (Time is in milliseconds, 5000 represents 5 sec), $i = 0$ (for accessing the array)
```

get the pcap file $pcapFile$ from user
open $pcapFile$ to read packets
for all packets in $pcapFile$ do
    if packet.timestamp > time then
        time = time + 5000
        i ++
    end if
    add packet to $storePacket[0]$
end for
ANALYSE($storePacket$)
NODEVISUALGRAPH($storePacket$)
exit
end function

function TRIANGLEAREA($\text{T}[]()$)
Initialize $A$, $k$ $\triangleright A$ indicates the area of a triangle
for $i \neq k$ do
    if $\text{T}[i] \geq 1600$ and $pl \geq 0.22$ then
        compute $A$ using Equation 1
    end if
    if $(A_1 \geq \delta A_h \mid A_1 \leq \delta A_h)$ then
        Triggers an alarm
    end if
end for
end function

a request to the edge router to drop the packet before entering the network. It also depends on the period between the attack pulses to overwhelm the target. We compute the triangle area iff the system finds consecutive three vertices with incidences greater than at least 1600 packets, and increased packet losses during the same time.

We also evaluate the TVis system using the MIT Lincoln Laboratory dataset to differentiate between legitimate and attack traffic. Figure 4 shows the visualized network traffic in the offline mode using MIT Lincoln Laboratory dataset. In Figure 4, we see that the presence or absence of attacks in the traffic. This is because it generates a sparse graph that enables us to identify legitimate traffic from time-periodic sampled data. Attack traffic instances are rarer than legitimate traffic instances.

In the online mode, we used the live network traffic of Assam Kaziranga University campus when executing attacks. We captured and visualized traffic for 5 second time windows for total 5 minutes as shown in Figure 9 when executing attacks in the testbed. TVis system can visualize the unique IP address and packets per protocol within a time window as shown in Figure 5, and 6, respectively. Also, this system shows the unique ports for TCP and UDP protocols in Figure 7 and 8, respectively. We can see that in the visualization, not all host in the network are shown, because it shows only those hosts that are active and either sending or receiving packets, including the target. Figure 9 exhibits dense graph to ensure TVis system for DDoS detection.

Following Moore et al. [12], we generate both low-rate and high-rate DDoS attack to validate TVis system. The attack traffic is generated using open-source attack codes with more than 1600 and less than 5000 packets per seconds for low-rate attacks and produces high-rate attack if it goes beyond these limits. However, this number will varies based on datasets and environments. Based on our experiment, we observe that TVis triggers an alarm for an attack when the normalized area of the triangle, $\delta A \geq 0.43$ with $P_k \geq 1600$ packets per seconds with...
5 second time window. Our system is significant because of the reasons given below. It is also superior to recent work [13]. Figure 10 reports the ROC curve of TVis system to detect DDoS attacks when using the MIT Lincoln Laboratory legitimate and the CAIDA DDoS datasets.

- TVis is cost-effective and can operate in both online and offline modes.
- It is fast and scalable, and is able to detect both low-rate and high-rate DDoS attacks effectively.

The triangle area $\delta_A$ goes lower for high-rate DDoS attacks and increases for low-rate attacks. High-rate attacks are more frequent towards a target with a high intensities of malicious incidences in a short time window. However, low-rate attacks are less frequent and are similar to legitimate traffic. For our experiment, we found best results when $\delta_{Ah} \geq 0.43$ and $\delta_{Al} \leq 0.61$, otherwise low-rate attacks. TVis shows $\delta_{Al,leg} < 0.43$ for legitimate traffic instances.

### C. Comparison with Competing Methods

Visualization-based methods for DDoS detection have been reported in the past. The main reason could be the increased size, speed, and complexity of networks have made proper visualization difficult. However, recently introduced software-defined networks to isolate network operations and make visualization of network flows useful again. Visualization techniques allow people to see and comprehend large amounts of complex data [16]. Graphics are used to assist IDS investigation and the reporting process by helping the analyst to identify significant incidents and reduce false alarms. Intricate patterns are displayed over time in a secure way, where each of them can be understand easily [16].
To assess the efficacy of TVis, we compare several methods such as sequence Fourier power spectral entropy (FPSE) [17], wavelet power spectral entropy (WPSE) [17], and the sequence alignment method [18]. The Fourier power spectral entropy (FPSE) and wavelet power spectral entropy (WPSE) methods [17] are information-theoretic methods to detect low-rate DoS attacks, and then achieve 95% detection accuracy. However, the sequence alignment method to detect synchronous low-rate DoS attacks using the Smith-Waterman algorithm to estimate similarity score of two sequences, and achieves 95.88% detection rate. TVis improves overall performance when experimented over the MIT legitimate and the CAIDA DDoS datasets by achieving 98.79% and 96.54% detection rates for low-rate and high-rate DDoS detection, respectively. Table III-B reports a comparison of existing DDoS detection methods.

### IV. CONCLUSION AND FUTURE WORK

In this paper, we presented the TVis system to visualize network traffic in real-time for detection of both low-rate and high-rate DDoS attacks. The use of appropriate data structures is helpful in running the system in near real-time to support successful detection of both low-rate and high-rate DDoS attacks. An undirected graph is generated, and TVis triggers an alarm based on the estimated triangle-area of three consecutive maximal incident vertices using Heron’s formula. The TVis system performs well in detecting four different classes of DDoS attacks, including syn flood, smurf, fraggle, and ping flood. However, we report the results for real-time syn flood attacks only in offline and online modes with extended scalability.

The TVis system is under future development so that it can evolve with new attacks. We are also working on adding datacenter infrastructure and service visualization features to monitor and prevent incidents in real-time and reduce downtime of applications.

### ACKNOWLEDGMENT

This work was supported by the Kempe post-doc fellowship via project no. SMK-1644, Sweden. Additional support was provided by the International Exchange Program of the National Institute of Information and Communications (NICT) and JST CREST Grant Number JPMJCR1783, Japan.

### REFERENCES


