ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 2.114

IJESRT
INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

Proficient Network Traffic Categorization Method with Strange Flow Detection

P.Thirugnanam*, M.Kalaimani
Assistant Professor, Department of Computer Science and Engineering, IFET College of Engineering, Tamilnadu, India

Abstracts

The area of Internet traffic measurement has advanced enormously in the last couple of years. This advancement is mostly based on the enormous growth in the number of users connected, in the increase in user access speeds and in the appearance of network-hungry applications. This change greatly affected the work of Internet Service providers and network administrators, which have to deal with increasing network users and capacity demands and abrupt traffic changes caused by new applications. This paper has been developed keeping in view of the distributed client server computing technology, in mind.

This paper provides a secure categorization through user interfaces which are browser specific to give distributed accessibility for the overall system. This paper provides security in the system by using the IP Trace back technique which traces the IP of the attacker and locates the position of the system. In addition to that, it also provides safe browsing features which are user specific and safe mailing feature with an ability to perform spam filtering

Keywords: Traffic flaws, IP trace back, spam filtering.

Introduction

TRAFFIC classification technique plays an important role in modern network security and management architecture. For instance, traffic classification is normally an essential component in the products for QoS control and intrusion detection. With the popularity of cloud computing, the amount of applications deployed on the Internet is quickly increasing and many applications adopt the encryption techniques. This situation makes it harder to classify traffic flows according to their generation applications. Traditional traffic classification techniques rely on checking the specific port numbers used by different applications or inspecting the applications' signature strings in the payload of IP packets. These techniques encounter a number of problems in the modern network such as dynamic port numbers, data encryption and user privacy protection. Currently, the state-of-the-art methods tend to conduct classification by analyzing flow level statistical properties. Substantial attention has been paid on the application of machine learning techniques to flow statistical features based traffic classification. However, the performance of the existing flow statistical feature based traffic classification is still unsatisfied in real world environments. In supervised traffic classification the flow classification model is learned from the labelled training samples of each predefined traffic class. The supervised methods classify any flows into predefined traffic classes, so they cannot deal with unknown flows generated by unknown applications. Moreover, to achieve high classification accuracy, the supervised methods need sufficient labelled training data. By contrast, the clustering-based methods can automatically group a set of unlabelled training samples and apply the clustering results to construct a traffic classifier. In these methods, however, the numbers of clusters have to be set large enough to obtain high-purity traffic clusters. It is a difficult problem of mapping from a large number of traffic clusters to a small number of real applications without supervised algorithm.

The existing traffic classification methods suffer from poor performance in the crucial situation where supervised information is insufficient and considerable unknown flows are present. Particularly, more and more new/unknown applications are emerging in the cloud computing based environment. Robust traffic classification is a big challenge in the real-world complex network. For instance, as the amount of new applications quickly increases, it can only collect and analyze an uncompleted training data set. Moreover, if the emerging applications are encrypted, it is almost impossible to analyze sufficient training samples through deep inspection in a limited time. These observations become the motivation of our work. This paper aims to tackle the problem of unknown flows in a semi-supervised framework. This work considers very few labelled training samples and investigates flow correlation in real world network environment, which makes it different to previous works. The major contributions of this paper are as follows.

This paper develops a system model to incorporate flow correlation into a semi-supervised method, which possesses the capability of unknown flow detection. And flow label propagation to automatically label relevant flows from a large unlabelled dataset in order to address the problem of small supervised training set.

This paper proposes the compound classification to jointly identify the correlated flows in order to further boost the classification accuracy. It provides the theoretical justification on performance benefit of applying these two new techniques to network traffic classification.

Implementation

IP Traceback System

The goal of this module is to trace the path of an IP packet to its origin. The most important usage of IP traceback is to deal with certain Denial-Of-Service (DoS) attacks, where the source IP address is spoofed by attackers. Identifying the sources of attack packets is a significant step in making attackers accountable. In addition, figuring out the network path which the attack traffic follows can improve the efficacy of defense measures such as packet filtering as they can be applied further from the victim and closer to the source. Each router maintains a different digest table for each of its neighbor routers. Packets coming from different neighbor routers (with different marks) can be recorded in corresponding digest tables simultaneously. That reduces the access time requirement by a factor of the number of neighbor routers.

E-Mail System

In this, it develops Email System to communicate with in organization. If any attack occurs mail will be send with the attacker's detail.
Injection Flaws – Injection flaws occur when software does not properly validate input. An attacker could craft malicious input that causes the Web Service software to perform operations on behalf of the attacker. Classes of injection flaws include Cross Site Scripting, SQL Injection, and XPath Injection.
Insecure Communications – Attackers can steal or modify information if not protected while in transit.
Insufficient Authentication - Web Services that perform sensitive functions should require authentication.

Distributed Caching

The proxies cooperate in the following way:
On a cache miss, a proxy attempts to determine if another proxy cache holds the desired Webpage. If so, a request is made to that proxy rather than trying to obtain that page from the Internet.

For such a scheme to be effective, proxies must know the contents of other proxy caches. In Summary Cache, to reduce message trace proxies do not transfer URL lists corresponding to the exact contents of their caches, but instead periodically broadcast Bloom filters that represent the contents of their cache. If a proxy wishes to determine if another proxy has a page in its cache, it checks the appropriate Bloom filter. In the case of a false positive, a proxy may request a page from another proxy, only to end that proxy does not actually have that page cached. In that case, some additional delay has been incurred. In this setting, false positives and false negatives may occur even without a Bloom filter, since the cache contents may change between periodic updates. The small additional chance of a false positive introduced by using a Bloom filter is greatly outweighed by the significant reduction in network trace achieved by using the succinct Bloom filter instead of sending the full list of cache contents. This technique is used in the open source Web proxy cache Squid, where the Bloom filters are referred to as Cache Digests.

Minimize false negative items

It shows that a false positive can trigger a deletion of a false positive item and result in at least one multi address item. Both cases cause an incorrect item deletion operation and lead to potential false negative items.

This reveals that the resulting false negative items are usually not fully exposed in consequent queries. It also measure the potential and exposed false negative items caused by an incorrect item deletion operation.

This paper proposes two principles to make potential false negatives unexposed whenever possible. Our design is able to increase the ratio of bits set to a value larger than one in a BF without decreasing the ratio of bits set to zero. And an enhanced BF scheme, which can reduce about 50-80 percent of exposed false negative items in BF. Through extensive experiments and mathematical analysis, the analysis shows the design achieves the desired properties.

Peer to Peer File Sharing System

Peer to Peer (P2P) file sharing an extremely popular method for swapping large files on the Internet, particularly music and videos. Unlike FTP, most P2P file sharing systems do not use any central servers but instead allow all computers on the network to function both as a client and a server. Numerous free P2P software programs exist each with their own technical advantages and loyal community following. Instant Messaging (IM) systems are a type of P2P application most commonly used for chatting, but all popular IM software also supports sharing files.

Online Sharing Services

Finally, numerous Web sites built for community file sharing exist on the Internet. Members post or upload their files to the site using a Web browser, and others can then download copies of these files using their browser. If attack occurs Mail Has Been sent with Attacker's Details.

Conclusion

In the faster development of the number of applications entering into a network, traffic classification faces more critical threats in the current advanced network and systems. This paper addressed the problem of unknown applications in the critical environment of small supervised training data. The system uses flow label propagation technique to automatically label unlabelled flows accurately, thereby increasing the capability of the nearest cluster based classifier and the other technique performs combining a number of flow predictions to make accurate classification of BoFs. In addition to the classification, the system attempts to introduce the same concept into the real-time environment where security issues are taken into consideration. The proposed system identifies the attacker in case of any attacks occurring in the system and also, it provides a secure browsing and mailing system that further enhances the security.

References

1. Jun Zhang, Chao Chen, Yang Xiag, Wanlei Zhou and Athanasios, "An effective network Traffic classification method with unknown flow detection" , IEEE Transaction, vol. 10, no. 2, JUNE 2013.
2. Jun Zhang, Yang Xiang, Yu Wang, Wanlei Zhou, Yong Xiang and Yong Guan, "Network traffic classification using correlation information", IEEE Trans. ParrallelDistrib. Syst., vol.24, no.1,pp.104-117,JAN 2013,
3. Yu Wang, Yang Xiang, Jun Zhang and Shunzheng Yu, "A novel semi-supervised approach for network traffic clustering", 2011 International conference on network and system security.
4. D. Bonfiglio, M. Mellia, M. Meo, D. Rossi, and P. Tofanelli, "Revealing Skype traffic: when randomness plays with you," in Proc. 2007 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications, pp. 37–48.
5. Andrew W. Moore and Denis Zuev, "Internet traffic classification using Bayesian Analysis Techniques", SIGMETRICS Perform. Eval. Rev., vol. 33, pp. 50-60, June 2005.
6. Y. Lim, H. Kim, J. Jeong, C. Kim, T. Kwon, and Y.Choi. "Internet Traffic Classification Demystified: On the Sources of the Discriminative Power". In ACM CoNEXT, December 2010.