Asian Journal of Computer Science and Technology
ISSN: 2249-0701 Vol.8 No.S2, 2019, pp. 61-65


# Network Traffic Analysis in Cloud:A Survey

**K. Ruckshana[1] and G. Ravi[2]**
[1]Research Scholar, [2]Associate Professor & Head
[1&2]Department of Computer Science, Jamal Mohamed College, Tiruchirappalli, Tamil Nadu, India
E-Mail: ruckshanasharief6912@yahoo.com

***Abstract*** **- A data center (DC) denotes to any huge faithful group of computers that is retained and operated by an organization. Data centers of numerous sizes are being made and hired for a dissimilar set of resolves today. On the one hand, big universities and isolated enterprises gradually consolidating their IT services within on-site data centers comprising hundreds to thousands of servers. On the other hand, huge online service providerssuch as Google, Microsoft, and Amazon quickly constructing geographically varied cloud data centers often have more than 10K servers; to offer a variation of cloud-based services such as Email, Web servers, Gaming, Storage, and Instant Messaging. Though there is great interest in planning developed networks for data centers, very little is identified about the network-level traffic characteristics of present data centers. In this paper, we focused on a study of the network traffic in data centers and defining the anomaly detection system in secure cloud computing environment.**
***Keywords:*** **Data Center, Cloud Computing, Network Traffic, Anomaly Detection, Risk Aware Network**

## I. INTRODUCTION

Cloud computing is a common term for everything that includes delivering hosted services over the Internet. These services are largely separated into three categories: Infrastructure-as-a-Service (IaaS), Platform-as-a-Service (PaaS) and Software-as-a-Service (SaaS). The name cloud computing was motivated by the cloud symbol that is frequently used to represent the Internet in flowcharts and diagrams.

A network is a gathering of computers, servers, mainframes, peripherals, network devices, or other devices associated with one another to permit the sharing of data. An excellent instance of a network is the Internet, which joins millions of people all over the world.

Network traffic or data traffic is the amount of data moving through a network at a given point of time. Network data in computer networks are typically encapsulated in network packets which offer the load on the network. Network traffic is the key component for network traffic measurement, network traffic control and simulation.

1. Network traffic control - managing, prioritizing, controlling or decreasing the network traffic
2. Network traffic measurement - evaluating the amount and type of traffic on a specific network
3. Network traffic simulation - to measure the effectiveness of a communications network

Accurate analysis of network traffic offers the organization with the network security as an advantage. An unusual amount of traffic in a network is a probable sign of an attack[1]. Network traffic reports offer valuable insights into preventing such attacks.

Traffic classification defines methods of classifying traffic based on features passively perceived in the traffic according to particular classification goals. One might have a coarse classification aim, i.e., whether it's transaction-oriented, bulk-transfer, or peer-to-peer file sharing;or one might have a fine-grained classification goal, i.e., the exact application denoted by the traffic. Traffic features could comprise the port number, application payload, or temporal, packet size, and addressing the characteristics of the traffic. Classification methods contain exact matching, e.g., port number or payload, heuristic or machine learning.

## II. RELATED WORK

Chuvakin *et al.*,[2] discovered the dissimilar types of correlation that could be applied for security analysis. One instance is the Rule based Correlation (RBC) technique which uses an If-Else statement to check for a particular attack. It needs exact and approximate knowledge about the attack. Later the attack is identified; it will be characterized using the If-Else statements which are defined. There can be dissimilar potential alerts that the administrator canget an action will be applied instantly. Using this model for the Project Coordinates can simply detect what type of attack occurred since it uses pre-defined knowledge of the attack. Applying such rules using If-Then statements would suitably consolidate the attack as soon as it is identified.

Pineda and Yatco *et al.*,[3] made a software-based Log Consolidation and Incident Management that can be amended to a particular event management. The system can build reports out of the logs, since SIEM products lack log consolidation and incident management structures.

Bernardo and Valencia *et al.*,[4] started a research about Artificial Intelligence-like IDS which can sort and generate rules from the logs of a honey pot. The proposed outcome is a self-creating signature-based IDS.

Gordon *et al.,* [5] conferred the execution of Alert Correlation. This study defined the simple variable between Alert Correlation and Event Correlation. It also discovered Alert Correlation as an executive for the Project Coordinate. The variance between the two is that former handles and examines any abnormalities or anomalies that occur inside the network. The second technique, examines neutral events that are captured and recorded. Alert Correlation is intended through the use of marking certain alerts and collecting them together. This includes header information like IP address, the number of alerts, and the destination stated in the alert. Event Correlation is designed through joining the logs if they are correlated or not. This would exactly group the connected logs into one event for the administrators to examine.

## III. TRAFFIC CLASSIFICATION

Traffic Classification is the first step to analyze and detect different types of applications flowing in a network. Through this technique, internet service providers or network operators can handle the complete performance of a network.

*A. Sensitive Traffic:* Sensitive traffic is a type of traffic in which the operator need transmit on time. This contains VoIP, online gaming, video conferencing, and web browsing. Traffic management schemes are usually customized in such a way that the quality of service of these particular uses is guaranteed, or at slightest priority over other classes of traffic. This can be accomplished by the lack of shaping traffic class, or by prioritizing sensitive traffic beyond other classes.

*B. Best-Effort Traffic:* Best effort traffic is a type of non-negative traffic. The traffic in which the ISP regards is not sensitive to Quality of Service metrics (jitter, packet loss, latency). A usual example would be peer-to-peer and email applications. Traffic management schemes are usually custom-made, so best-effort traffic gets what is left after sensitive traffic.

*C. Undesired Traffic:* This category is usually limited to the delivery of spam and traffic generated by worms and other malicious attacks. In some networks, this description can contain such traffic as non-local VoIP (for example, Skype) or video streaming services to defend the market for the 'in-house' services of the same type. In these cases, traffic classification mechanism detects traffic permitting the network operator to either block this traffic completelyor strictly hamper its operation.

## IV. TRAFFIC CLASSIFICATION PARAMETERS

Organize activity parameters are usually considered in the investigation of bundle and movement characterization methods.

*A. Packet Size:* Bundle size is one form of movement characterization. The massive majority of the activity volumes on the Internet can be ordered into either little parcels or expansive bundle sizes. The massive parcel size is usually connected with greater connection use. Basically 20% of the associations on the Internet are in charge of 80% of the activity, most of the part comprised of elephant bundles.

*B. Duration:*Term of bundle streams is another form of parcel arrangement. Contingent upon the application, a fleeting bundle can last from a duo of milliseconds up to a duo of minutes. Lasting parcelscan last from a duo of minutes up to limited hours. There are instant relations between bigger parcel sizes and longer lengths.

*C. Confidence Interval (CI):*CI is a populace associated parameter, which is an interim estimator. Certainty interims are used to provide a gauge on how a sample is dependable. For an outrageous numerous specimen space (for an instance) the activity designs on the Internet spine is probable that one wants to screen the lines for a drawn out stretch of time (e.g., months or years) and after that run movement order method above the spared follows, or use the little example space with a guide of a certainty interim estimator. A certainty interim is greater than 95% of usualdecent estimation.

## V. TRAFFIC CLASSIFICATION METHODS

*A. Port Based Approach*

Port-construct technique is located in light of the element that particular application administrations use IANA relegated port numbers. This strategy practices the accompanying insufficiencies. In the first place, P2P applications use arbitrary or dynamic port numbers. Secondly, regular administration ports might be used by different administrations for instance malware. At Third, there are port numbers other than relegated. At Fourth, it is coarse-grained. At last, port numbers can be protected by transport layer or IP parcel encryption.

*B.Payload Based Approach*

Payload based strategy refers to deep bundle review DPI system, which uses static application marks as a part of the payload to differentiate conventions. Statefull Packet Inspection creates utilization of measurable properties of payload in packets. DPI is extremely affected by encryption since the plaintext marks turn imperceptible.

*1. Deep Packet Inspection (DPI):* Right now IT industry increasingly perceives and exploits the esteem usefulness of bundle level investigation moreover called Deep Packet Inspection, for differentiating the genuine source, nature of system, application unwavering quality and execution problems. Deep Packet Inspection (DPI) is mainly used to

audit the substance of bundle and organize the system applications.

*a) Mac Address Identification:* In this process a Media Access Control (MAC) deliversare used to remarkably identify hubs on an Ethernet organization. This strategy uses Media Access Control (MAC) addresses data of a gadget to shape a profile for the observed organize hub.

*b) TCP/UDP Port Number Identification:* Port examining is a standout between the most prevalent methods to discover the benefits that can endeavor to break into frameworks. TCP/UDP port numbers can be used to perceive control framework applications. Each of these conventions comprises a 16-bit source and goal port ID number.

c) *Arrange Payloads Identification:* This method defines organize proprietor to observe activity through the system constantly and to isolate them as per their payloads. The Payload is bits of significant information that is being carried in bundles to the client over the system.

*C. Statistical Based Classification*

Factual characterization mostly refers to the strategies in a view of measurable properties of an activity; in which machine learning is the most famous one. The insights can be usually isolated into bundle level and stream level.

## VI. TRAFFIC ANALYSIS ATTACK

There are different kinds of traffic analysis attack. Some of them are stated below,

*A. Passive Attack:* A passive attack observes insecure traffic such as sensitive information, password, and account number. These used to defend the required data that are used by the user in day to day life. Passive attack offers defense or security from the attacker.

*B. Active Attack:* In the cases of active attack, the attacker attempts to breakdown the security of the system through a virus. Active attacker contains malicious virus, steal and alter the data.

*C. Password Attack:* In these cases, an attacker attempts to discover the password that is stored in a network or host or IP address of the user. By the help of the user data or password the attack attempts to identify the significant data the user store with it.

## VII. COMPARATIVE STUDY OF VARIOUS PROPOSALS

Zhengbing Hu *et al.*,[6] targeted to determine developed anomaly detection system in secure cloud computing environment; to display its theoretical description and conduct appropriate simulation. The result determines that the established system provides a great percentage (>90%) of anomaly detection in secure cloud computing environment. The model can be used to construct data centers in diverse areas. In addition, a model was established for the discovery of anomalies secure environment ―cloud computing based on the concept of Big Data. Also, they were completed the study of modern methods of finding anomalies and taking into account their defects.There was established a hybrid system anomaly detection using the method Decision Tree, signature module Snort, technology Big Data (HDFS, YARN, MapReduce, Spark) and databases KDDCup99 can identify anomalies in traffic secure environment ofcloud computing. This proposed work experimentally examined anomaly finding module in the Weka tool, which showed extremely accurate algorithm. The practical value is the facility to assimilate the developed system anomaly detection in the secure environment of cloud computing and growing the percentage of recognition over the use of signature module that can identify well-known attacks.

Weigang Hou *et al.*,[7] emphasis on a risk-aware VNE framework, because a blind VNE operation would effect in severe information outflow among co resident VMsintheserver.ByassessingVMthreatand vulnerability, risky VMs are identified and rendered to investigation results. To execute physical isolation among risky and security VMs, a risk-aware VNE heuristic algorithm is suggested. The simulation outcomes show that our heuristic algorithm achieves better than the benchmark in terms of sustaining ODCN security and getting rental revenue. In this paper, they have designed a novel RA model to find risky VMs throughout a specific future time epoch. The simulation outcomes have proved that RVNE well assured the ODCN security with a worthy average safety ratio 0.75. More specifically, their RVNE has acquired a greater number of safe servers related to the benchmark. The algorithm solution is very nearby to the upper bound derived from them, which has well established algorithm optimality.

Rajesh Kumar D *et al.*,[8] proposed a light-weight, fast, inexpensive and dispersed agent technology based on security answer beside the black hole attack for Wireless Sensor Networks (WSNs). The projected work is to reserve against Black Hole Attack corruption multiple Base Stations arranged in a network by exploiting mobile agents. The enactment of the proposed approach has been inspected through simulations. The simulation outcomes show that EEMA model offers greater performance with associate enhancement of black hole attack detection probability rate by 17% and conjointly condensed energy intake of the data packet gathering method by 24% when related to state-of-the-art works.

Justin David Pineda *et al.,* [9] discovers dissimilar correlation techniques that recognize patterns based on definite components in the logs. The researchers also present Tree Correlation, a newly-created correlation technique that can be used to help in defining possible

attacks that can occur by examining a series of logs based on header, content and behavior. Among the three types of correlation techniques, the greatest solution to ease the chance of having false positives and refining the operation of the SIEM is the Tree correlation. Tree correlation technique can be used as a way to sort and collect information from the logs to work as a guide to define the accuracy of the alerts. It can handle such attacks like sudden ping sweeps and payload interceptions to identify the received attacks. It can also work as a real-time detection and inspection of the alerts and potential attacks upcoming in the network.

Paolo Costa *et al.*,[10] proposed Network-as-a-Service (NaaS), a structure that incorporates current cloud computing contributions with direct, yet secure, tenant access to the network infrastructure. Using NaaS tenants can simply organize custom routing and multicast protocols. Additionally, by altering the content of packets on path, they can efficiently implement advanced network services such as network data aggregation, elimination, redundancy and smart caching. Their early simulation study proposes that even with inadequate processing capability at network switches, NaaS can significantly raise application output and decrease network traffic.

Theophilus Benson *et al.*,[11] conduct an experimental study of the network traffic in 10 data centers belonging to three dissimilar types of organizations including enterprise, university and cloud data centers. They collect and examine SNMP statistics, topology, and packet-level traces. They observe the range of applications organized in these data centers, their placements, the flow-level and packet level diffusion properties of these applications, their effect on link utilization, network utilization, congestion, and packet drops. They define the effects of the experiential traffic patterns for data center, internal traffic engineering as well as for newly projected architectures for data center networks.Table I Shows the Comparative Study of Various Models Proposed By Various Authors on Network Traffic Analysis.

TABLE I COMPARATIVE STUDY OF VARIOUS PROPOSED WORKS

| Author | Model | Algorithm | Outcome | Result | Environment |
|---|---|---|---|---|---|
| Zhengbing Hu *et al.*,[6] | Data Center Model (DCM) | Decision Tree | Detection of anomalies | 99.65% | Cloud |
| WeigangHou *et al.*,[7] | A Novel RA Model | Risk-Aware VNE Heuristic Algorithm | Find risky VMs and ODCN security | Average Safety Ratio 0.75 | Distributed Data Centers |
| Rajesh Kumar D *et al.*,[8] | EEMA Model | Mobile Agent Based Clone Attack Detection (MABCAD) | Black Hole Attack Detection | Black Hole Attack Detection - 17% Data packet gathering method -24% | Dynamic Network Environment |
| Justin David Pineda *et al.*, [9] | Security Incident Event Management (SIEM) | Tree correlation | Monitoringmalicious and anomalous traffic | reduce the false Reduce the false positives accurately | Security Operation Centers (SOC) |
| Paolo Costa *et al.*, [10] | NaaS Model | INPEs | Better Scalability, Performance Isolation and Programmability | Flow completion time reduced by 63.18% (respectively 93.07% ) | Cloud |
| Theophilus Benson *et al.*,[11] | Centralized Routing Mechanisms | OpenFlow architecture | Moderate link utilizations | 10% delay overhead on most flows | Cloud |

## VIII. CONCLUSION

Cloud computing is the user oriented technology in which user looks a group of virtualized computer resources. In computer networks,devices are computedwith each other using relations (data links) between nodes. Instead referred to as network traffic; traffic is a term used to define all data communications on a computer or computer network. During high traffic periods a computer or computer network may slow down and become crowded if not sufficient for the load. In some cases, too much traffic may avert a computer or network device from operating. Traffic classification is an exact significant calculated and arithmetical tool in communications and computer networking, which is utilized to discover average and arithmetical information of the traffic passing above certain pipe or hub. The results attained from an appropriate deployment of a traffic analysis method to offer valuable insightcomprising: (a) how busy a link is?, (b) the average end-to-end delays, and (c) the average packet size. These valuable information bits will help engineers to show robust networks, evade probable congestions, and foresee future development. In this paper, we have surveyed network traffic and traffic classification in detail.

## REFERENCES

[1] Aug (2017) [Online] Available at: https://en.wikipedia.org/wiki/ Network_traffic

[2] A. Chuvakin, "Security Event Analysis through Correlation", *Information Systems Security*, Vol. 13, No. 2, pp. 13-18, 2004.

[3] J. Pineda and R. Yatco, "Adaptable Software-based Log Consolidation and Incident Management for a Security Information

Event Management System AdLCIM", *Manila: Undergraduate Thesis, De La Salle University-Manila, Philippines*, Dec. 2010.

[4] J. Bernardo and A. Valencia, "Network Threat Detection System,"*Makati, Undergraduate Thesis, De La Salle University-Manila, Philippines*, 2016.

[5] D. Gordon, "Extending Intrusion Detection with Alert Correlation and Intrusion Tolerance", *Goteborg: Thesis for the Degree of Licentiate of Engineering, Chalmers University of Technology, Sweden*, 2003.

[6] Zhengbing Hu Sergiy Gnatyuk, Oksana Koval, Viktor Gnatyuk and Serhii Bondarovets, "Anomaly Detection System in Secure Cloud Computing Environment", *I. J. Computer Network and Information Security*, pp. 10-21, 2017.

[7] Weigang Hou, Zhaolong Ning, Lei Guo, Zhikui Chen, and Mohammad S. Obaidat, "Novel Framework of Risk-Aware Virtual Network Embedding in Optical Data Center Networks ", *IEEE systems journal*, Vol. 12, No. 3, pp. 2473-2482, Sep. 2018.

[8] D. Rajesh Kumar and A. Shanmugam, "Energy Efficient and Trust Based Black Hole Attack Identification Model in Wireless Sensor Networks", *Journal of Network Security Computer Networks*, Vol. 2, No. 3, pp. 1-9, 2017.

[9] Justin David Pineda, Joanna De Guzman and Adrian Tobias,"Classifying Multi-Layered Attacks Using Correlation of Relevant Data in Network Access Technologies", *Proceedings of the 17th Philippine Computing Science Congress*, pp. 73-78, 2017.

[10] Paolo Costa, Matteo Migliavacca, Peter Pietzuch, and Alexander L. Wolf, "NaaS: Network-as-a-Service in the Cloud", In*Hot-ICE'12 Proceedings of the 2nd USENIX conference on Hot Topics in Management of Internet, Cloud, and Enterprise Networks and Services*, 2012.

[11] Theophilus Benson, Aditya Akella and A. David Maltz, "Network Traffic Characteristics of Data Centers in the Wild", In IMC '10 *Proceedings of the 10th ACM SIGCOMM conference on Internet measurement*, p. 267-280, 2010.