IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661, p- ISSN: 2278-8727Volume 16, Issue 2, Ver. III (Mar-Apr. 2014), PP 100-106
www.iosrjournals.org

Detecting a Proficient Flow Label Propagation from Network Traffic Classification

S. K. Murugaraja[1], R. Rameshkumar[2], K.Sarathkumar[3]

[1]*head Of The Department / Computer Science And Engineering Department, Anna University Chennai, Gnanamani College Of Technology, Rasipuram, Namakkal -637214, India.*

[2]*assistant Professor / Computer Science And Engineering Department, Anna University Chennai, Gnanamani College Of Technology, Rasipuram, Namakkal -637214, India.*

[3] *M.E Computer Science And Engineering Department, Anna University Chennai, Gnanamani College Of Technology, Rasipuram, Namakkal -637214, India.*

Abstract: *The system model in a traffic classification method introduces the threefunctions that are flow label propagation, nearest cluster based classifier and compound classification. For some special cases, the flow label propagation is critical to require that the labeled flows and unlabelled flows must be captured on the same network in a short period of time. In flow label propagation the peer whose initial flow we have labeled may be off-line, so no other peer will connect to it and flow label propagation is useless for this case. Same for servers, say, and FTP server is popular on a network, and then you will correctly propagate flow information to all other hosts connecting to that server. However, if the server changes its IP address (without changing its DNS name), then will fail to propagate its label. For these type of case propose a alter flow label propagation and also detect the unknown flow in the network*

Keywords: *FTP (file transfer protocol) , IP (Internet protocol*

I. Introduction

Traffic classification technique plays an important role in modern network security and management architectures. For instance, traffic classification is normally an essential component in the products for Quos control and intrusion detection. With the popularity of cloudcomputing, 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 unlabelled training samples and apply the clustering results to construct a traffic classifier. In these methods, however, the number of clusters has 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 information.

For some special cases, the flow label propagation is critical to require that the labeled flows and unlabelled flows must be captured on the same network in a short period of time. A detailed reasoning is as follows. Flow label propagation involves looking at 3-tuple including IPs and ports. Hence, during, say a 1-hour period we may catch many Bit Torrent peers swarming together, and if we label one flow, due to the mushiness nature of the Bit Torrent overlay we can succeed in catching them all. However, if we move the next day, the peer whose initial flow we have labeled may be off-line, so no other peer will connect to it and label propagation is useless for this case. Same for servers, say, and FTP server is popular on a network, then you will correctly propagate flow information to all other hosts connecting to that server. However, if the server changes its IP address (without changing its DNS name), then

will fail to propagate its label. In addition, the methodology cannot work across networks: i.e., a labeled flow in isp cannot be useful in wide trace unless the server is popular in both setups. It should be pointed out that this paper does not address the problem of traffic classification across networks.

Develop a system model to incorporate flow correlation into a semi-supervised method, which possesses the capability of unknown flow detection. Propose flow label propagation to automatically label relevant flows from a large unlabelled dataset in order to address the problem of small supervised training set. The compound classification to jointly identify the correlated flows in order to further boost the classification accuracy. If the flow label may be a off-line, then a new alternate flow label have been detect to reduce the traffic in a network.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.

A number of supervised classification algorithms and unsupervised clustering algorithms have been applied to network traffic classification. In supervised traffic classification, the flow classification model is learned from the labeled 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 labeled 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 number of clusters has 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 information.

For some special cases, the flow label propagation is critical to require that the labeled flows and unlabelled flows must be captured on the same network in a short period of time. A detailed reasoning is as follows. Flow label propagation involves looking at 3-tuple including IPs and ports. Hence, during, say a 1-hour period we may catch many Bit Torrent peers swarming together, and if we label one flow, due to the mushiness nature of the Bit Torrent overlay we can succeed in catching them all. However, if we move the next day, the peer whose initial flow we have labeled may be off-line, so no other peer will connect to it and label propagation is useless for this case. Same for servers, say, and FTP server is popular on a network, then you will correctly propagate flow information to all other hosts connecting to that server. However, if the server changes its IP address (without changing its DNS name), thenwill fail to propagate its label. In addition, the methodology cannot work across networks: i.e., a labeled flow in isp cannot be useful in wide trace unless the server is popular in both setups. It should be pointed out that this paper does not address the problem of traffic classification across networks.

Develop a system model to incorporate flow correlation into a semi-supervised method, which possesses the capability of unknown flow detection. Propose flow label propagation to automatically label relevant flows from a large unlabelled dataset in order to address the problem of small supervised training set. The compound classification to jointly identify the correlated flows in order to further boost the classification accuracy. If the flow label may be a off-line, then a new alternate flow label have been detect to reduce the traffic in a network.

1.1domaindescription

Network management refers to the broad subject of managing computer networks. There exists a wide variety of software and hardware products that help network system administrators manage a network. Network management covers a wide area, including

Security: Ensuring that the network isprotected from unauthorized users.
Performance: Eliminatingbottlenecks in the network.
Reliability: Making sure thenetwork is available to users and responding to hardware and software malfunctions

1.2 Network Management

OSI management is required for a number of purposes. These requirements are categorized into a number of functional areas:

- Fault management
- Accounting management
- Configuration management
- Performance management
- Security management

1.3 Specific Management Functions

Specific management functions within these functional areas are provided by OSI management mechanisms. Many of the mechanisms are general in the sense that they are used to fulfill requirements in more than one functional area. Similarly, managed objects are general in the sense that they may be common to more than one functional area. Each of these functional areas is described briefly below. The lists of functions are not necessarily exhaustive. **Specific Management Functions**

Specific management functions within these functional areas are provided by OSI management mechanisms. Many of the mechanisms are general in the sense that they are used to fulfill requirements in more than one functional area. Similarly, managed objects are general in the sense that they may be common to more than one functional area. Each of these functional areas is described briefly below. The lists of functions are not necessarily exhaustive.

Fault Management

Fault management encompasses fault detection, isolation and the correction of abnormal operation of the OSI environment. Faults cause open systems to fail to meet their operational objectives and they may be persistent or transient. Faults manifest themselves as particular events (e.g. errors) in the operation of an open system. Error detection provides a capability to recognize faults. Fault management includes functions to

- Maintain and examine error logs
- Accept and act upon error detection notification
- Trace and identify faults
- Carry out sequences of diagnostic tests
- Correct faults.

Accounting Management

Accounting management enables charges to be established for the use of resources in the OSIE, and for costs to be identified for the use of those resources. Accounting management includes functions to

- Inform users of costs incurred or resources consumed
- Enable accounting limits to be set and tariff schedules to be associated with the use of resources
- Enable costs to be combined where multiple resources are invoked to achieve a given communication objective.

Configuration Management

Configuration management identifies, exercises control over, collects data from and provides data to open systems for the purpose of preparing for, initializing, starting, providing for the continuous operation of, and terminating interconnection services. Configuration management includes functions to

- Set the parameters that control the routine operation of the open system;
- Associate names with managed objects and sets of managed objects;
- Initialize and close down managed objects;
- Collect information on demand about the current condition of the open system;
- Obtain announcements of significant changes in the condition of the open system;
- Change the configuration of the open system.

Performance Management

Performance management enables the behavior of resources in the OSIE and the effectiveness of communication activities to be evaluated. Performance management includes functions to

- Gather statistical information;
- Maintain and examine logs of system state histories;
- Determine system performance under natural and artificial conditions; and
- Alter system modes of operation for the purpose of conducting performance management activities.

Security Management

The purpose of security management is to support the application of security policies by means of functions which include

- The creation, deletion and control of security services and mechanisms;
- The distribution of security-relevant information; and
- The reporting of security-relevant events.

Note - Recommendation X.800 provides further information on the placement of OSI management functions within the overall security architecture.

Service Management

The critical and complex nature of today's business applications has made it very important for IT organizations to monitor and manage application service levels at high standards of availability. Problems faced in an enterprise include service failures and performance degradation. Since these services form an important type of business delivery, monitoring these services and quickly correcting problems before they can impact business operations is crucial in any enterprise.

Service-level agreements are used to evaluate service availability, performance, and usage. By constantly monitoring the service levels, IT organizations can identify problems and their potential impact, diagnose root causes of service failure, and fix these in compliance with the service-level agreements.

Enterprise Manager provides a comprehensive monitoring solution that helps you to effectively manage services from the overview level to the individual component level. When a service fails or performs poorly, Grid Control provides diagnostics tools that help to resolve problems quickly and efficiently, significantly reducing administrative costs spent on problem identification and resolution. Finally, customized reports offer a valuable mechanism to analyze the behavior of the applications over time.

Grid Control monitors not only individual components in the IT infrastructure, but also the applications hosted by those components, allowing you to model and monitor business functions using a top-down approach, or from an end-user perspective. If modeled correctly, services can provide an accurate measure of the availability, performance, and usage of the function or application they are modeling.

1.4 Defining Services In Enterprise Manager

A "service" is defined as an entity that provides a useful function to its users. Some examples of services include CRM applications, online banking, and e-mail services. Some simpler forms of services are business functions that are supported by protocols such as DNS, LDAP, POP, or SMTP.

Grid Control allows you to define one or more services that represent the business functions or applications that run in your enterprise. You can define these services by creating one or more service tests that simulate common end-user functionality. Using these service tests, you can measure the performance and availability of critical business functions, receive alerts when there is a problem, identify common issues, and diagnose causes of failures.

You can define the following service types: Generic Service, Web Application, Forms Application, and Aggregate Service. Web applications, a special type of service, are used to monitor Web transactions. Forms applications are used to model and monitor Forms transactions.

The following elements are important to understanding Grid Control's Service Level Management feature:

- **Service:** Models a business process or application.
- **Availability:** A condition that determines whether the service is considered accessible by the users or not.
- **Service Test:** The functional test defined by the Enterprise Manager administrator against the service to determine whether or not the service is available and performing.
- **System:** A group of underlying components, such as hosts, databases, and application servers, on which the service runs. For more information on systems, refer to the "Managing Systems" section in this chapter.
- **Beacons:** Functionality built into Management Agents used to pre-record transactions or service tests.
- **Performance and Usage:** Performance indicates the response time as experienced by the end users. Usage refers to the user demand or load on the system.
- **Service Level:** Operational or contractual objective for service availability and performance.
- **Root Cause Analysis:** Diagnostic tool to help determine the possible cause of service failure

Service management is the handling of the service components of a service system through the service system life-cycle activities. Services are deeds, processes, and performances. A service is a time-perishable, intangible experience performed for a customer acting in the role of a co-producer.

1.5 Definition Of Service Firms

Service enterprises are organizations that facilitate the production and distribution of goods, support other firms in meeting their goals, and add value to our personal lives.

Networked Service System Consist of service components. A service component is a software component executing on a node.

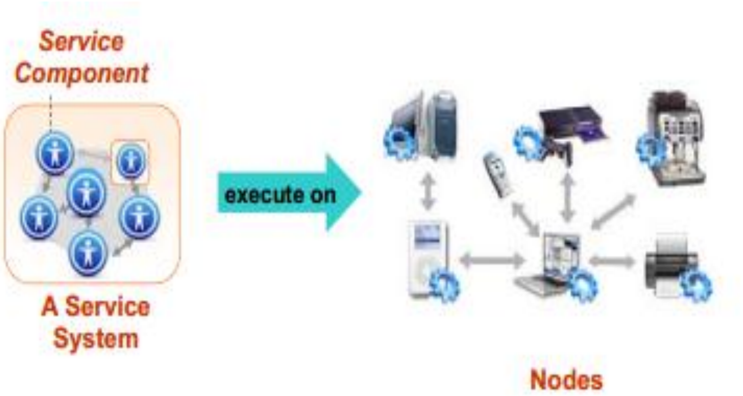


Fig:1.1 Service Components

Capability: An inherent property of a node, which defines the ability to do something. A capability is a feature available to implement services.

- **Resources** (physical or logical components with finite capacity, e.g. processing resources, storage resources, communication resources, addresses, etc.)
- **Functions** (pure software or combined software/hardware components, which perform particular tasks, e.g. encryption functions or special programs available for general use),
- **Data** (just data, the interpretation, validity and lifespan of which depend on the context of the usage, e.g. user login and access rights)

This five-part series of articles will outline the strategic importance of service assurance to CSPs and highlight a clear path towards achieving the right level of service assurance practice appropriate to the organization.

Each of the articles will focus on one of the four clear steps that we have identified that transform service assurance from an operational tool to a strategic asset. These four stages are:

❖ Primary
❖ Operational
❖ Tactical
❖ Strategic

Completing each stage delivers benefits to CSPs, ensuring consistent ROI from the beginning of an implementation of the process.

Achieving service management excellence is a journey that never ends. In a dynamic and fluid network environment, it is a discipline that is essential to the smooth operation of networks, services and operational systems. When implemented, a fully-fledged service management and assurance solution makes a significant contribution to the top and bottom line of any CSP's business.

Service assurance is an on-going process. Just as an organization can never have enough sales, so they can never stop paying attention to service assurance. But with service management and assurance having such a critical role for CSPs, how can they both achieve optimal service assurance delivery and implement supporting processes to ensure that best practice continues to be observed. This series of articles will help you solve that issue.

Service management and assurance requires the unification of different systems into a seamless network, ensuring consistency and that CSPs fully benefit from their deployed assets, leveraging the information they generate to ensure full optimization and quality.

The adoption of eTOM/ITIL standards to provide a framework to achieve this is now internationally recognized as best practice for CSPs. NetVision from ISPM is a compliant eTOM service assurance solution that enables the CSP to move from network maintenance to service excellence.

By leveraging ISPM's expertise and deploying the appropriate elements of the NetVision solution, CSPs can cost-effectively embrace the strategic service assurance approach and benefit at every step in the path.

1.6 Network Management Service Features

Performance management 24/7/365 by Century Link skilled operations engineers, including continuous, proactive network monitoring and real-time notification of managed devices

Easy-to-use Control Center interface to access network performance statistics, view alerts, and create and manage trouble tickets

Fault management and escalation through analysis of activities to isolate and correct unusual operational behaviors

Configuration management for configuration faults and customer-requested changes for configuration integrity

Full suite of online reporting capabilities available through Control Center Comprehensive device management service level agreements (SLAs) including performance thresholds for incident identification and response, policy change request acknowledgement, policy change request implementation, and emergency change request implementation.

II. System Design

Flows: These are represented by 5 tuples, {source_ip, destination_ip, source port, destination port, protocol} and are classified as unidirectional, bidirectional and full flows. Flow statistical Features: Features are properties or characteristics of flows calculated over multiple packets. Some of the properties are maximum or minimum packet length in each directions, minimum or maximum packet arrival time, minimum or maximum number of bytes transferred in forward and backward directions.

Performance metrics: Terms used are True Positive (TP), True Negative (TN), and False Positive (FP), and False Negative (FN), accuracy (flow and byte), precision, recall, classification error, F-measure [10].Common metrics to find classifier‟s accuracy are as follows [2]:

 True Positive: The number of features belonging to class Y classified as class Y.

 False Positive: The number of features of other classes incorrectly classified as belonging to class Y.

- True Negative: The number of features of other classes correctly classified as not belonging to class Y.
- False Negative: The number of features of other class Y incorrectly classified as not belonging to class Y.
- Accuracy: It is defined as the fraction of the number of correctly classified flows or bytes over the size of the data.
- Classification Error: It is defined as the raw count of flows which were correctly classified divided by the total number of flows. This metric is used to find classifier's accuracy for the whole system.
- F-measure: *2 * Precision * recall /(precision +recall).*This metric is used to rank and compare the per-application performance of ML algorithms.

ML uses the following metrics:

- Recall: The numbers of features of class Y correctly classified as belonging to class Y.
- Precision: The numbers of those instances that truly have class Y, among all those classified as class Y.

Confusion matrix is used to show the relationship between performance metric terms.

III. Implementation

Multihop wireless networks (MWNs), or the next-generation wireless networks, can significantly improve network performance and deployment and help implement many novel applications and services. However, when compared to wired and single-hop wireless networks, MWNs are highly vulnerable to serious security threats because packets may be relayed through integrated networks and autonomous devices. My research has been focusing on developing security protocols for securing MWNs. Specifically, we are interested in securing route establishment and data transmission processes, establishing stable routes, and preserving users‟ anonymity and location privacy.

3.1 File Selection

A user can select a file from local machine. When the form is submitted (perhaps together with other form data), the file is uploaded to the web server. The class can process multiple files selected with the file or text form. The descriptions are picked from the value of a form text field that is submitted with the file field data.

3.2 Ip Address Scan

IP address scanning is a procedure for identifying active node on a network. Scanning procedures, such as IP address and port scans, return information about which IP addresses map to live node that are active on the network. A port scan is a series of messages sent by someone attempting to break into a computer to learn which computer network services, each associated with a "well-known" port number, the computer provides.

3.3 Flow Label Propagation

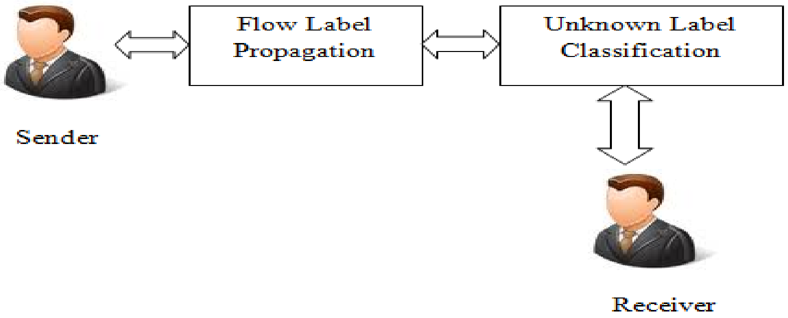
The proposed method aims to classify traffic flows based on the flow level statistical properties. A flow consists of successive IP packets having the same 5-tuple:{ source ip, source port, destination ip, destination port, transport protocol }. Traffic flows are constructed by inspecting the headers of IP packets captured by the system on a computer network. For the purpose of classification, each flow can be represented using a set of flow level statistical properties such as number of packets and packet size.

3.4 File Transfer

The file transfers describe a files and data conversion for the data transfer function. Moving or copying a file from one computer to another (whether of the same or different type) over a network.After scanning the IP address, we select the IP to transfer a data or file. The files are transferred to selected IP address.

3.5 Unknown Traffic Detection

To carry out a number of experiments to evaluate the capability of unknown traffic detection which is proposed to tackle the problem of unknown applications in the proposed method and Erman's method. Two new measures are introduced to study the accuracy of unknown traffic detection quantitatively. False detection rate is defined as the ratio of the sum of known flows inaccurately detected as unknown to the sum of known flows. True detection rate is defined as the ratio of the sum of flows accurately detected as unknown to the sum of all testing unknown flows



IV. Conclusion

Traffic classification encounters more critical problems in current advanced network and system, especially in cloud computing environment. A alter flow label propagation and also detect the unknown flow in the network. In flow label propagation the peer whose initial flow labeled may be off-line, so no other peer will connect to it and flow label propagation is useless for this case. Same for servers, say, and FTP server is popular on a network, then you will correctly propagate flow information to all other hosts connecting to that server. However, if the server changes its IP address (without changing its DNS name), then will fail to propagate its label. By using the proposed method we would not fail to propagate the label when the peer is in offline.

References

[1] Mahmoud And Shen: A Secure Payment Scheme With Low Communication And Processing, Overhead For Multihop wireless networks., IEEE Transactions On Parallel And Distributed Systems, Vol. 24,No. 2, pp. 209-224, February 2013.

[2] G. Shen, J. Liu, D. Wang, J. Wang, and S. Jin, "Multi-Hop Relay for Next-Generation Wireless Access Networks," Bell Labs Technical J., vol. 13, no. 4, pp. 175-193, 2009.

[3] C. Chou, D. Wei, C. Kuo, and K. Naik, "An Efficient Anonymous Communication Protocol for Peer-to-Peer Applications Over Mobile Ad-Hoc Networks," IEEE J. Selected Areas in Comm., vol. 25, no. 1, pp. 192-203, Jan. 2007.

[4] H. Gharavi, "Multichannel Mobile Ad Hoc Links for Multimedia Communications," Proc.IEEE, vol. 96, no. 1, pp. 77-96, Jan. 2008.

[5] S. Marti, T. Giuli, K. Lai, and M. Baker, "Mitigating Routing Misbehavior in Mobile Ad Hoc Networks," Proc. MobiCom‟00, pp. 255-265, Aug. 2000.

[6] G. Marias, P. Georgiadis, D. Flitzanis, and K. Mandalas, "Cooperation EnforcementSchemes for MANETs: A Survey," Wiley‟s J. Wireless Comm. and Mobile Computing, vol. 6, no. 3, pp. 319-332, 2006.

[7] Y. Zhang and Y. Fang, "A Secure Authentication and Billing Architecture for Wireless Mesh Networks," ACM Wireless Networks, vol. 13, no. 5, pp. 663-678, Oct. 2007.

[8] L. Buttyan and J. Hubaux, "Stimulating Cooperation in Self-Organizing Mobile Ad Hoc Networks," Mobile Networks and Applications, vol. 8, no. 5, pp. 579-592, Oct. 2004.

[9] Y. Zhang, W. Lou, and Y. Fang, "A Secure Incentive Protocol for Mobile Ad Hoc Networks," ACM Wireless Networks, vol. 13, no. 5, pp. 569-582, Oct. 2007.

[10] A. Weyland, "Cooperation and Accounting in Multi-Hop Cellular Networks," PhD thesis, Univ. of Bern, Nov. 2005.