ISSN: 2319-8753

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 11, November 2014

Enhancing the Traffic Privacy with Leakage Detection and Optimization

Sarathi.S[1], Balasubramaniam.R[2]

P.G Student, Department of Computer Science Engineering, Kathir College of Engineering, Coimbatore, T.N, India[2]

Assistant Professor, Department of Computer Science Engineering, Kathir College of Engineering, Coimbatore, T.N, India[2]

ABSTRACT: With recent experimental works, it is found that there is a information leakage in the packetized flow of data due to user activity and traffic content. We aim at understanding how complex when the information leaked by packet traffic features namely packet length, direction and times. Technique to call this type of feature is called traffic masking. Here, we define a security model to find the ideal target of masking which removes any leakage. Further, we investigate there is a tradeoff between traffic privacy protection and making cost, namely required amount of overhead and realization complexity feasibility. Major findings are that 1). Masking is a security model to find the ideal target of masking and it removes the leakage of packets in the network. Masking achieves similar overhead values with padding only and in case fragmentation is allowed and 2) Optimized statistical masking attains only moderately better overhead than simple fixed pattern masking does, while still leaking correlation information that can be exploited.

KEY WORDS: Privacy, Traffic Masking, Traffic flow classification, Padding, Fragmentation

I. INTRODUCTION

From the last few years an extensive experimental evidence within a secure channel, a packetized flow leaks information through observation of features of traffic flows, for example, the ordered sequence of packet lengths, packet inter arrival times and packet directions. Based on the above information when the flow is carried within a secure channel (e.g., SSL/TLS or SSH connections), to identify the type of service or application protocol run among a given set of alternatives to classify the traffic. Other privacy breaking attacks based on analysis of packet flow features have been demonstrated, for example, to profile web access, to infer language of phone calls or even conversation transcripts .This communication privacy break is a positive proof that ciphering does not conceal all relevant information of a packetized application flow; hence, we aim at investigating protection of privacy against traffic analysis. Besides being a privacy issue, traffic analysis tools can be useful to network administrators and operators for enforcement of security policies and traffic filtering, or to support quality of service mechanisms. We term this as traffic masking.

II. RELATED WORK

During the years there were various countermeasures have been developed ie, Wright et al. It makes use of the convex optimization techniques to modify the source packet lengths distribution to look like a target distribution with the minimum overhead values. It does not provide any explicit solution.

Next technique is Yu et al. It implements a new strategy of packet padding aiming at offering perfect anonymity on web browsing. The proposed solution allows in reducing the overhead by exploiting as padding web pages that the user is expected to download in the future.

Luo et al,to modify statistical features of a traffic flow by using a set of transformations both at the application and

DOI: 10.15680/IJIRSET.2014.0311027

ISSN: 2319-8753

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 11, November 2014

at the transport layer. In a recent work, Dyer et al, shows that it is still possible to classify traffic flows after masking. They consider nine masking countermeasures applied to web pages, and show that they can still identify which webpage is being downloaded with standard algorithms, achieving accuracy as high as 98 percent.

As for contributions of this work, we define formally the problem of privacy against traffic classification, thus finding what the ideal traffic masking should do. Then, we define the achievable performance bounds for a masking algorithm, by defining an optimization problem to find an ideal masking algorithm that minimizes overhead cost.

III. GENERAL MODEL OF AN APPLICATION FLOW

Any application traffic flow between an initiator entity A and the responder entity B (e.g., client and server for the given flow, respectively) can be cast into a sequence of N _ 1 message bursts.[2] Each burst consists of one or more messages in one direction (A ! B or B ! A). Bursts in the two opposite directions alternate, starting from the initial burst sent by the initiator A to the responder B.

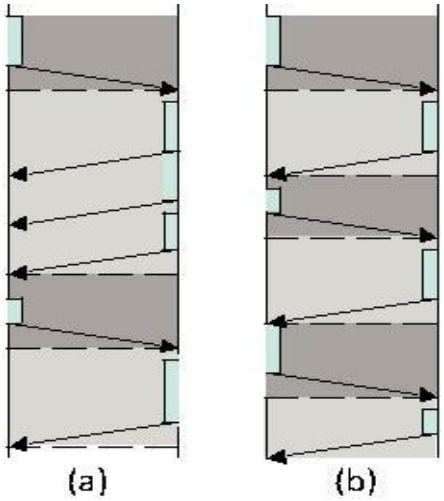

Fig 1.Examples of message exchanges of two application flows

(a) One way data transfer, like http; (b) Alternate messages, like most signaling and control protocols.

IV. DATA LEAKAGE

Data leakage is the accidental or unintentional distribution of private data to an unauthorized entity. Sensitive data in organizations and companies includes patient information, credit-card data and other information based on the business and industry. It possesses a serious issue for companies to the number of incidents and the cost to get increase. By the fact, it is enhanced that the transmitted data (both inbound and outbound), including emails, instant messaging, website forms and file transfer are largely unregulated and unmonitored by the way to their destinations.

Furthermore, in many cases, sensitive data are shared among various stakeholders such as employees working from outside the organization's premises (e.g., on laptops), business partners, and customers. This increases the risk that confidential information will fall into unauthorized hands.

ISSN: 2319-8753

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 11, November 2014

V. MODEL OF TRAFFIC FLOW MASKING

Let us consider two end points (hosts) in a packet network running in a given application. The information flow between the two hosts is made to be ciphered. Though there is a secure channel for communication still it shows the information leakage between the two hosts.

VI. DATA MASKING AND THE CLOUD

In the latest years, organizations develop their new applications in the cloud more and more often, whether final applications will be hosted in the cloud. The Cloud Solutions allow organizations to use Infrastructure as a Service or Iaas, Platform as a Service or Paas, and Software as a Service or Saas.

There are various modes of creating test data and moving it from on-premises databases to the cloud, or between the different environments within the cloud. Data masking invariably becomes the part of these processes in SDLC.

VII.METHODOLOGY

It describes a methodology for identifying the sources of the discriminative power in traffic classification based on the performance metrics and machine learning algorithms.

VIII. PERFORMANCE METRICS

To evaluate the traffic classification performance based on the machine learning algorithms, it can be classified as five metrics: overall accuracy, precision, recall, F-measure, and classification speed:

Overall accuracy: It shows that the ratio of the number of correctly classified traffic flows to the total number of all flows in a particular given trace. This metric is to measure the accuracy of a classifier on the whole trace set.

Precision: Precision shows that the ratio of True Positives over the sum of True Positives and the False Positives or the percentage of flows that are properly attributed to a given application.

Recall: It shows that the ratio of True Positives over the sum of True Positives and False Negatives or the percentage of flows in an application class that are correctly identified.

F-measure: It is widely-used as a metric in information retrieval and classification, it considers both precision and recall in a single metric by taking their harmonic **mean (2 × precision × recall precision + recall)**.We use this metric to measure the per-application classification performance of machine learning algorithms.

Classification speed: It shows that the number of classification decisions performed per second in a given application.

Technique Used:

Traffic Masking:

Analyze the leakage information and Packet Features like packet length, packet direction and times.Alert has been sent to the admin , when an unauthorized user leaks data. Users can also find the Hacker details when a unknown person trying to access their details with own user name and password.

DOI: 10.15680/IJIRSET.2014.0311027

ISSN: 2319-8753

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 11, November 2014

Guilt Model Technique:

Occurs due to more traffic and less bandwidth in the network.Detecting the leaked data and displaying the leaked person details with their system information. Leakage removal has been performed. Providing the probability chart to identify the leaker within then given application.

IX. PRIVACY AGAINST FLOW CLASSIFICATION

In general, the traffic masking operation includes introducing dummy flows, to modify the a priority probabilities as P_j into new values as Q_j, and transforming each flow sent through the network and the flow transformation between the different hosts implies message padding, fragmenting, insertion of dummy messages, and message delaying.

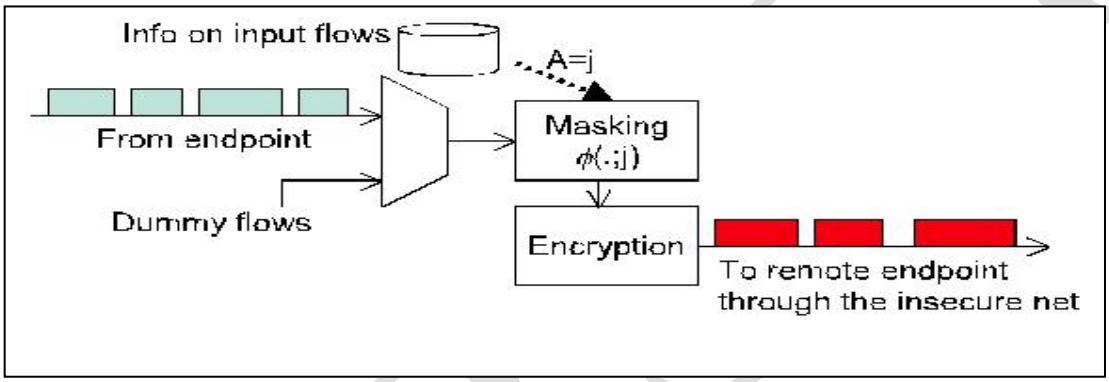

Fig2. Dummy flows are added to modify the priority of generating applications.

From the input for the flow transformation can be considered as a A=j and from the end point this flow can be masked as ω(j).We assume that when an eavesdropping adversary, aiming at flow classification. The adversary can be observing ciphered and masked flows (including dummy flows) .It can detect the feature vector y for each observed flow.

In other words, the adversary can collect samples y of the random variables Y. An overall scheme of the masking plus enciphering at sending side is shown in Fig. 2. The reverse operations (deciphering and de masking) take place at receiver side. Probability of correctly classified ob-served flows to get its theoretical minimum 1=M. Given the adversary guilt model above, we mean removing any information leakage that could be exploited by the adversary to classify observed flows.

X. CONCLUSION

Role of the project is to define a security model for the data leakage based on the traffic and user activity in a particular organization. The information leakage in the packetized flow of data due to user activity and traffic is analyzed and detected by using the above mentioned techniques. The admin can detect the traffic in the network by using the network bandwidth.

Maintaining a probability chart for the agents to find the highest probability of the agents and to show the leaker name. Performing the security for the separate agents when some unknown person trying to access system with their own username and password. The system information and login details including time will be displayed in their agent's hacker list.

DOI: 10.15680/IJIRSET.2014.0311027

ISSN: 2319-8753

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 11, November 2014

REFERENCES

[1] Yeon-supLimMA,Hyun-chul,JeongSeoul,Chong-kwon,"Taekyoung",YangheeChoiSeoul ,"Internet Traffic Classification Demystified": On the Sources of the Discriminative Power.

[2] "Data Leakage Detection"Sandip A. Kale1, Prof. S.V.Kulkarni2 Department Of CSE, MIT College of Engg, Aurangabad, Dr.B.A.M.University, Aurangabad (M.S), India 1,2."International Journal of Advanced Research in Computer and CommunicationEngineering ", Vol. 1, Issue 9, November 2012 .

[3] Sridhar Gade1, Kiran Kumar Munde2, Krishnaiah.R.V." Data Allocation Strategies for Leakage Detection", IOSR Journal of Computer Engineering (IOSRJCE) ISSN: 2278-0661, ISBN: 2278-8727 Volume 5, Issue 2 (Sep-Oct. 2012),"

[4] "Michael BackesSaarland University and MPI-SWS ,Goran Doychev IMDEA Software Institute ,Boris Kopf IMDEA Software Institute boris.koepf@imdea.org," Preventing Side-Channel Leaks in Web Trac: A Formal Approach".

[5] Alice Este, Francesco Gringoli, Luca Salgarelli DEA, " On the Stability of the Information Carried by Traffic Flow Features at the Packet Level" ,Universita degli Studi di Brescia, Italy.

[6] 1S.Jenila, 2K.Sivasankari,3R.Arudselvi,4J.Maria Monica,5B.Saranya." Guilt Model Process for Identifying Data Leakage and Guilty Agent in Data transmission"

[7] Alfonso Iacovazzi, Andrea Baiocchi " Ideal Packet Length Masking against Traffic Classification ".