The Design, Data Flow Architecture, and Methodologies for a Newly Researched Comprehensive Hybrid Model for the Detection of DDoS Attacks on Cloud Computing Environment

Anteneh Girma, Kobi Abayomi and Moses Garuba

Abstract As cloud computing services are becoming more practical and popular for the sake of its convenience and being more economical, its' security vulnerability has been the continuous threat for both cloud service providers and clients. The more the financial benefits of these services became so attractive and the need for uninterrupted services grows, the distributed denial of services (DDoS) attacks that degrade and down its' service availability has been the major security concern. While researchers try to address this security threat and come up with lasting solutions for early detection of DDoS attacks, the degree of these attacks is getting higher and very sophisticated. The changing and aggressive nature of the attacks make it very severe threat and difficult to easily find remedy. On this research paper, we are presenting the design and dataflow architectures for a solution model that could contribute in resolving one of the major cloud security threat, the early detection of DDoS attacks, using its hybrid approach.

Keywords Cloud security · Cloud design architecture · Detection of DDoS attacks · Conditional entropy

1 Introduction

A Distributed Denial of Service (DDoS) attack is a flooding attack on a certain target network or server system that is launched through many compromised

A. Girma(✉) · M. Garuba
Computer Science Department, Howard University, Howard, USA
e-mail: agirma@howard.edu

K. Abayomi
Mathematics Department, Howard University, Howard, USA

S. Latifi (ed.), *Information Technology New Generations*,
Advances in Intelligent Systems and Computing 448,
DOI: 10.1007/978-3-319-32467-8_34

systems called zombies. DDoS attacks are launched with the intention of service disruption by depleting either the network resources or the host servers. DDoS attacks are caused by sending flood of requests and preventing legitimate users from accessing network services or data center resources. In order for attackers to create large botnets of computers (Zombies) under their control, they have two options: the more common option of using specialized malware to infect the machines of users who are unaware that their machines are compromised, or the relatively newer option of amassing a large number of volunteers willing to use DDoS programs.

Anytime attackers want to launch a DDoS attack, they can send messages to their botnet's C&C servers with instructions to perform an attack on a particular target, and any infected machines communicating with the contacted C&C server will comply by launching a coordinated attack.

2 Related Works

Many researches have been conducted and as many number of different DDoS detection techniques have been proposed. Among these was a simple and efficient hidden markov model scheme for host based anomaly intrusion detection [1]. An entropy based anomaly detection system to prevent DDoS attacks in cloud was reviewed, explored, investigated and proposed as an alternative solution [2]. After investigating the correlativity changes of monitored network features during flood attacks, a covariance-Matrix modelling and detecting various flooding attacks was proposed [3].

An experimented result was also analyzed and presented to support a model that was instrumental to propose a model to detect flood based DoS attack in cloud environment. It provided research results that support how effectively the flood attacks are detected [4]. Researchers also discussed how entropy based collaborative detection of DDoS attacks on community networks could effectively works in theory by applying information theory parameter called entropy rate [5]. Different types of DDoS attacks at different layers of OSI model were discussed and presented, and finally, analyzed the impact of DDoS attacks on cloud environment [6]. The analysis of covariance model for DDoS Detection was discussed and the researchers described how the method can effectively differentiate the traffic between the normal and attack traffic. They also showed how the linear complexity of the method makes its real time detection practical [7].

Another detecting solution framework to predict multi-step attacks before they pose a serious security risk is by using hidden markov model. The study based the real time intrusion prediction on optimized alerts since alerts correlations play a critical role in prediction [8]. The design of two independent architectures for HTTP and FTP which uses an extended hidden semi-markov model to describe the browsing habits of web searchers and detecting DDoS attacks were discussed

and investigated [9]. A survey of different mechanism of DDoS attacks, its detection, and the various approaches to handle them was discussed and explored, to enable the clients review and understand those different parameters having impacts in their decision making process while selecting the right DDoS detecting scheme [10].

The scopes of DDoS flooding attack problems and attempts to combat them have been explored by categorizing the DDoS flooding attacks and classifying existing countermeasures based on different parameters [11]. A comprehensive survey presented DDoS attacks, detection methods, detection tools used in wired networks and internet, and future research direction [12]. The Security problem associated with cloud computing becomes more complex due to entering of new dimensions in problem scope related to its own main attributes. Researchers also proposed a detection scheme based on the information theory based metrics. The proposed scheme has two phases: Behavior monitoring and Detection. Based on the observation, Entropy of requests per session and the trust score for each user is calculated [13]. DDoS attacks could be detected using the application of Dempster Shafer Theory. The theory was applied to detect DDoS threat in cloud environment. It is an approach for combining evidence in attack conditions [14]. The effectiveness of an anomaly based detection and characterization system highly dependent on accuracy of threshold value setting. And this approach described a novel framework that deals with the detection pf variety of DDoS attacks [15]. Cloud specific Intrusion Detection System was proposed and described a defense mechanism against the DDoS attacks. This defense mechanism discusses how to detect the DDoS attack before it succeeds [16]. Effectively detecting the bandwidth limit of a cloud network and the bandwidth currently in use helps to know when a DDoS attacks begin [17]. An approach described based on fundamentals of information theory specifically Kolmogorov complexity to detecting distributed denial of service (DDoS) attacks was proposed. Despite its complexity the scheme enabled early detection [18].

3 Strategy of DDoS Attacks

Even though there are many different types of DDoS attacks, all these attacks have the same attacking strategies and involve four stages of scenarios: ***Selecting agents*** where agents that will perform the attack is chosen, ***Compromising agents*** where the vulnerabilities of the zombie machines will be exploited and the attack codes are transferred, ***Communication*** where attackers will communicate with the handlers to determine which zombie machines are up and running and decide the other parameters, and finally ***Attac*****k** where the hackers initiate the attacks and other attacking parameters could be adjusted.

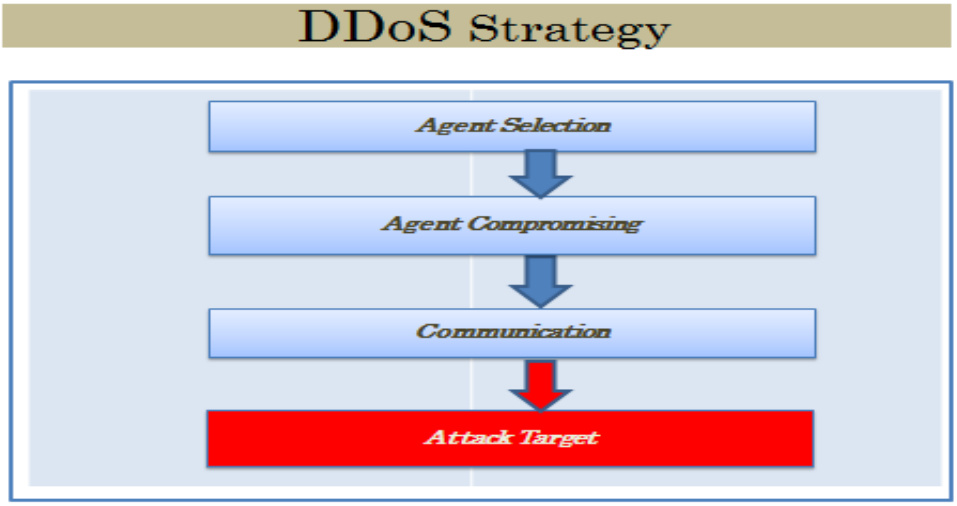

Fig. 1 Description of DDoS Strategy

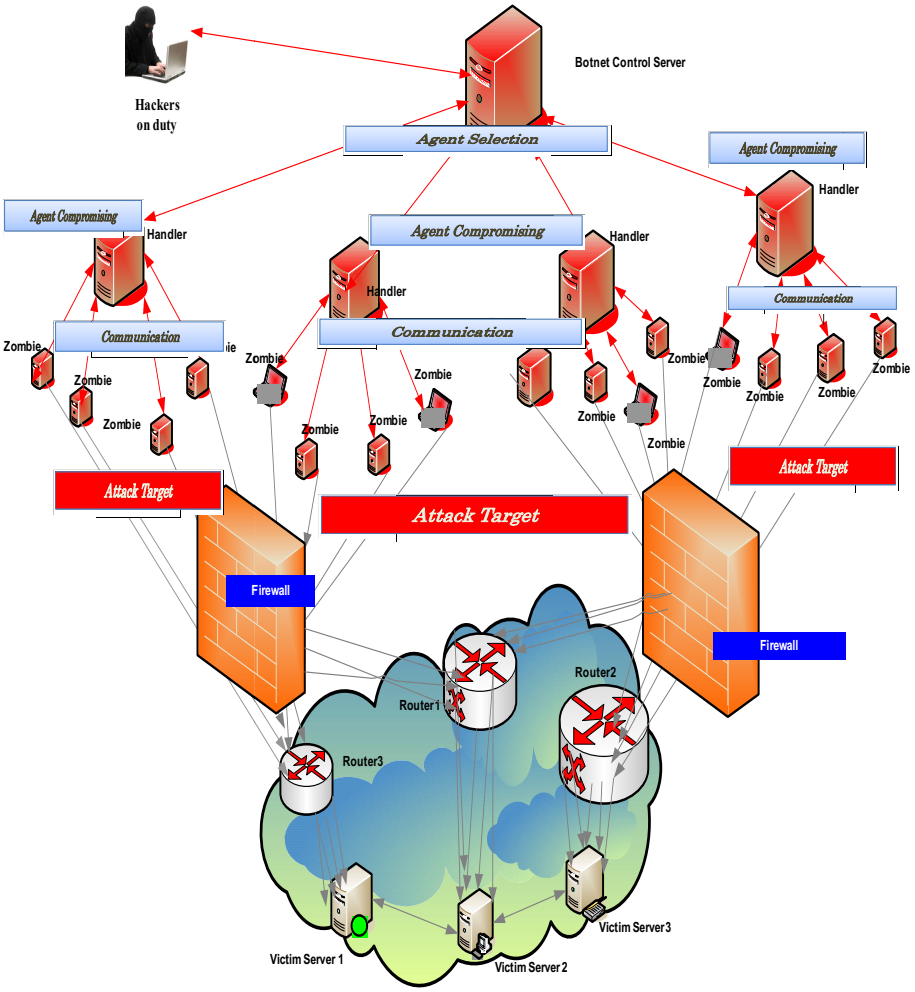

Fig. 2 Detain diagram Sources of an organized flooded DDoS attacking strategies [19]

4 Design Architecture of the New Detection System

The design Architecture includes all entities and structures involved in the overall DDoS attack and detection processes. It clearly shows where the attack packet get launched from, filtered and analyzed, and finally categorized as an attack or not attack.

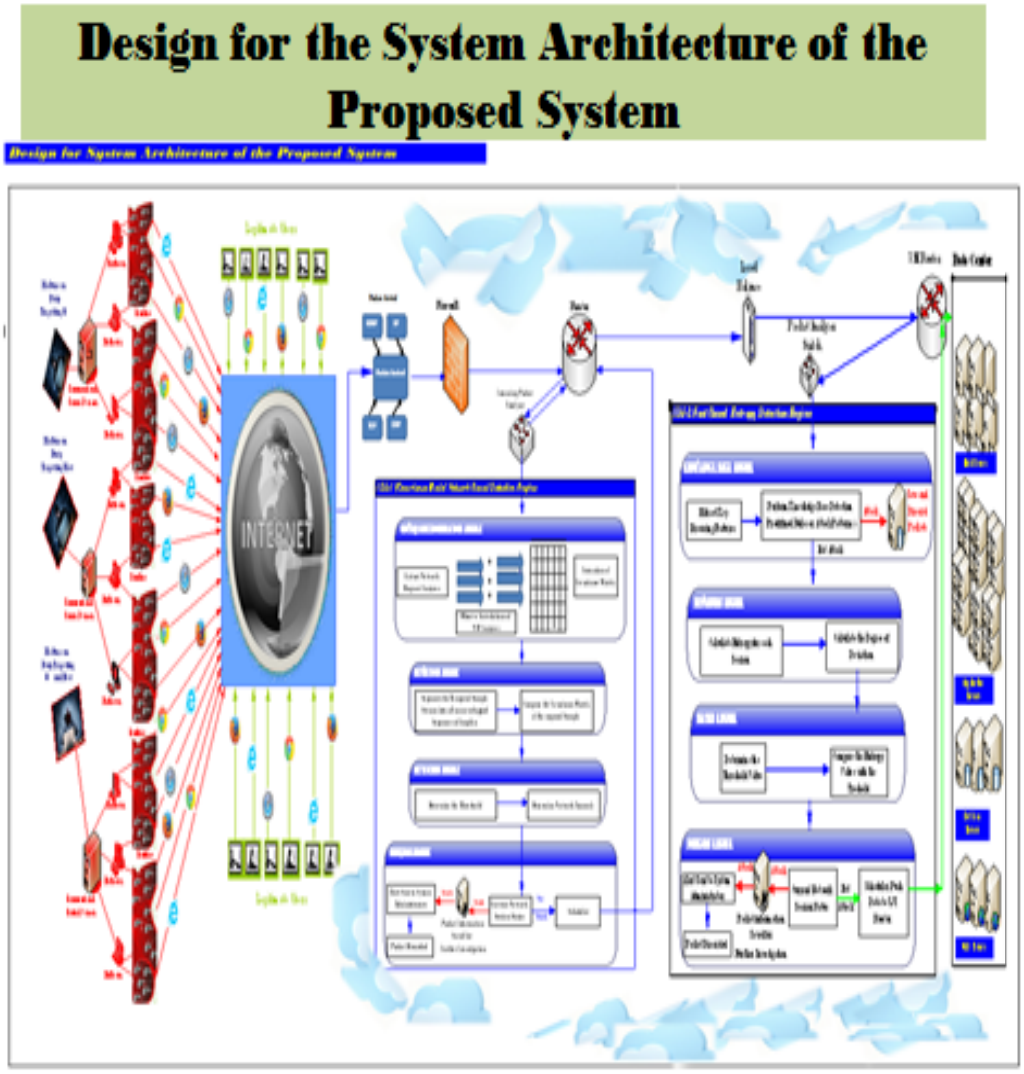

Fig. 3 Design Architecture of the Detection system

5 Data Flow Architecture of the Proposed System

The data flow architecture is designed on the following data flow diagram (figure 4) in more simplified way. The major functional procedures are mentioned and included and give a clear indication for any reviewers regarding how the packets get traversed throughout the system.

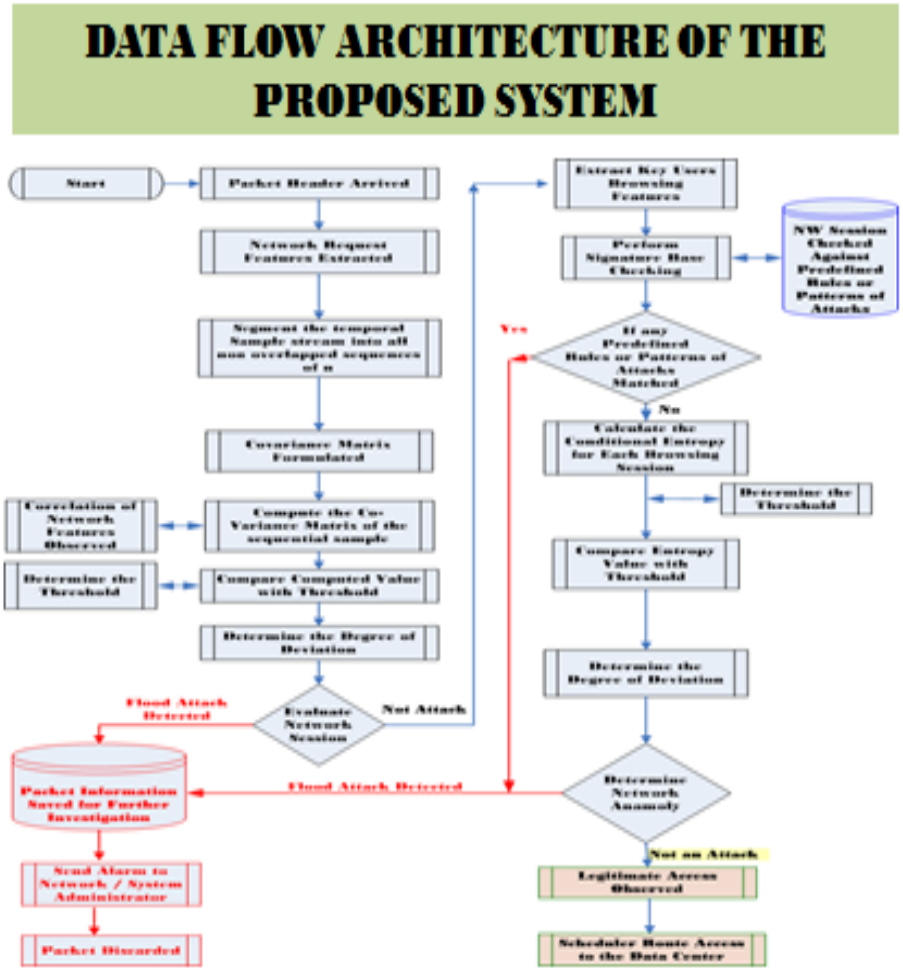

Fig. 4 Data Flow Architecture

6 Algorithm

Our Hybrid Algorithm is consists of two stochastic models. The first one is applied on the selected sample network features and passes the results (attack packet) to the second one that will analyze the same packet but with different selected browsing features that handles mainly the false alarm rate which is one of our main targets. The packet will finally be granted access to the cloud data center if it found not to be an attack after the second packet analysis procedure.

IP Packet Observed

P^* Sample Network features selected
Q^* Sample Browsing features

For Each session of P^* Network features selected

Compute the Kendall tau value using Concordance and Dis-concordance (Using formula 3)

Covariance Matrix C(X) will be modulated (Using formula 4)

The Deviation Function D(.) is computed

Let V = D(.)
Assign V a value equals to 0 or 1

If (V = 0) (Packet not an attack) then

(Perform the next packet analysis procedure)
Browsing features will be considered

For range of interval time observation (t = 1..T)

Compute entropy $H(\mathbf{X})=\sum_{\mathrm{X}} f(\mathbf{X})\ log(f(\mathbf{X})$

Compare the entropy rate $H(\mathbf{X})$ with the threshold

If $H(\mathbf{X})$ is equal or less than the threshold value

Packet discarded
Attack Alarm will be sent to Admin.

End If
End

Else (If V = 1) (Packet an attack)

(False Alarm Analysis is done using conditional entropy)
Compute Conditional Entropy (Using formula 12)

$H(\overrightarrow{\mathbf{x}}|\mathbf{X})=H(\overrightarrow{\mathbf{x}},\mathbf{X})-H(\mathbf{X})$

If $H(\overrightarrow{\mathbf{x}}|\mathbf{X})$ is equal or less than the threshold

Packet discarded
Attack Alarm will be sent to Admin

Else
Grant Access to the Cloud Data Center

End If
End If

End

7 Detection Approaches

Let $\mathbf{X}$ be a $p \times T$ multivariate vector where p is number of network 'features' or variables and T is the number of (discrete time interval) observations. Our objective is to identify the presence of atypical dependence in a sample after establishing a baseline dependence.

In statistical terms we evaluate the dependency among X for an ordinary, non-attack, regime – Let's call it $T_0 = T(\mathrm{X}_0)$, where T is some statistic of the multivariate data and X_0 is the baseline or non-attack, training, data. Then the task is to evaluate the 'distance', via this statistic, between the training data and 'new' data; large values of this distance indicate an 'attack'. In notation

$$D(T_0, T(\mathbf{X})) \tag{1}$$

7.1 Multivariate Correlation

Out of $p^* \leq p$ features the covariance matrix is only taking two features at a time, then testing after imposing a threshold and/or using 0, 1 as the distance. For ordinary $p \times p$ correlation/covariance matrix has entries proportional to:

$$Cov(X_i, Xj) = \boldsymbol{E}[X_i \cdot X_j] \tag{2}$$

there are $\frac{p\ (p+1)}{2}$ unique entries in such a matrix where i and j are now multivariate indices, for each of t: $len\,(i) = \mathrm{len(j)}$. The expectation above, then, should be a scalar and the matrix collecting these will be of dimension $p \times p$.

7.1.1 Kendall's Tau for Bivariate Correlation, Multiple Correlation

The population version of Kendall's Tau is:

$$\tau = P(Concordance) - P(Discordance) \tag{3}$$

These can be collected pairwise into a matrix as well, say, by calculating:

$$\tau_{ij} = P(\{Concordance\} - P(\{Discordance\})[X_i, X_j] \tag{4}$$

or into a matrix of reduced dimension

$$\tau_{ij} = P(\{Concordance\} - P(\{Discordance\})[X_i, X_j] \tag{5}$$

in analogy with the covariance and the same thresholding, can be done. Even to the use of Chebyshev's inequality to impose probability limits on the statistical values.

The sample version of Kendall's Tau is calculated by letting

$$P(\{Concordance\}) = \#\{X_i, X_j > 0 \ \& \ X_i, X_j < 0\} \tag{6}$$

$$P(\{Disconcordance\}) = \#\{X_i > 0, X_j < 0 \ \& \ X_i < 0, X_j > 0\} \tag{7}$$

Where # is the number of instances over all indices i, j.

7.1.2 Entropic Approach

The entropy of a multivariate, discrete, random variable is:

$$H(\mathbf{X}) = \sum_{\mathrm{X}} f(\mathbf{X}) \, log \, (f(\mathbf{X})) \tag{8}$$

This is the entropy across all features of the data, in terms of the model across the entire dimension of the multivariate vector H (.) is a function from $\mathrm{R}^p \rightarrow \mathrm{R}$. In the world of the data which for the moment we assume arrive from a stable process but possible 'attack' regime - we have $\mathbf{x}1, \mathbf{x}2, ..., \mathbf{x}T$ as the multivariate observations. Call these $\overrightarrow{\mathbf{x}}$. So our estimate of the entropy will rely on $f(\cdot)$ - which can be a particular model or an empirical estimator - and be estimated across $t = 1, ..., T$ units of time indexed observations is given by:

$$H\,(\overrightarrow{\mathbf{x}}) = \sum_{t=1}^{\mathrm{T}} f(\mathbf{x}_t) \, log \, (f(\mathbf{x}_t)) \tag{9}$$

Consider setting

$$H(\mathbf{X}) = T_0(\mathbf{X}) \tag{10}$$

as the *baseline* entropy among the features. Then we can consider

$$D(T_0, T\,(\overrightarrow{\mathbf{x}})) = H(\mathbf{X}) - H\,(\overrightarrow{\mathbf{x}}) \tag{11}$$

7.1.3 Conditional Entropy

Alternately we compute:

$$H(\overrightarrow{\mathbf{x}} | \mathbf{X}) = H(\overrightarrow{\mathbf{x}}, \mathbf{X}) - H(\mathbf{X}) \tag{12}$$

as measures of (an increase in) dependency among suspected attack data. In (12), a large distance between the baseline entropy and the entropy of the data (i.e. calculated across some time steps) is the signal for an attack. In (13), we use the entropy itself (i.e. the dependency between the features under the model and those of the data) as the measure of an attack. In (13) this measure is function of the model $f(\mathbf{X})$ on the 'training'

data; perhaps an empirical estimator $\hat{f}(\overrightarrow{\mathbf{x}})$ or the probability of the data given the model $f(\overrightarrow{\mathbf{x}})$. If we eschew an empirical estimator and calculate $f(\overrightarrow{\mathbf{x}})$ in (13) we can think of this as similar to a *likelihood based approach*, but where we access the likelihood via the entropy function. In a sense, this method is more 'complete': the mass from the entire probability distribution (via the model f and estimator $\hat{f}$) is used and not just the expectation.

8 Conclusion

On this paper we have presented and discussed the design architecture, data flow architecture, and algorithm of our newly hybrid approached solution model that could be a better choice in terms of detecting the DDoS flood attacks. The preliminary results showed that our detection technique is very encouraging and promising. We shall present the details of our research results, data analysis, and our recommendation for future works ideas on the next upcoming conference.

References

1. Hu, J., Yu, X., Qiu, D., Chen, H.-H.: A simple and efficient hidden markov model scheme for host-based anomaly intrusion detection. IEEE Network (February 2009)
2. Syed Navaz, A.S., Sangeetha, V., Prabhadevi, C.: Entropy based anomaly detection system to prevent DDoS attacks in cloud. International Journal of Computer Applications (0975-8887) (January 2013)
3. Yeung, D.S., Wang, X.: Covariance-matrix modeling and detecting various flooding attacks. IEEE Transactions on Systems, MAN, Cybernetics- Part A: Systems and Humans **37**(2) (March 2007)
4. Ismail, M.N., Aborujilah, A., Musa, S., Shahzad, A.: Detecting Flooding based DoS attack in cloud computing environment using covariance matrix approach. In: ICUIMC (IMCOM) (2013)
5. Yu, S., Zhou, W.: Entropy-Based Collaborative detection of DDoS attacks on community networks. In: Sixth annual IEEE International Conference on Pervasive Computing and Communications (2008)
6. Sha, J.J., Malik, L.G.: Impact of DDoS attacks on cloud environment. International Journal of Research in Computer and Communication Journal **2**(7) (July 2013)
7. Jin, S., Yeung, D.S.: A covariance analysis model for DDoS attack detection. In: IEEE Communications Society (2004)
8. Sendi, A.S., Dagenais, M., Jabbarifar, M.: Real time Intrusion prediction based on optimized alerts with hidden markov model. Journal of Networks **7**(2) (February 2012)
9. Ankali, S.B., Ashoka, D.V.: Detection architecture of application layer DDoS attack for internet. Advanced Networking and Applications **3**(1), 984–990 (2011)

10. Er. Kakkar, S., Er. Kumar, D.: A survey on distributed denial of services (DDoS). International Journal of Computer Science and Information Technologies **5**(3) (2014)
11. Patcha, A., Park, J.-M.: An overview of anomaly detection techniques: Existing solutions and latest technological trends. Science Direct, Computer Networks **51** (2007)
12. Bhuyan, M.H., Kashyap, H.J., Bhattacharyya, D.K., Kalita, J.K.: Detecting distributed denial of service attacks: methods, tools and future directions (December 2012). http://www.garykessler.net/library/ddos.html
13. Renuka Devi, S., Yogesh, P.: Detection of application layer DDoS attacks using information theory based metrics. CS & IT-CSCP, pp. 217–223 (2012)
14. Lonea, A.M., Popescu, D.E., Tianfield, H.: Detecting DDoS attacks in cloud computing environment. International Journal of Computing and Communication **8**(1), 70–78 (February 2013). ISSN 1841-9836
15. Gupta, B.B., Misra, M., Joshi, R.C.: An ISP level solution to combat DDoS attacks using combined statistical based approach. Journal of Assurance and Security **2**, 102–110 (2008)
16. Goyal, U., Bhatti, G., Mehmt, S.: A dual mechanism for defeating DDoS attacks in cloud computing model **2**(3) (March 2013)
17. Panda, B., Bhargava, B., Pati, S., Paul, D., Lilien, L.T., Meharia, P.: Monitoring and managing cloud computing security using denial of service bandwidth allowance (2012)
18. Kulkarni, A.B., Bush, S.F., Evans, S.C.: Detecting distributed denial-of-service attacks using kolmogorov complexity metrics. In: GE Research & Development Center (February 2002)
19. Girma, A., Abayomi, K., Garuba, M., Li, J., Liu, C.: Analysis of DDoS attacks and an introduction of a hybrid statistical model to detect DDoS attacks on cloud computing environment. In: ITNG-2014 (April 2014)