Distinguishing DDoS Attacks from Flash Crowds Using Probability Metrics

Ke Li, Wanlei Zhou, Ping Li, Jing Hai and Jianwen Liu
School of Engineering and Information Technology
Deakin University
{ktql, wanlei, pingli, jinghai, jianwen.liu}@deakin.edu.au

Abstract – *Both Flash crowds and DDoS (Distributed Denial-of-Service) attacks have very similar properties in terms of internet traffic, however Flash crowds are legitimate flows and DDoS attacks are illegitimate flows, and DDoS attacks have been a serious threat to internet security and stability. In this paper we propose a set of novel methods using probability metrics to distinguish DDoS attacks from Flash crowds effectively, and our simulations show that the proposed methods work well. In particular, these mathods can not only distinguish DDoS attacks from Flash crowds clearly, but also can distinguish the anomaly flow being DDoS attacks flow or being Flash crowd flow from Normal network flow effectively. Furthermore, we show our proposed hybrid probability metrics can greatly reduce both false positive and false negative rates in detection.*

Keywords: DDoS, Flash crowd, Probability metrics

1. Introduction

Nowadays, DDoS (Distributed Denial-of-Service) attacks still are one of the most destructive attacking means and the sources of mass disruption on internet. DDoS attacks typically occur when a large number of internet packets from compromised hosts (zombies) flood the bandwidth or resources of a single target (victim), and the flood of incoming messages to the victim essentially forces it to respond so slowly as to be rendered effectively unavailable and even to shut down, thereby coursing denial of service for legitimate users of the targeted system [5].

Flash crowds are large surges of legitimate traffic focus on some specify sites on internet over the relatively short period of time. In general, Flash crowds usually occur on popular web sites when hundreds and thousands of requests access to the web servers simultaneously. Flash crowds are quite similar with DDoS attacks in terms of network anomaly and traffic phenomenon; in fact sometimes also can cause a web site or target to slow down its service for users or even temporarily close due to the significantly increased traffic [1, 11, 16, 18].

Although Flash crowds and DDoS attacks are very much alike in their traffic behaviors, there are still some main differences between them in their nature and origin such as their access intents and the distributions of their source IP address and the increased and decreased speeds of traffic between them. In this paper, we take full advantage of these differences to distinguish DDoS attacks from Flash crowds using our proposed approaches effectively and quickly [2, 8].

In statistical theory, the probability metric [15, 19] is a numerical function on the space or distance of distributions of random elements. It must satisfy the following conditions and properties:

Let the probability metric be $D(x, y)$, $\forall x, y, z \in R$. We have:

1. Identity property: $D(x, y)=0$, while $x = y$.
2. Symmetry property: $D(x, y) = D(y, x)$.
3. Triangle inequality: $D(x, y) \le D(x, z) + D(z, y)$.

The probability metric includes many classes such as the total variation metric and the Bhattacharyya metric which both are the most important probability metrics. The total variation metric mainly measures the difference of two discrete probability distributions, and the Bhattacharyya metric is then mainly used to measure the similarity of two discrete probability distributions. We can take full advantage of the properties of the total variation metric and the Bhattacharyya metric to distinguish DDoS attacks from Flash crowds.

The main contributions of this paper are:

1. It proposes using a hybrid metric of the total variation metric and the Bhattacharyya metric to distinguish clearly DDoS attacks from Flash crowds. The proposed hybrid metric also can reduce the false positive rate greatly.
2. It proposes the idea and concept of using the hybrid probability metric to distinguish not only DDoS attacks from Flash crowds but also DDoS attacks from Normal network flow and even Flash crowds from Normal network flow, and shows that the proposed metric works well.

The remainder of this paper is organized as follows: We analyses the traffic models in section 2; in section 3 we provide the detection algorithm and system analysis. In section 4 we analyses the effectiveness of our proposed metric using real datasets. Section 5 analyses the system's detection sensitivity and section 6 presents our simulation

978-0-7695-3838-9/09 $26.00 © 2009 IEEE
DOI 10.1109/NSS.2009.35

results. Finally we conclude our paper and present the future work in section 7.

2. Traffic Model Analysis

We know that Flash crowds and DDoS attacks are very similar in traffic behavior from macroscopic observation; however there are also several essential differences in the aspects of access intents, distributions of source IP address and speed of the increased and decreased traffic.

At first, their access intents are quite different. Flash crowds are the results of the legitimate users respond to special events such as breaking news or popular products (movies, music and software) release. All the users just want to obtain the information or material they wanted from the server; they expect their access are successful and quick, do if the server is slowed down or even shut down are the things that they are extremely unwilling to see. However, DDoS attacks are not social events and all the requests are launched by attackers and are illegitimate. The attackers have only one aim that is to shut down the server quickly, or to cause the server not available for legitimate users.

Secondly, the distributions of the source IP address are also quite different between Flash crowds and DDoS attacks. The users of Flash crowds are just interested in the specified events in the server, they can come from the whole community network or the whole internet, so the distribution of source IP addresses in Flash crowds is very dispersive; If we aggregate these IP addresses, the distribution of source IP addresses will be subject to the fractional Gaussian noise distribution [10, 14] approximately but is more dispersive than it, so the waveform of the distribution of source IP addresses after aggregation in Flash crowds is more flat than the waveform of the fractional Gaussian noise distribution. However, in DDoS attacks, all the users are illegitimate as all the requests are launched by attackers or zombies. Therefore, the distribution of source IP addresses is concentrated relatively according to the limited attackers or zombies. If we aggregate these source addresses, the number of the source IP addresses in DDoS attacks will be decreased quickly and the distribution of source IP addresses will be subject to the Poisson distribution [7, 17].

Thirdly, there is a big difference in the increased and decreased speed of traffics between them. In Flash crowds, all users are impossible to access simultaneously the same server at the beginning, because the messages or news need take time to spread among the users; So the number of requests to the server is increased gradually then to the peak; similarly, at the end stage of the Flash crowds, all users will not lose their interesting to the server simultaneously, so the number of requests to the server will be decreased gradually from the peak. However, in DDoS attacking, the attackers or zombies must launch a large number of requests to the server simultaneously or within a very short time difference to achieve the desired attack effect, it is not difficult that because the DDoS attacks are controlled by machines completely, so to launch attacks at the same time can be completed automatically by machines; Therefore, the number of requests to the server is increased sharply to reach the peak, and then will be decreased sharply also at the end stage of the DDoS attacks.

3. Detection Algorithm and System Analysis

The total variation [3, 13] is one of the most important divergences in mathematical statistics; it can measure the largest possible difference between two probability distributions which can assign to the same event. Given two complete probability distributions $P= (p_1, p_2, \ldots, p_n)$ and $Q= (q_1, q_2, \ldots, q_n)$ with $\sum_{i=1}^{n} p_i = \sum_{i=1}^{n} q_i = 1$, $1 \ge p_i \ge 0$, $1 \ge q_i \ge 0$, $i=\{1, 2, \ldots, n\}$. We have the total variation between them as follows:

$$T(P,Q) = \sum_{i=1}^{n} | p_i - q_i |$$

$T(P,Q)=0$, while $P=Q$.

Obviously, $T(P,Q)=T(Q,P)$, so the total variation is symmetric measure and belongs to the probability metric.

The value of the total variation is increasing from zero denotes that the difference between two distributions is expending.

The Bhattacharyya coefficient [4, 6, 9] is another most popular statistical measure and sometimes it is also called the similarity coefficient and it measures the similarity of two discrete probability distributions. Obviously, on the other hand it is also a divergence-type measure and can be used to measure the dissimilarity of any two classes in classification. Given two complete probability distributions P and Q, the conditions are same as the above. The Bhattacharyya coefficient between P and Q is denoted by $\rho(P,Q)$, the definition as follows:

$$\rho(P,Q) = \sum_{i=1}^{n} \sqrt{p_i q_i}$$

The properties of the Bhattacharyya coefficient:

1. $0 \le \rho(P,Q) \le 1$
2. $\rho(P,Q)=1$, while $P=Q$.
3. $\rho(P,Q)=0$, while P is orthogonal to Q.

The Bhattacharyya coefficient has the symmetric property and is also a probability metric as $\rho(P,Q) = \rho(Q,P)$. The value of the Bhattacharyya coefficient indicates the similarity between two probability distributions, unit indicates the strongest similarity, and on the contrary, zero indicates the weakest similarity between the distributions. So the Bhattacharyya coefficient is also named the similarity coefficient.

Based on above discussion and analysis, we design the DDoS detection system shown as Figure 1 which uses the hybrid probability metric of total variation and similarity coefficient to distinguish DDoS attacks flow from Flash crowd flow and DDoS attacks flow from Normal network flow, and even also can distinguish the anomaly being DDoS attacks flow or Flash crowd flow from Normal network flow very well.

The detection system includes five parts such as flow anomaly detector, flow distribution estimator, hybrid probability metric calculator and decision device. The main function of each part is described as follows. The flow anomaly detector is a multi-input and bi-output device which detects the flows anomaly of the incoming flows in a specified router, and only having the anomaly flow; it can take effect and output two flows which include one abnormal flow at least. The flow distribution estimator is used to sample the flow's distribution according to its recognizable characteristics in the sampling period to obtain the flow probability distribution. The hybrid probability metric calculator is used to compute the values of the total variation and the similarity coefficient of two flows in parallel. The decision device is used to distinguish DDoS attacks flow from Flash crowd flow or Normal network flow, and deside the anomaly flow being DDoS attacks flow or flash crowd flow from Normal network flow by the combined using the grouping thresholds GT_T and GT_S, in this paper, the grouping threshold GT_T and GT_S express the values of decision in detecting DDoS attacks using the total variation metric and the similarity coefficient metric respectively, if it decides the anomaly flow being DDoS attacks flow, it can discard the attacks flow immediately, otherwise pass the flow to the destination or the downstream routers.

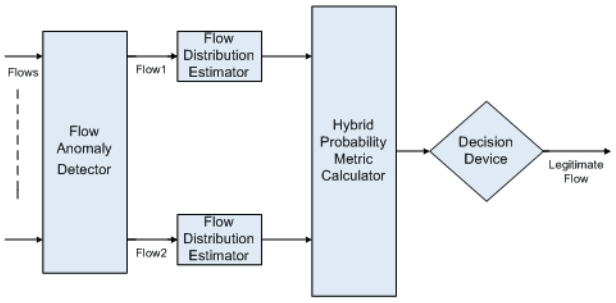

Figure 1. DDoS attacks detection system of using the hybrid probability metric

Listing 1 is the detection algorithm of using the hybrid probability metric which can not only detect DDoS attacks flow form Flash crowd flow and Normal network flow clearly, but also can decide the anomaly flow being DDoS attacks flow or Flash crowd flow from Normal network flow very well. It can reduce both false positive rate and false negative rate as well.

Listing 1. The detection algorithm of using the hybrid probability metric

The DDoS attacks detection algorithm:

1. Set the sampling frequency as *f*, the sampling period as *T*, and the grouping thresholds as GT_T and GT_S.
2. In the router after aggregation of traffic, sampling the network flows come from the upstream routers.
3. Calculate the numbers of packet which has various recognizable characteristics (such as the source IP address or the packet's size, etc.) in each sampling time interval τ ($\tau = \frac{1}{f}$) within *T*.
4. Calculate in parallel the probability distributions of the sampled network flows.
5. Calculate in parallel the values of the total variation and the similarity coefficient among each of the pair using the formulas as follows:

   $$T(P,Q) = \sum_{i=1}^{n} | p_i - q_i |$$

   and $$\rho(P,Q) = \sum_{i=1}^{n} \sqrt{p_i q_i}$$
6. If the value of the total variation of any two distributions is more than the lower bound of the grouping threshold GT_T (1.1045) and the value of the similarity coefficient is less than the upper bound of GT_S (0.7220), then the system detected the DDoS attacks from Flash crowds, and begins to raise alarms and discard attack packets.
7. If the value of total variation is located in the grouping threshold GT_T (the lower bound: 0.5921, and the upper bound: 1.1045) and the value of the similarity coefficient is located in GT_S (the lower bound: 0.7220, and the upper bound: 0.8708), then the system detected the DDoS attacks from Normal network flow, and begins to raise alarms and discard attack packets.
8. If the value of the total variation of any two distributions is less than the upper bound of the grouping threshold GT_T (0.5921) and the value of the similarity coefficient is more than the lower bound of GT_S (0.8708), then the system detected the Flash crowds from Normal network flow, and begins to raise alarms.
9. Otherwise the router forwards the packets to the destination or the downstream routers.
10. Return to step 2.

4. Effectiveness Analysis

To analyze the effectiveness of our proposed metric in detecting DDoS attacks, we use the real datasets in our experiment as follows. Firstly, we use the MIT Lincoln Laboratory Scenario (DDoS) 2.0.2 dataset [20] as the incoming DDoS attacks flow, in which the scenario includes a distributed denial service attack run by a novice attacker and is performed over multiple networks. We downloaded and used the inside tcpdump sensor dataset where the sniffer is on the "inside" network to do the experiment. In the inside tcpdump data-set, we used the dataset with TCP protocol only filter and obtained a TCP SYN flooding data set, the partial attack scenario is shown as Figure 2; this is because that we know the DDoS attack in this scenario only contains a TCP SYN flooding attack, so the filtered data-set can simplify our execution.

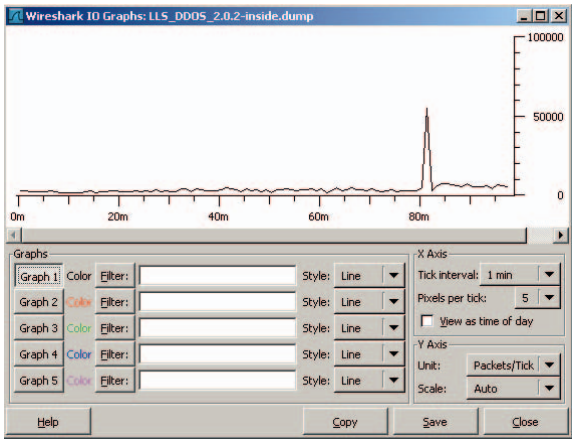

Figure 2. TCP SYN flooding scenario from MIT/LL, X-axis denotes tick interval (minute), Y-axis denotes packets/tick

From Figure 2 we know that the DDoS attacks were launched from the 80th minute and end to the 82nd minute by an attacker in the attacks scenario, obviously, it is a small and simple DDoS attack.

Secondly, we filter the DDoS attacks flow dada from the same MIT Lincoln Laboratory Scenario (DDoS) 2.0.2 dataset and can obtain the Normal network flow dataset to simulate approximately the Normal network flow in our experiment.

And thirdly, we use a day of HTTP logs dataset from a very busy WWW server as the incoming Flash crowds flow approximately. This dataset contains a day's worth of all HTTP requests to the SDSC WWW server located at the San Diego Supercomputer Center in San Diego, California [21].

Finally, we process, replace and classify the DDoS attacks dataset, the Normal network flow dataset and the Flash crowd flow dataset by their source addresses according to requirements, and sample out the comparable sample points which hold the uniform length and interval from the classified datasets by the certain period. Their probability distribution of source IP addresses is shown as Figure 3.

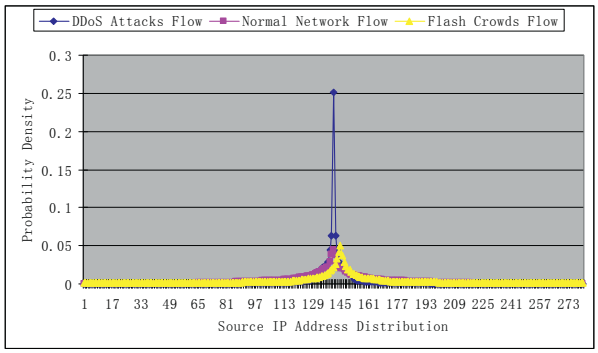

Figure 3. Probability density distribution of source IP addresses of DDoS attacks flow, Normal network flow and Flash crowds flow

We test our proposed metric using the above datasets; the experimental results are shown as Table 1.

Table 1. Values of total variation and similarity coefficient among DDoS attacks flow, Normal network flow and Flash crowds flow

Metrics \ Flows	Flash Crowds vs. DDoS Attacks	Normal Flows vs. DDoS Attacks	Flash crowds vs. Normal Flows
Total Variation	0.8469	0.7053	0.4596
Grouping Threshold GT_T	Lower Bound: 0.7761	Upper Bound: 0.7761 and Lower Bound: 0.5825	Upper Bound: 0.5825
Similarity Coefficient	0.8476	0.9047	0.9526
Grouping Threshold GT_S	Upper Bound: 0.8762	Upper Bound: 0.9287 and Lower Bound 0.8762	Lower Bound: 0.9287

From Table 1 we know that our proposed metric can not only detect DDoS attacks flow from Normal network flow effectively but also can distinguish DDoS attacks flow from

Flash crowd flow clearly, and even can detect Flash crowd flow from Normal network flow very well, it means that it can decide the anomaly flow being DDoS attacks flow or Flash crowd flow from Normal network flow.

5. Detection Sensitivity Analysis

In this section we focus on discussing the false positive rate of the detection system. The false negative rate has the same principle. We can predigest the detection system shown as Figure 4 for the aim of analysis.

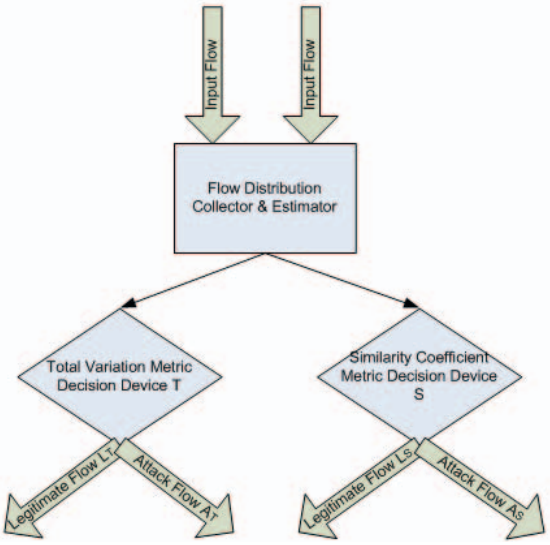

Figure 4. The simplified DDoS attacks detection system of using the hybrid probability metric

In the simplified detection system, the flow distribution collector & estimator collects and estimates the input flows and exports the flow distributions towards decision devices in which the total variation metric decision device T calculates the input flow distributions using the total variation metric and distinguishes out the attack flow A_T from the legitimate flow L_T, in parallel, the similarity coefficient metric decision device S calculates the input flow distributions using the similarity coefficient metric and distinguishes out the attack flow A_S from the legitimate flow L_S.

We take advantage of the set concept which is shown as Figure 5 to discuss the false positive and negative rate of the detection system. We set the output legitimate flows L_T and L_S as L_T set and L_S set respectively, the output attack flow A_T and A_S as A_T set and A_S set respectively, U set is the universal set of the output flows, A set and L set are the universal sets of the attack flows and the legitimate flows respectively. Therefore, we have

$U = L_T + A_T = L_S + A_S$

$A = A_T \cup A_S = A_T \cap A_S + (A_T - A_T \cap A_S) + (A_S - A_T \cap A_S)$

$L = U - A = L_T \cap L_S = U - A_T \cap A_S - (A_T - A_T \cap A_S) - (A_S - A_T \cap A_S)$

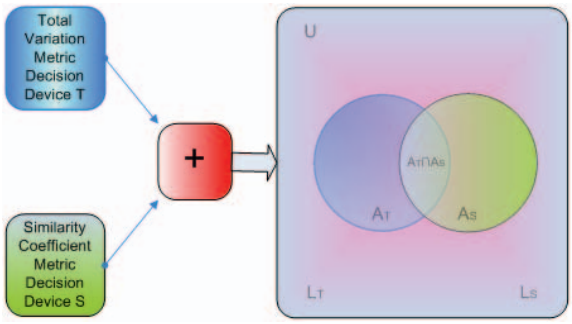

Figure 5. Output sets of the detection system using the hybrid probability metric

Let η_T and η_S be the false positive rates of using the total variation metric and the similarity coefficient metric in DDoS attacks detecting respectively, so if the attack flows are detected by both the total variation metric and the similarity coefficient metric synchronously, then the false positive rate of using the hybrid probability metric will be $\eta = \eta_T * \eta_S$, namely, the detected attack flows all are located in the set $A_T \cap A_S$. If the attack flow is detected only by the total variation metric decision device or the similarity coefficient metric decision device, then it will be decided further by checking the information of IP package's header, if it is empty, then the flow will be decided as the attack flow. Let η_H be the false positive rate of using IP package's header checking metric, therefore, the false positive rates after checking the information of IP package's header will be $\eta_T * \eta_H$ or $\eta_S * \eta_H$ respectively, namely, the detected attack flow is located in the set $A_T - A_T \cap A_S$ or the set $A_S - A_T \cap A_S$. Obviously, if we also check the information of IP package's header after the attack flows are detected by both the total variation metric and the similarity coefficient metric at the same time, the false positive rate of the detection system will be decreased to $\eta_T * \eta_S * \eta_H$.

Therefore, our proposed hybrid metric can reduce the false positive rate greatly in DDoS detecting, the same, the false negative rate can be reduced clearly too.

The process of flow processing and deciding in detection system is shown as Figure 6. The incoming flows are collected and estimated to various probability distributions by the flow distribution collector & estimator; the devices of total variation metric decision and similarity coefficient metric decision calculate the values of the total variation and the similarity coefficient between the various probability distributions respectively. The collaborative grouping thresholds decision device is a core part in our proposed detection system, it will distinguishes the attack flows from the legitimate flows correctly according to the

algorithm and approach we discussed above, for example, if the calculated values of total variation and similarity coefficient both are within their own grouping thresholds (the detailed data is shown as Table 1&2, for the aim of discussion, we assume that the boundary values of sets A_T and A_S in Figure 5 are the grouping thresholds values of using the total variation metric and the similarity coefficient metric respectively), namely, the value of total variation is located in the set A_T, and the value of similarity coefficient is located in the set A_S in Figure 5, the decision device will decide that the detected flow is a attack flow by the $\eta = \eta_T * \eta_S$ false positive rate and then discard it; otherwise, if the values of total variation and similarity coefficient both are located in the set L = U – A, the decision device will decide that the detected flow is a legitimate flow and pass it to the destination or the downstream routers; however, if the values are not both within their own grouping thresholds, namely, the values both can be located in the set (A_T - $A_T \cap A_S$) or set (A_S - $A_T \cap A_S$) in Figure 5, the collaborative grouping thresholds decision device will hand in the flows to the IP header checking device to check whether the flows' IP headers are empty or not, if they are empty, the final decision device will decide that the detected flow is a attack flow by the $\eta_T * \eta_H$ or $\eta_S * \eta_H$ false positive rate and then discard it; otherwise, the device will decide that the detected flow is a legitimate flow and pass it to the destination or the downstream routers.

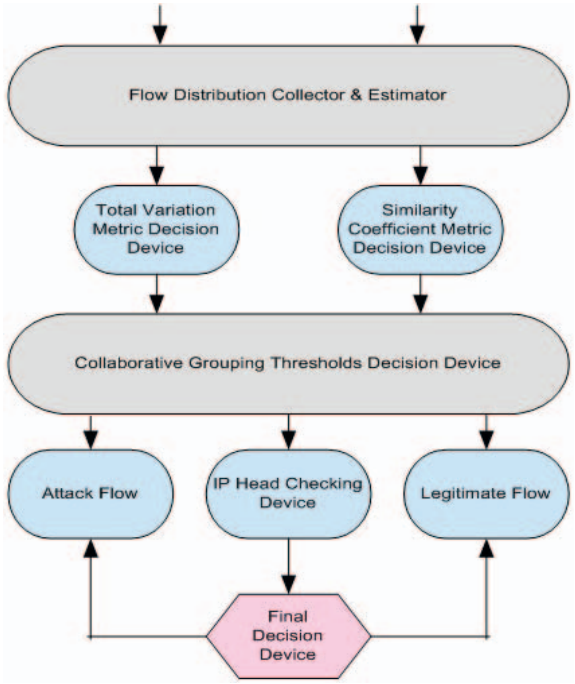

Figure 6. Flows processing and deciding in the detection system of using the hybrid probability metric

6. Simulation and Experiment Results

6.1 Simulation Results

Although the MIT Lincoln Laboratory data-sets have been used to evaluate lots of attack detection systems and algorithms by a number researchers, there still is a lack of standard and appropriate evaluation data-set that can be used to simulate the realistic network environments currently. As indicated by McHugh [12] etc. that the methodology used to generate the data by MIT Lincoln Laboratory and the data itself are not appropriate for simulating the realistic network environments. Furthermore, the Flash crowd dataset we used in the experiments is not perfect, as the data providers have removed all of the client IP addresses and replaced them with unique but unclear identifier to ensure the privacy of each individual that visited the SDSC WWW server.

Therefore, in our simulations, we consider using the Poisson distribution to simulate the DDoS attacks flow [7], using the fractional Gaussian noise distribution to simulate the Normal network flow [10, 14] and using the adjusted fractional Gaussian noise distribution (the number of source IP addresses increased 7 times and the value of probability of distributions decreased 5 times than in the Normal network flows) to simulate the Flash crowd flows approximately [1]. The simulation results are described in Table 2.

Table 2. Values of total variation and similarity coefficient among DDoS attacks flow, Normal network flow and Flash crowd flow

Flows / Metrics	Flash Crowds vs. DDoS Attacks	Normal Flows vs. DDoS Attacks	Flash crowds vs. Normal Flows
Total Variation	1.1556	1.0533	0.1309
Grouping Threshold GT_T	Lower Bound: 1.1045	Upper Bound: 1.1045 and Lower Bound: 0.5921	Upper Bound: 0.5921
Similarity Coefficient	0.6995	0.7445	0.9970
Grouping Threshold GT_S	Upper Bound: 0.7220	Upper Bound: 0.8708 and Lower Bound 0.7220	Lower Bound: 0.8708

From Table 2, we can see that the value of total variation achieves the maximum (1.1556) between Flash crowds and DDoS attacks in this three pairs which are Flash crowds vs. DDoS attacks, Normal flows vs. DDoS attacks and Flash crowds vs. Normal flows. And the value of total variation achieves the minimum (0.1309) between Flash crowds and Normal flows. Therefore, it fully shows that the distributions of source IP addresses of Flash crowds and Normal flows are very similar, but the distributions of Flash crowds and DDoS attacks are quite different. So, we can use these properties to distinguish DDoS attacks from Flash crowds clearly and quickly. On the other hand, we can see that the value of similarity coefficient achieves the maximum (0.9970) between Flash crowds and Normal flows in the three pairs, and the value of similarity coefficient achieves the minimum (0.6995) between Flash crowds and DDoS attacks, so it also shows that the distributions of source IP addresses of Flash crowds and Normal flows are very similar, and the distributions of Flash crowds and DDoS attacks are quite different. The same, we can use this property to distinguish DDoS attacks from Flash crowds clearly.

In this paper, we use the hybrid probability metric to detect DDoS attacks for the aim of reducing the false positive rate and the false negative rate. We compute the values of total variation and similarity coefficient of any two distributions of the network flows at first, if the value of total variation is more than the lower bound GT_T (1.1045) and the value of similarity coefficient is less than the upper bound GT_S (0.7220), we can decide that the two distributions are the flows of Flash crowds and DDoS attacks, and then to discard the DDoS attacks flows, if the value of total variation is more than the lower bound GT_T (0.5921) and less than the upper bound GT_T (1.1045), and the value of similarity coefficient is more than the lower bound GT_S (0.7220) and less than the upper bound GT_S (0.8708), we can decide the two distributions are the DDoS attacks flow and the Normal network flow, the same, we can decide the two distributions are the Flash crowds flow and the Normal network flow if the value of total variation is less than the upper bound GT_T (0.5921) but is not closed to zero it means that they are not the same distributions and the value of similarity coefficient is more than the lower bound GT_S (0.8708) but is not closed to unit it means that they are not the same distributions.

6.2 Experiment Results

We also test the proposed metric in the simulative network environment using the standard network software. At first, we simulate the DDoS attacks flow, Flash crowds flow and the Normal network flow are shown in Figure 7. In our network simulation, we set up a campus network firstly which has three different network sizes to simulate approximately DDoS attacks flow, Flash crowds flow and Normal network flow respectively through the different network loads of the server.

In Figure 7, the blue curve represents the Normal flow, while the campus network is a small network size which has a few of nodes; the red curve represents the DDoS flow, while the campus network is a large network size which has a large number of nodes, all nodes are regarded as the attack nodes and connected by only several routers, and we allocate the different source IP addresses to the entire nodes according to the different routers and the number of the different IP addresses is the same number as the routers; the green curve represents the Flash Crowds flow, while the campus network also is a large network size which has a large number of nodes and the all nodes are regarded as the Flash crowds nodes in which we allocate the different source IP addresses for each of all nodes.

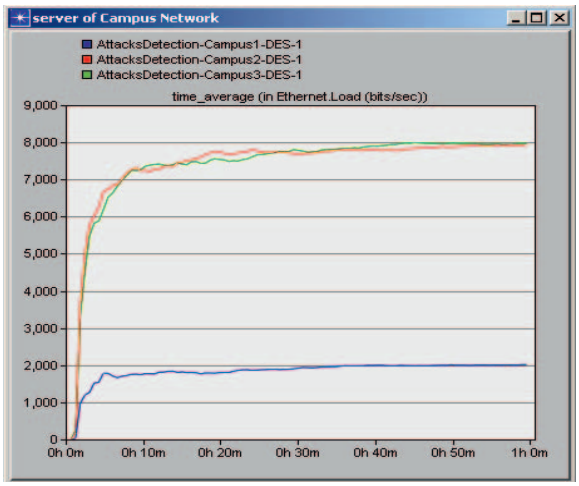

Figure 7. Simulative DDoS attacks flow, Flash crowds flow and Normal network flow in campus network environment

Secondly, we test the proposed metric among the incoming flows of DDoS attacks flow, Flash crowds flow and Normal network flow. Figure 8&9 show the decision results between these flows by using the proposed total variation metric and similarity coefficient metric respectively.

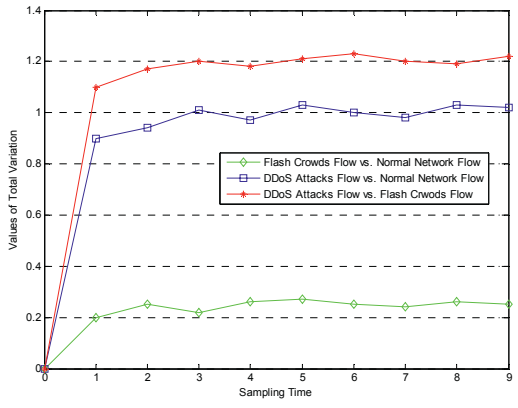

Figure 8. Decision results among the DDoS attacks flow, Flash crowds flow and the Normal network flow using the proposed total variation metric

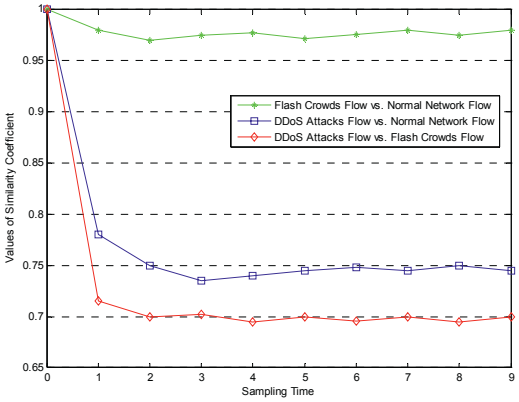

Figure 9. Decision results among the DDoS attacks flow, Flash crowds flow and the Normal network flow using the proposed similarity coefficient metric

Therefore, from above experimental results we can know that the proposed metric can not only distinguish out different network flows clearly (there are big gaps among these flows) but also early (can distinguish out flows at the first sampling period). During the DDoS attacks detecting, while network is in the normal network traffic environment, if the detection system detected the incoming flows are DDoS attacks flow and Normal network flow, the system will discard the attacks flow immediately; if the system detected the incoming flows are Flash crowds flow and Normal network flow, the system will forward the Flash crowds flow to the destination or the downstream routers; and while network is in Flash crowds traffic environment, if the system detected the incoming flows are DDoS attacks flow and Flash crowds flow, the system will discard the attacks flow immediately and forward the Flash crowds flow to the destination or the downstream routers.

7. Conclusion and Future Work

In this paper, we propose using hybrid probability metrics to detect DDoS attacks, and through experiment and simulation we show that the proposed metric can not only detect DDoS attacks from the normal flows, but also can distinguish DDoS attacks from Flash crowds clearly, and even can distinguish the anomaly flow being DDoS attacks flow or Flash crowds flow from the Normal network flow very well; Furthermore, our proposed hybrid metric can reduce both the false positive rate and the false negative rate greatly, so it can improve the detection sensitivity clearly. Our further research will be to verify the proposed hybrid metric in the real network situation, and to find more recognizable characteristics of IP packets and flows to achieve better detection effects in DDoS attacks detecting.

References

1. I. Ari, B. Hong, E.L. Miller, S.A. Brandt, and D.D.E. Long. *Managing flash crowds on the Internet*. in *Modeling, Analysis and Simulation of Computer Telecommunications Systems, 2003. MASCOTS 2003. 11th IEEE/ACM International Symposium on*. 2003.
2. P. Barford and D. Plonka, *Characteristics of network traffic flow anomalies*, in *Proceedings of the 1st ACM SIGCOMM Workshop on Internet Measurement*. 2001, ACM: San Francisco, California, USA.
3. A.R. Barron, L. Gyorfi, and E.C. van der Meulen, *Distribution estimation consistent in total variation and in two types of information divergence.* Information Theory, IEEE Transactions on, 1992. **38**(5): p. 1437-1454.
4. A. Bhattacharyya, *On a measure of divergence between two statistical populations defined by probability distributions.* Bull. Calcutta Math. Soc., 1943. **vol. 35**: p. pp. 99–109.
5. R.K.C. Chang, *Defending against flooding-based distributed denial-of-service attacks: a tutorial.* Communications Magazine, IEEE, 2002. **40**(10): p. 42-51.
6. A. Djouadi, O. Snorrason, and F.D. Garber, *The quality of training sample estimates of the Bhattacharyya coefficient.* Pattern Analysis and Machine Intelligence, IEEE Transactions on, 1990. **12**(1): p. 92-97.
7. P. Du and S. Abe, *IP Packet Size Entropy-Based Scheme for Detection of DoS/DDoS Attacks.* IEICE Transactions on Information and Systems, 2008: p. E91-D (5):1274-1281.
8. J. Jung, B. Krishnamurthy, and M. Rabinovich, *Flash crowds and denial of service attacks: characterization and implications for CDNs and web sites*, in *Proceedings of the 11th international conference on World Wide Web*. 2002, ACM: Honolulu, Hawaii, USA.
9. T. Kailath, *The Divergence and Bhattacharyya Distance Measures in Signal Selection.* Communication Technology, IEEE Transactions on, 1967. **15**(1): p. 52-60.
10. S. Ledesma and D. Liu, *Synthesis of Fractional Gaussian Noise Using Linear Approximation for Generating Self-Similar Network Traffic.* Computer Communication Review, 2000. **vol.30**.
11. D. Mayur, A. Abhishek, C. Mason, V. Nalini, and M. Sharad. *Flashback: A Peer-to-PeerWeb Server for Flash Crowds*. in *Distributed Computing Systems, 2007. ICDCS '07. 27th International Conference on*. 2007.

12. J. McHUGH, *Testing Intrusion detection systems: a critique of the 1998 and 1999 DARPA intrusion detection system evaluations as performed by Lincoln Laboratory.* ACM Transactions on Information and System Security (TISSEC) 2000. **3**(4): p. 262 - 294
13. A. Muller, *Integral Probability Metrics and Their Generating Classes of Functions.* Advances in Applied Probability, 1997. **29**(2): p. 429-443.
14. E. Perrin, R. Harba, R. Jennane, and I. Iribarren, *Fast and exact synthesis for 1-D fractional Brownian motion and fractional Gaussian noises.* Signal Processing Letters, IEEE, 2002. **9**(11): p. 382-384.
15. S.T. RACHEV, *Probability metrics and the stability of stochastic models*. Wiley series in probability and mathematical statistics. , ed. WILEY. 1991, USA: John Wiley & Sons Ltd.
16. D. Rubenstein and S. Sahu, *Can unstructured P2P protocols survive flash crowds?* Networking, IEEE/ACM Transactions on, 2005. **13**(3): p. 501-512.
17. W. Willinger. *Traffic modeling for high-speed networks: Theory versus practice*. in *Stochastic Networks*. 1995: Springer-Verlag.
18. W. Zhou. *Keynote III: Detection and traceback of DDoS attacks*. in *Computer and Information Technology, 2008. CIT 2008. 8th IEEE International Conference on*. 2008.
19. V.M. Zolotarev, *Probability Metrics.* Theory of Probability and its Applications, 1984. **28**(2): p. 278-302.
20. *http://www.ll.mit.edu/mission/communications/ist/corpora/ideval/data/2000/LLS_DDOS_2.0.2.html.*
21. *http://ita.ee.lbl.gov/html/contrib/SDSC-HTTP.html*