# Rank Correlation for Low-Rate DDoS Attack Detection: An Empirical Evaluation

Arindom Ain[1], Monowar H. Bhuyan[1], D. K. Bhattacharyya[2] and J.K. Kalita[3]
*(Corresponding author: Monowar H. Bhuyan)*

Department of Computer Science & Engg., Kaziranga University[1]
Jorhat-785006, Assam, India
Department of Computer Science & Engg., Tezpur University[2]
Tezpur-784028, Assam, India
Department of Computer Science, University of Colorado[3]
Colorado Springs, Co 80918, USA
(Email: {ainarindom,monowar.tezu}@gmail.com, dkb@tezu.ernet.in, jkalita@uccs.edu)
*(Received Nov. 26, 2014; revised and accepted )*

## Abstract

**A low-rate distributed denial of service (DDoS) attack has the ability to obscure its traffic because it is very similar to legitimate traffic. It can easily evade current detection mechanisms. Rank correlation measures can quantify significant differences between attack traffic and legitimate traffic based on their rank values. In this paper, we use two rank correlation measures, namely, Spearmen Rank Correlation (SRC) and Partial Rank Correlation (PRC) to detect low-rate DDoS attacks. These measures are empirically evaluated using three real-life datasets. Experimental results show that both measures can effectively discriminate legitimate traffic from attack traffic. We find that PRC performs better than SRC in detection of low-rate DDoS attacks in terms of spacing between malicious and legitimate traffic.**

*Keywords: DDoS attack, low-rate, network traffic, packet, rank correlation*

## 1 Introduction

With the rapid growth in the number of applications on Internet-connected computers and the devices and the rise in the sophistication of attacks on the application, early detection of Internet-based attacks is essential to reduce damage to legitimate user's traffic. A DDoS attack is a DoS attack that uses multiple distributed attack sources. Typically, attackers use a large number of compromised computers, also called zombies, to launch a DoS attack against a single target or multiple targets with the intention of making one or more services unavailable to intended users. Botnets have become a powerful way to control a large number of hosts, allowing the launching of sophisticated and stealth DDoS attack on target host(s) quickly [3].

In the recent past, botnets have become more intelligent and capable, and as a consequence the amount of attack traffic has increased targeting servers and components of Internet infrastructure such as firewalls, routers, DNS servers as well as network bandwidth. Regardless of how well secured the victim system may be, its susceptibility to DDoS attacks depends on the state of security in the rest of the global Internet [6]. A lot of different tools are used by attackers to bypass security systems, and as a result, researchers have to upgrade their approaches to handle new attacks simultaneously. Some defense mechanisms concentrate on detecting an attack close to the victim machine, because the detection accuracy of these mechanisms is high. Network traffic comes in a stream of packets and it is difficult to distinguish legitimate traffic from attack traffic. More importantly, the volume of attack traffic can be much larger than the system can handle. The behavior of network traffic is reflected by its statistical properties [9] because such properties summarize behaviour. Correlation measures can be used on the traffic summary to identify malicious traffic.

A network or host can be compromised with DDoS attacks using two types of traffic, namely, high-rate DDoS traffic and low-rate DDoS traffic. High-rate traffic is similar to flash crowd, i.e., when a large amount of unexpected legitimate traffic comes to a smallest server, and on the other hand, low-rate traffic is similar to legitimate traffic. So, it is very difficult to identify and mitigate either type of DDoS attack within a short time

period [1].

Correlation coefficient is a measure that can be used to identify linear relationship between malicious and legitimate traffic. In this paper, we attempts to use rank correlation to detect low-rate DDoS attacks. We use, two rank correlation techniques, namely, SRC and PRC.

The rest of the paper is organized as follows: Section II provides related work and observations. Section III presents the detection mechanism for low-rate DDoS attacks using rank correlation. Experimental results are reported in Section IV. Section V presents concluding remarks and future work.

## 2 Related work

A DoS attack is characterized by an explicit attempt to prevent the legitimate use of a service [6]. A DDoS attack deploys multiple attacking entities to attain this goal. Much research has been devoted to the detection of DDoS attacks [7]. Wei et al.[10] propose a rank correlation based approach to detect reflection DDoS attacks. Once suspicious flows are found, it estimates the rank correlation between flow pairs and generates a final alert according to preset thresholds. Sheng et al. [8] discuss a measure based on Hurst coefficient to detect low-rate DDoS attacks. Bhuyan et al. [1] present an empirical evaluation of the suitability of various information metrics to detect both low-rate and high-rate DDoS attacks. Jin and Yeung [2] propose a covariance analysis model for detecting SYN flooding attacks. The method can accurately detect DDoS attacks with different intensities. It can also detect DDoS attacks which are similar to legitimate traffic. Mathew and Katkar [5] propose a light-weight software-based approach for low-rate DoS (LDoS) attack detection, and integrated it with an existing intrusion detection system. It does not require any change in existing infrastructure and protocol. Xiang et al. [11] present a generalized information metric to detect both low-rate and high-rate DDoS attacks. They consider the spacing between legitimate traffic and attack traffic in terms of an information distance measure. We observe the following based on literature survey.

- Although a large number of methods have been introduced to detect high-rate DDoS attacks, the number of methods to detect low-rate DDoS attacks is small. Most methods to detect low-rate DDoS attacks suffer from significant large percentage of false alarms.
- Most published detection methods, attempt to detect at the packet level for low-rate DDoS attacks. Though NetFlow traffic analysis is faster than packet level analysis.

## 3 Rank Correlation for low-rate DDoS Attack Detection

Rank correlation has been found suitable as a potential metric to differentiate legitimate traffic from attack traffic [10]. Low-rate attacks exploit TCP retransmission time-out (RTO) to slowly reduce network throughput. An attacker causes legitimate TCP flow by entering the RTO state repeatedly. In a compromised host, it reduces the throughput significantly also reducing the bandwidth of the network simultaneously. A low-rate DDoS attack strategy is given in Figure 1. $T$ is the total time interval for a period. $T_w$ indicates the height of the attack burst, i.e., the strength of the attack traffic and $T_x$ represents the burst length that indicates the pulse width. $T_y$ is the time interval between two consecutive attack pulses, i.e., RTO + 2 round trip time (RTT). $T_z$ is the interval between two pulses of high-rate traffic, i.e., legitimate traffic, $N_a$, $N_b$, and finally $A_a, A_b$, $A_c$ are the high-rate attack traffic pulses towards a target from common effort of different attackers. The average volume of attack traffic can be calculated as $T_w * T_x/T$, which is much less than the legitimate TCP traffic [4]. So, it is difficult to detect attack traffic within short interval of time. In this work, two measures are used to detect low-rate DDoS attack, namely, Spearmen rank correlation and Partial rank correlation. Table I describes the symbols used to describe the method.

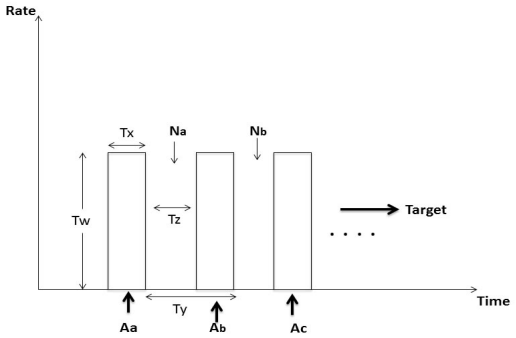


Figure 1: A low-rate DDoS attack strategy

### 3.1 Spearmen Rank Correlation (SRC)

Spearman's correlation coefficient measures the strength of association between two random variables better [10]. Spearman rank correlation coefficient is computed as

$$r_{X,Y} = \frac{E(X,Y) - E(X)E(Y)}{\sqrt{E(X^2) - E^2(X)}\sqrt{E(Y^2) - E^2(Y)}}. \quad (1)$$

The coefficient $r_{X,Y}$ is the covariance value normalized by standard deviation, and $E$ is the expected value. The

Table 1: Symbol used

| Symbols | Definition |
|---|---|
| T | time interval for processing |
| $T_w$ | strength of the attack traffic |
| $T_y$ | burst length that indicates pulse width |
| $T_x$ | time interval between two consecutive attack pulse |
| $T_z$ | interval between two pulses of high-rate traffic |
| $N_a$, $N_b$ | legitimate traffic |
| $A_a$, $A_b$, $A_c$ | high-rate attack traffic |
| $t_i$ | $i^{th}$ time interval within T |
| $x_i$ | $i^{th}$ instance within $x$ |
| $\delta$ | threshold for attack detection |
| S | sample traffic |
| N | total number of packets within full time interval T |
| n | represents number of packets within the smaller time interval t within T |

use of rank measure correlation using characteristics that cannot be expressed quantitatively but that lend themselves to being ranked. A perfect linear relationship between the ranks yields a rank correlation coefficient of +1 for positive relationship (or -1 for a negative relationship) and no linear relationship between the ranks yields a rank correlation coefficient of 0.

### 3.2 Partial Rank Correlation (PRC)

Partial rank correlation computes correlation between two random variables keeping one or more variables constant. The partial correlation between X and Y with a given set of n controlling variables Z = $Z_1, Z_2, Z_3, ., Z_n$, written as $r_{xy.z}$, is the correlation between residual $r_x$ and $r_y$ resulting from the linear regressions of X with Z and of Y with Z, respectively.

$$r_{xy.z} = \frac{r_{xy} - r_{xz}r_{yz}}{\sqrt{(1 - r_{xz}^2)(1 - r_{yz}^2)}} \quad (2)$$

where $r_{xy}$ denotes the correlation between X and Y with Z constant. The rank correlation coefficient values vary from -1 to +1, where +1 indicates complete linear relationship, -1 indicates a negative linear relationship and 0 indicates no relationship. Partial correlation is a measure of the degree of association between two random variables keeping the third variable constant.

As stated in the algorithm, the sample period T considered for experimentation is divided into n intervals with, N being the total time interval. Three different network traffic instances, namely, $x_i$, $x_j$ and $x_k$ are considered.

**Algorithm 1** The low-rate DDoS attack detection

Input: $x$ represents network traffic with respect to time window T and thresholds $\delta_1$ and $\delta_2$.
Output: alarm information (attack or legitimate).

1) initialization: sample period T = $t_1, t_2, t_3, ..., t_N$ where N is the full time interval. $x_i$, $x_j$, $x_k$ represent three different network traffic instances
2) sample the network traffic $x$ received from upstream router R based on sampling period T
3) compute rank correlation coefficient using Equation (1) or (2) for each sample within T sampling period of ith sample based on traffic features (i.e., source IP, destination IP and protocol).
4) Check whether $RC(x_i, x_j) \geq \delta_1$ or $RC(x_i, x_j, x_k) \geq \delta_2$, if so generate alarm; otherwise, router sends the packets to the next level routers.
5) go to step 2.

Rank correlation coefficient is calculated for each sample using Equation (1) or (2) within a sampling period T of the $i$th sample based on source IP, destination IP and protocol. If the rank correlation of $x_i$ and $x_j$ is greater than threshold $\geq \delta_1$ or rank correlation of $x_i, x_j, x_k$ is greater than threshold $\geq \delta_2$, an alarm will be generated, else the router will send the packet to the next level of routers.

### 3.3 Complexity Analysis

Both spearman rank correlation and partial rank correlation work in quadratic time, $O(n^2T)$, where $n$ is the number of traffic instances within a sample, T is the time interval. Though the complexity is high the rank correlation reveals that:

1) It can discriminate legitimate traffic from attack traffic correctly.
2) PRC can significantly identify low-rate DDoS attack with high linear correlation value.

## 4 Experimental Analysis

In this section, experimental results are presented for both the rank correlation measures using benchmark datasets.

### 4.1 Datasets

The evaluation of any detection method is extremely important before deployment in a real-time network. Two different datasets are used, namely: (i) MIT Lincoln Laboratory and (ii) CAIDA DDoS 2007 dataset. The MIT dataset contains pure legitimate traffic in tcpdump format. It does not contain any attack traffic. Even though

it is old it is still useful and widely used [7]. The CAIDA DDoS 2007 dataset contains 5 minutes of anonymized traffic from a DDoS attack on August 4, 2007. This traffic trace contains only traffic to the victim and responses from the victim. If more than 10,000 attack packets per second are forwarded to the victim machine, it is known as high-rate attack traffic [7]. If up to 1000 attack packets per seconds are forwarded to the victim machine, it is considered low-rate attack traffic [7]. So, low-rate attack may be similar in nature with legitimate traffic.

## 4.2 Results

Initially the total time interval is splits into 10 second subintervals. Three packet attributes are used during the experiment, namely, source IP, destination IP and protocol. For a victim-end based detection system, source IP is important, especially to find source hosts even though they may be spoofed. The destination IP is also important to identify and to estimate the traffic flowing to a particular target. The attribute protocol is added to identify the attack type. Each sample is processed one at a time. The rank correlation measure is applied to find the linear relationship between legitimate and attack traffic.

Figures 1, 2 and 8 show the probability density of legitimate and attack traffic when using the MIT dataset as legitimate traffic and the CAIDA dataset as attack traffic. Following Xiang et. al [11], the MIT dataset is considered legitimate traffic in our experiment. Spearmen rank correlation and partial rank correlation are computed on the two different datasets, namely, MIT legitimate and CAIDA attack dataset. These attack traffic instances are assumed to satisfy the low-rate attack properties. Results for both legitimate and attack traffic instances are reported for both rank correlation measures in Figures 3 and 4 for PRC and 5, 6 and 7 for SRC. Correlation values for legitimate traffic and attack traffic are reported in Table 2. While using Spearmen rank correlation and partial rank correlation, Figures 7 and 8 report results for mixed traffic (i.e. both legitimate and attack traffic). We see that PRC can discriminate effectively legitimate traffic from attack traffic with rank correlation with min = 0.8 and max = 1.0. Figure 9 reports the attack and legitimate traffic rank values when using SRC and PRC. Figure 10 shows the spacing between legitimate traffic and attack traffic when using SRC and PRC. It seems that PRC has higher spacing than SRC. Better results are observed for those ranges of rank correlation values and are reported in Table 2, when detecting low-rate DDoS attacks.

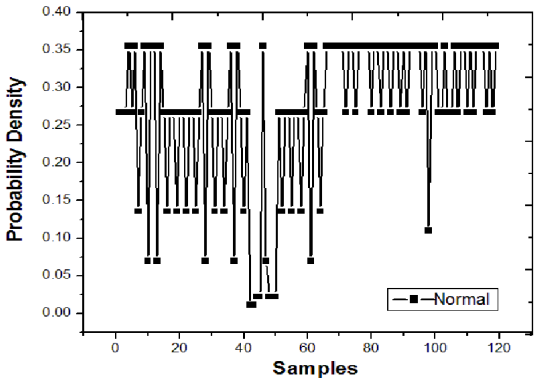


Figure 2: Probability density for legitimate traffic

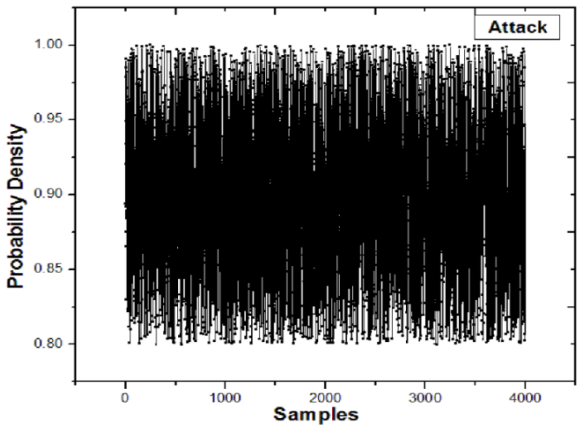


Figure 3: Probability density for attack traffic

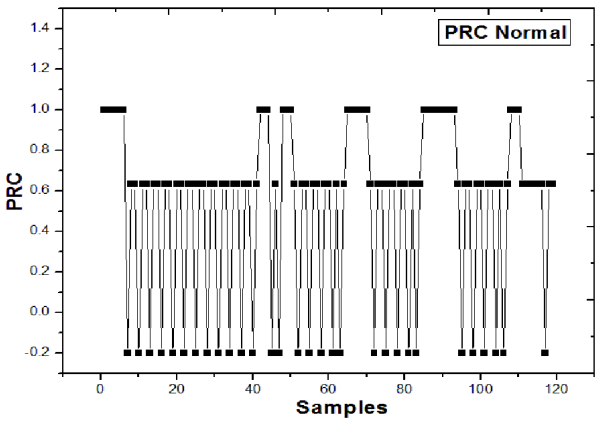


Figure 4: Partial rank correlation for legitimate traffic

## 4.3 Discussion

Based on the analysis, we make the following observations.

1) SRC uses a small number of parameters to estimate rank correlation.

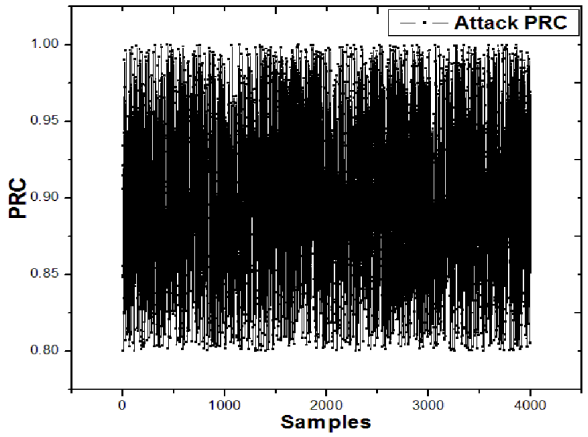


Figure 5: Partial rank correlation of attack traffic

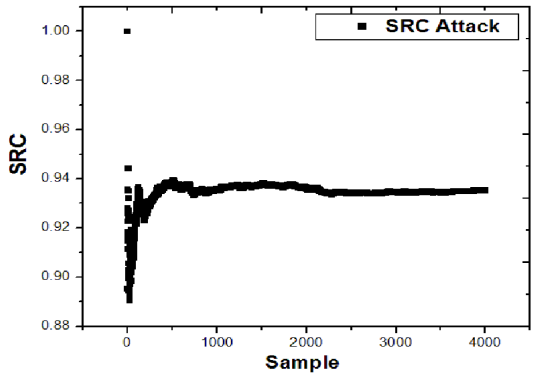


Figure 8: Spearmen rank correlation for attack traffic

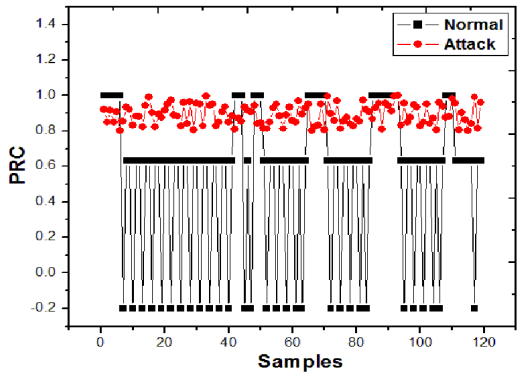


Figure 6: Partial rank correlation of legitimate and attack traffic

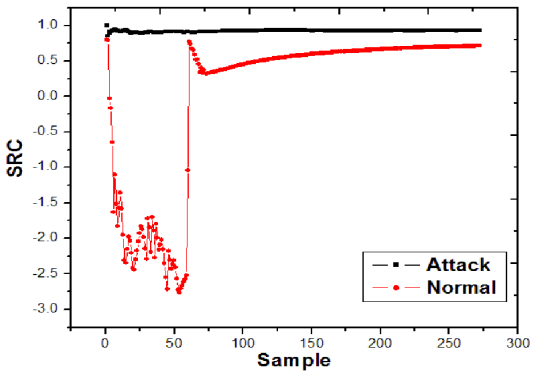


Figure 9: Spearmen rank correlation of legitimate and attack traffic

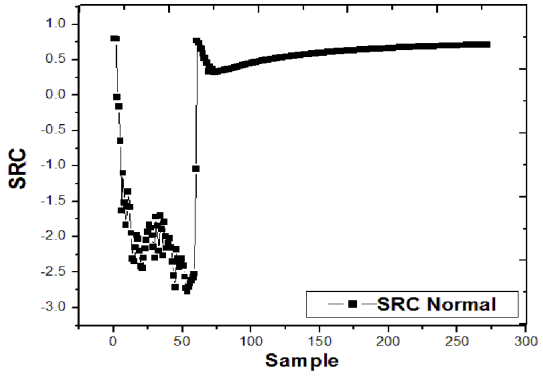


Figure 7: Spearmen rank correlation for legitimate traffic

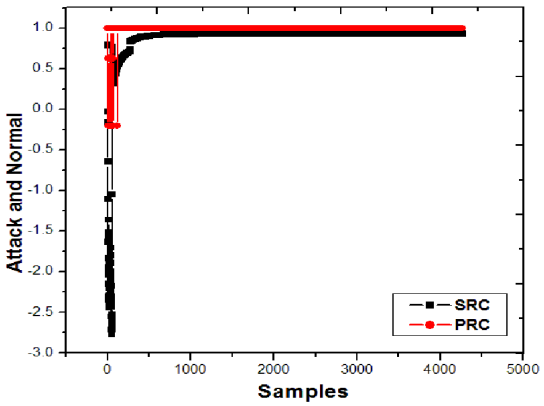


Figure 10: Rank Correlation for attack and legitimate traffic using SRC and PRC

2) Both correlation measures are capable of differentiating legitimate traffic from malicious traffic correctly.

3) The partial rank correlation measure is effective in reducing false alarms in a victim-end defence system. It is due to higher spacing between legitimate and attack traffic.

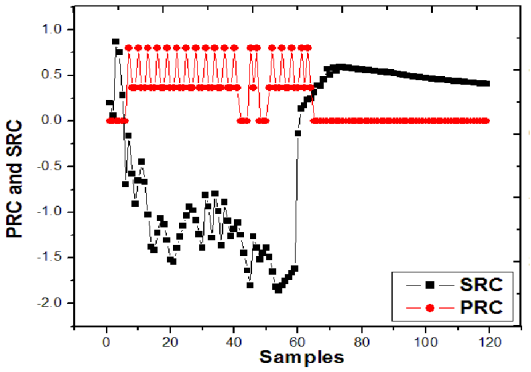


Figure 11: Shows difference between PRC and SRC

Table 2: Ranges of correlation values

| Rank correlations | Traffic type | Minimum | Maximum |
|---|---|---|---|
| PRC | normal-normal | -0.2 | 1.0 |
| SRC | normal-normal | -2.7 | 0.8 |
| PRC | attack-attack | 1.0 | 0.8 |
| SRC | attack-attack | 0.8 | 1.0 |
| PRC | normal-attack | -0.2 | 1.0 |
| SRC | normal-attack | -2.8 | 0.9 |

## 5 Conclusion and Future Work

In this paper, we have presented an empirical study of rank correlation used to detect low-rate distributed DoS attacks. PRC and SRC both are used to effectively differentiate legitimate traffic from malicious traffic. Our experimental study, show that PRC is more effective than SRC in differentiating legitimate traffic from attack traffic. Development of a traceback mechanism to support low-rate DDoS attack is underway.

## References

[1] Monowar H Bhuyan, D K Bhattacharyya, and J K Kalita, "An empirical evaluation of information metrics for low-rate and high-rate ddos attack detection," *Pattern Recognition Letters*, vol. 51, pp. 1–7, 2015.

[2] Shuyuan Jin and Daniel S Yeung, "A covariance analysis model for ddos attack detection," in *Communications, 2004 IEEE International Conference on*, vol. 4, pp. 1882–1886, 2004.

[3] Haiqin Liu and Min Sik Kim, "Real-time detection of stealthy DDoS attacks using time-series decomposition," in *Communications (ICC), 2010 IEEE International Conference on*, pp. 1–6. IEEE, 2010.

[4] Gabriel Maciá-Fernández, Jesús E Díaz-Verdejo, and Pedro García-Teodoro, "Mathematical model for low-rate dos attacks against application servers," *Information Forensics and Security, IEEE Transactions on*, vol. 4, no. 3, pp. 519–529, 2009.

[5] Rejo Mathew and Vijay Katkar, "Software based low rate dos attack detection mechanism," *International Journal of Computer Applications*, vol. 20, no. 6, pp. 14–18, 2011.

[6] Jelena Mirkovic and Peter Reiher, "A taxonomy of ddos attack and ddos defense mechanisms," *ACM SIGCOMM Computer Communication Review*, vol. 34, no. 2, pp. 39–53, 2004.

[7] David Moore, Colleen Shannon, Douglas J Brown, Geoffrey M Voelker, and Stefan Savage, "Inferring internet denial-of-service activity," *ACM Transactions on Computer Systems (TOCS)*, vol. 24, no. 2, pp. 115–139, 2006.

[8] Zhang Sheng, Zhang Qifei, Pan Xuezeng, and Zhu Xuhui, "Detection of low-rate ddos attack based on self-similarity," in *2010 Second International Workshop on Education Technology and Computer Science*, pp. 333–336, Washington, DC, USA, 2010.

[9] Zhiyuan Tan, Aruna Jamdagni, Xiangjian He, Priyadarsi Nanda, and Ren Ping Liu, "Denial-of-service attack detection based on multivariate correlation analysis," in *Neural Information Processing*, pp. 756–765, 2011.

[10] Wei Wei, Feng Chen, Yingjie Xia, and Guang Jin, "A rank correlation based detection against distributed reflection dos attacks," *Communications Letters, IEEE*, vol. 17, no. 1, pp. 173–175, 2013.

[11] Yang Xiang, Ke Li, and Wanlei Zhou, "Low-rate ddos attacks detection and traceback by using new information metrics," *Information Forensics and Security, IEEE Transactions on*, vol. 6, no. 2, pp. 426–437, 2011.

**Arindom Ain** received his Master's degree in Computer Application from Dibrugarh University, Assam, India in 2011. He is System Administrator at Kaziranga University, India. He is pursuing his Ph.D. from Sikkim Manipal University.

**Monowar H. Bhuyan** is an assistant professor in the Department of Computer Science and Engineering at Kaziranga University, Jorhat, Assam, India. He received his Ph.D. in Computer Science & Engineering from Tezpur University in 2014. His research areas include data mining, cloud security, computer and network security. He has published 20 papers in international journals and referred conference proceedings.

**Dhruba K. Bhattacharyya** received his Ph.D. in Computer Science from Tezpur University in 1999. Currently,

he is a Professor in the Computer Science & Engineering Department at Tezpur University. His research areas include data mining, network security and bioinformatics. Prof. Bhattacharyya has published 220 research papers in leading international journals and conference proceedings. Dr. Bhattacharyya also has written/edited 10 books.

**Jugal K. Kalita** is a professor of Computer Science at the University of Colorado at Colorado Springs. He received his Ph.D. from the University of Pennsylvania in 1990. His research interests are in natural language processing, machine learning, artificial intelligence, bioinformatics and applications of AI techniques to computer and network security. He has published 150 papers in international journals and referred conference proceedings and has written two technical books.