ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 5, May 2017

A Review on: Network Traffic Classification using Bi Graph and Projection

Prital Mahadev Kadam, Prof. Prajakta Satarkar

Department of Computer Science, Solapur University / SVERI's COE, Pandharpur, India

ABSTRACT: A continuous increase of internet service is becoming critical to analyze network traffic. Due to Internet traffic growth complexity of network traffic analysis has been increased; it has become an increasingly crucial task to understand behavior patterns of internet user for different internet services and network applications. In proposed paper presents a novel approach based on behavioral graph analysis to study the behavior similarity of Internet end-hosts. Specifically, we use bipartite graphs to model host communications from network traffic and build one-mode projections of bipartite graphs for discovering social-behavior similarity of end-hosts. By applying simple and efficient clustering algorithms on the similarity matrices and clustering coefficient of one-mode projection graphs, we perform network-aware clustering of end-hosts in the same network prefixes into different end-host behavior clusters and discover inherent clustered groups of Internet applications. Proposed experiment results based on real datasets show that end-host and application behavior clusters exhibit distinct traffic characteristics that provide improved interpretations on Internet traffic. Finally, Proposed method demonstrate the practical use of understanding behavior similarity in profiling network behaviors, discovering emerging network applications, and detecting anomalous traffic patterns.

KEYWORDS: Traffic Analysis, Bipartite Graph, Bipartite Network Projection, Clustering, Internet Traffic Classification

I. INTRODUCTION

This work focuses on groups of end-hosts in the same network prefix, while some earlier studies are interested in significant individual hosts.Many insignificant hosts might not be selected for pro filing due to low traffic volume; however these hosts in the same pre fixes will be collectively analyzed in this work. The early work constructs e-mail communication graphs and employs interest-clustering algorithms for discovering e-mail users with particular interests or expertise. Reference develops an inference algorithm to search botnet communication structures from the background communication graphs constructed from the collected network traffic. Inspired by these studies, our work also uses graph analysis to uncover the social-behavior similarity among end-hosts and Internet applications.

II. OBJECTIVE OF WORK

1. Creation of source and destination communication network
2. Crating bipartite graph for network traces.
3. Generate one mode projection of source and destination network IP traces.
4. IP traces communication classification.
5. Thisdemonstrateanalysis of behavior similarityof source and destination IP traces for separately identifying network flow i n the samenetworkprefixesusingclustering.
6. One mode projectiongraphs shows communication between sources connected to same destination from the underlyingnetwork traffic.
7. Bipartite analysis has scope for showing network communication between source and destination network traces. Our work shows network traffic analysis for better network flow traces understanding.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 5, May 2017

III. LITERATURE REVIEW

Recent years have noticed rapid development of innovative and disruptive Internet services such as video streaming and peer-to-peer applications. As network traffic of these applications continues to increase, it has become a challenging task to understand their communication patterns and traffic behavior of end hosts engaging in these applications. This paper present a novel technique based on behavioral graph analysis to study social behavior of Internet applications based on bipartite graphs and one-mode projection graphs. Using a vector of graph properties including coefficient clustering that capture social behaviors of end users, we discover the inherent clustered groups of Internet applications that not only exhibit similar social behavior of end user, but also have similar features in the summarized traffic. We show the usage of the proposed approach in detecting need for applications and anomalous traffic patterns towards Internet applications[1]. Knowing the structure and dynamics of the user behavior networks for web traffic that connect users with servers across the Internet network is a key to modeling the network and designing future application. The Web-visited bipartite networks, called the user behavioral networks, display a natural bipartite structure: two types of nodes coexist with links only between nodes of different types. We obtained the result that the out-degree distribution of users the in-degree distribution of servers and the strength distribution are approximately power-law, whose exponential is between 1.7 and 3.4. The clustering coefficient of clients and servers is larger than that in randomized, degree preserving versions of the same graph, which indicate a modular structure of UBNWT. Finally, based on the algorithm of finding the community structure in bipartite network, we divided the clients into different communities, through manual examination of hosts in these communities, the typical normal and abnormal communities were found. Interestingly, the loyalty of clients belonging to the same community in different time is higher than 80%. The structure analysis of UBNWT is very helpful for the network management, resource allocation, traffic engineering and security[2].Identifying groups of Internet hosts with a similar behavior is very useful for many applications of Internet security control, such as DDoS defense, worm and virus detection, detection of botnets, etc. There are two major difficulties for modeling host behavior correctly and efficiently: the huge number of overall entities, and the dynamics of each individual. In this paper, This paper present and establish the Internet host network profiling problem using the header data from public packet traces to select relevant features of frequently-seen hosts for profile creation, and using hierarchical clustering techniques on the profiles to build a containing all the hosts. The well-known agglomerative algorithm is used to discover and combine similarly-behaved hosts into clusters, and domain-knowledge is used to analyze and evaluate clustering results. This paper shows the result of applying the proposed classificationmethodology to a data set from NLANR-PMA Internet traffic archive with more than 60,000 active hosts. On this dataset, our approach successfully identifies clusters with significant and interpretable features. Proposed system next uses the created host profiles for detecting anomalous behavior during the Slammer worm spread. The experimental results show that our profiling and clustering approach can successfully detect Slammer outbreak and identify majority of infected hosts[3].Extensive research has been conducted since then and many applications have been developed. We have reviewed an extensive number of studies with emphasis on network flow applications. First, we provide a brief introduction to sFlow, NetFlow and network traffic analysis. Then, we review the state of the art in the field by presenting the main perspectives and methodologies. Our analysis has revealed that network security has been an important research topic since the beginning. Newapproaches, such as machine learning, have been very promising. This workexplore a critique of the studies surveyed about data sets, perspectives, methodologies, challenges, future directions and ideas for potential integration with other Information Technology infrastructure and methods. Finally, we concluded this survey [4]. Graph-based techniques and analysis have been used for IP network traffic analysis. The objective of this paper is to study the hosts' interaction behavior and use graph clustering Algorithm, the Markov clustering algorithm, to group (cluster) hosts which have interaction using the HTTP protocol. Using real network traces, the clustering results show that MCL algorithm successfully group the hosts to their corresponding clusters. Analyzing the clustering results, it is showed that communications between one source IP address to one destination IP address, one source IP address to several (different) destination IP addresses, and several (different) source IP addresses to one destination IP address, are grouped to their own clusters[5]. Recent spates of cyber-attacks and frequent emergence of applications affecting Internet traffic dynamics have made it imperative to develop effective techniques that can ex-tract, and make sense of, significant communication patterns from Internet traffic data for use in network operations and security management. In this paper, we present a general

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 5, May 2017

methodology for building comprehensive behavior profiles of Internet backbone traffic in terms of communication patterns of end-hosts and services. Relying on data mining and entropy-based techniques, the methodology consists of significant cluster extraction, automatic behavior classification and structural modeling for in-depth interpretive analyses. System results verify the methodology using data sets from the core of the Internet [6].

IV. RELATED WORK

Bipartite graph and one mode projection focuses on groups of internet end-hosts in the same network area, while some earlier studies are interested in significant individual hosts.Many insignificant hosts might not be selected for profiling because of low internet traffic volume; however these hosts in the same network prefixes will be collectively analyzed in the proposed system. The early work constructs email communication graphs and employs interest-clustering algorithms for discovering e-mail users with particular interests or expertise. Proposed work develops an intermediate algorithm to search end host communication structures from the background communication graphs constructed from the collected network traffic. Proposed work also uses graph analysis to recover the social behavior similarity between internet end-hosts and Internet application service.

Internet network source and destination hostandapplications are increasingly day by day, itbecomes crucial task toknow thetraffic behaviorof end-hosts and network applications for efficient network management and security monitoring. A number of research studieshave workedon traffic behavior analysis of individual hosts and applications. However, a growing huge amoun to fend-hosts, awide diversity of applications, and massive traffic dataposes significant challenges for such traffic analysis for back bone networks or enterprise networks.

V. PROPOSED WORK

System formulates the standard bipartite graph representation of communication patterns in computer network traffic. Specifically, Let GB = (V;U;E B), where vertices v2V represent source IPs (srcIP), vertices u2Urepresent destination IPs (dstIP), and edges fv;ug2EB represent data flows from sources v to destinations u. In constructing such a graph GB from data, we assign a vertexv2Vfor every unique IP address that played a role as aProposed algorithm uses an edge fv; u g to EB if and only if there is a flow in our data from v to u. as per rules, multiple edges arise when there are having multiple flows between source and destination IP address However, proposed cluster identify reliable way in which use this structure by assigning edge weights equal to the multiplicity of an edge. Following diagram shows description of such a graph GB, belong to small connected component used from the dataset. There are 10 source nodes (i.e., nodes 1 through 10), 4 destination nodes (i.e., nodes a through d), and 13 edges, corresponding to 25 flows, with weights ranging from 1 to 4. Note that sources share destinations to varying extents. For example, sources 1;2;3, and 10 all talk only to destination a, while sources 4 and 6 talk only to destination c. This technique has purpose to the traffic in each of the two subsets of corresponding flows, and hence recommends in turn that some portion of same network prefix which may connect to these two sets of sources. However, while source 9 also communicates to Destination and source 7 also talks to destination c, they each talk as well to destination b. This observation suggests that, while sources 7 and 9 participate in the communities defined around destinations a and c, they do not necessarily belong to those communities. The first is a one mode projection of GB, in format of an undirected graph GP= (V; EP), where nodes vi and vj are connected if and only if they share at least one common destination. The one-mode projection of the bipartite graph in Figure Note that under this representation the types of ‘communities’ we identified, i.e., source nodes that all communicate with a common destination, exhibit a distinct topological structure in GP , in the form of cliques. For example, nodes 1;2;3;9; and 10 form a five-clique, while nodes 4;6, and 7 form a three-clique.

Wedemonstrate practical benefitsofexploring behaviorsimilarity ofInternet end-hosts inprofilingnetwork pre- fixesandemergingapplications anddetectinganomalous trafficpatternssuchas scanningactivities,worms,or denial-of-service attacksthroughsynthetictraffictraces.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 5, May 2017

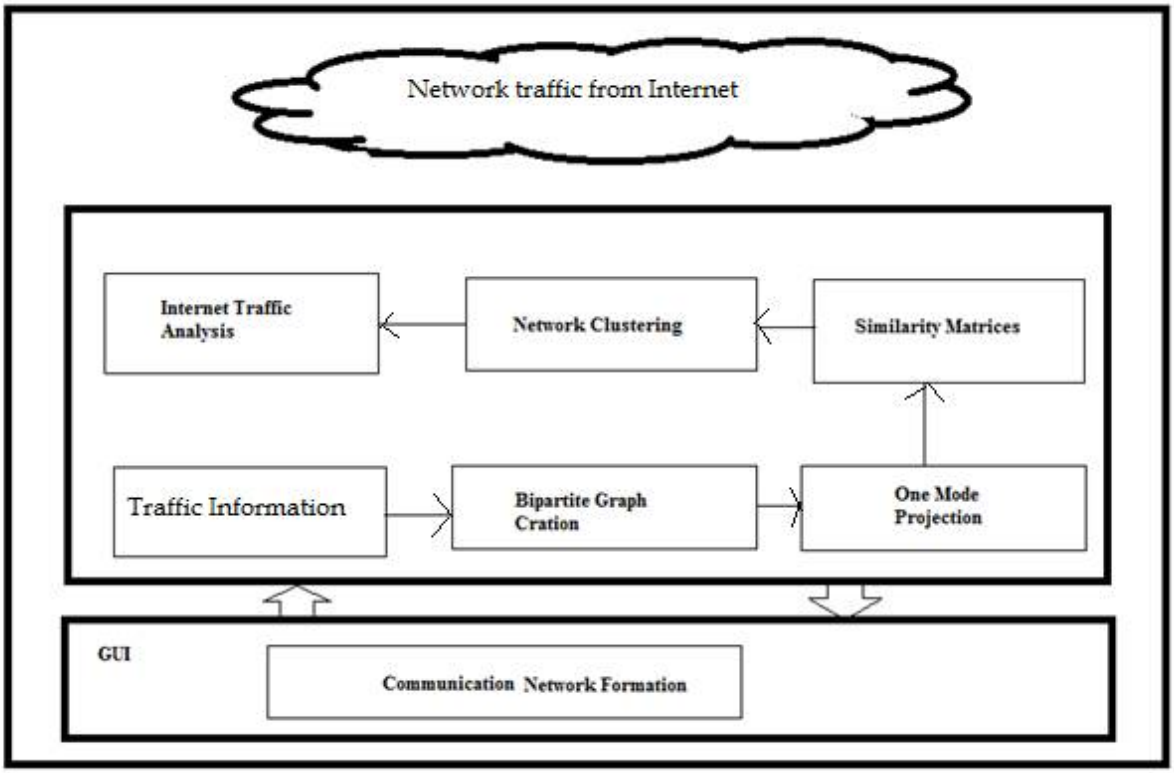

Figure01 : - ArchitectureofSystem

VI. CONCLUSION AND FUTURE WORK

In this application we propose bipartite graphs and one-mode projection graphs to determine traffic for social behavior of end hosts communicating in the same Internet applications. By making use of end host classification and other graph properties, we search novel similarity matrices of social behavior among different internet end host, and then apply a simple category clustering algorithm to group applications with similar social behavior into different clusters. For designing source address and destination address communication pattern graphs for each application port. Hence in our proposed method implement a way to understand the social behavior of source and destination hosts used in the same applications for different communication. Moreover, classification of these applications depends on coefficient classification of source and destination behavior graphs into different grouping help to understand unknown applications and source and destination address that follows same communication patterns with well-known applications. To implement the quality of the clustering results, proposed work use traffic features of application clusters and compare the similarity in network traffic features from different destination end host ports in the same clusters as well as the dissimilarity among ports in different clusters.

REFERENCES

[1] K. Xu, F. Wang, and L. Gu, "Network-aware behavior clusteringof Internet end hosts," in Proce. IEEE INFOCOM, Apr. 2011, pp.2078–2086.
[2] K. Xu and F. Wang, "Behavioral graph analysis of internet applications," in Proc. IEEE GLOBECOM, Dec. 2011, pp. 1–5.
[3] S. Wei, J. Mirkovic, and E. Kissel, "Profiling and clustering internet hosts," in Proc. Int. Conf. Data Mining, 2006, pp. 269–275.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 5, May 2017

[4] K. Xu,Z.-L. Zhang, andS.Bhattacharyya, "Profiling internet backbone traffic: Behavior models and applications," in Proc. ACM SIGCOMM, Aug. 2005, pp. 169–180.
[5] H. Jiang, Z. Ge, S. Jin, and J.Wang, "Network prefix-level traffic profiling: Characterizing, modeling, and evaluation," Comput. Netw., vol.54, no. 18, pp. 3327–3340, 2010.
[6] Y. Jin, E. Sharafuddin, andZ.-L. Zhang, "Unveiling core network-wide communication patterns through application traffic activity graph decomposition," in Proc. ACM SIGMETRICS, Jun. 2009, pp. 49–60.