International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056
Volume: 04 Issue: 06 | June -2017 www.irjet.net p-ISSN: 2395-0072

Network Traffic Analysis by Bipartite Graph and One mode Projection

Ms. Kadam Prital M.[1] , Prof. Mrs Satarkar Prajakta A.[2]

[1]ME-Computer SVERI's COE, Pandharpur, University Solapur, Maharashtra,India
[2]Asst. Prof. SVERI's COE, Pandharpur, Solapur University Solapur, Maharashtra,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.*

Key Words: Traffic Analysis, Bipartite Graph, Bipartite Network Projection, Clustering, Internet Traffic Classification

1. 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.

1.1 MOTIVATION

Bipartite graphs are used for modeling data communication in network traffic and the one-mode projection for capturing behavior similarity of end-hosts.Clustering algorithm is used to calculate similarity matrices and their coefficient for discovering behavior clusters of end-hosts in the same prefixes or engaging in the same applications.The different features of end user behavior cluster within the same network area and uses behavior similarity to discover traffic patterns in network area and detect anomalous behaviors.

1.2 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 1 social behavior similarity between internet end-hosts and Internet application service.

Internet network source and destination host and applications are increasingly day by day, it becomes crucial task to know the traffic behavior of end-hosts and network applications for efficient network management and security monitoring. A number of research studies have worked on traffic behavior analysis of individual hosts and applications. However, a growing huge amount of end-hosts, a wide diversity of applications, and massive traffic data poses significant challenges for such traffic analysis for backbone networks or enterprise networks.

2. PROBLEM STATEMENT

To analyze the traffic behavior for network prefixes that include end-hosts with the same network bits in their IP addresses, we could further decompose the bipartite graph of all the traffic into a set of smaller disjoint bipartite sub graphs such that each bipartite sub graph captures the host communications for a single source or destination IP prefix x, e.g., source behavior graph (SBG) and destination behavior graph (DBG) representing the bipartite sub graphs of host communications for the source IP prefix and the destination IP prefix x , respectively.

2.1 PROPOSED SYSTEM

We demonstrate practical benefits of exploring behavior similarity of Internet end-hosts in profiling network pre-fixes and emerging applications and detecting anomalous traffic patterns such as scanning activities, worms, or denial-of-service attacks through synthetic traffic traces.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 a Proposed 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.

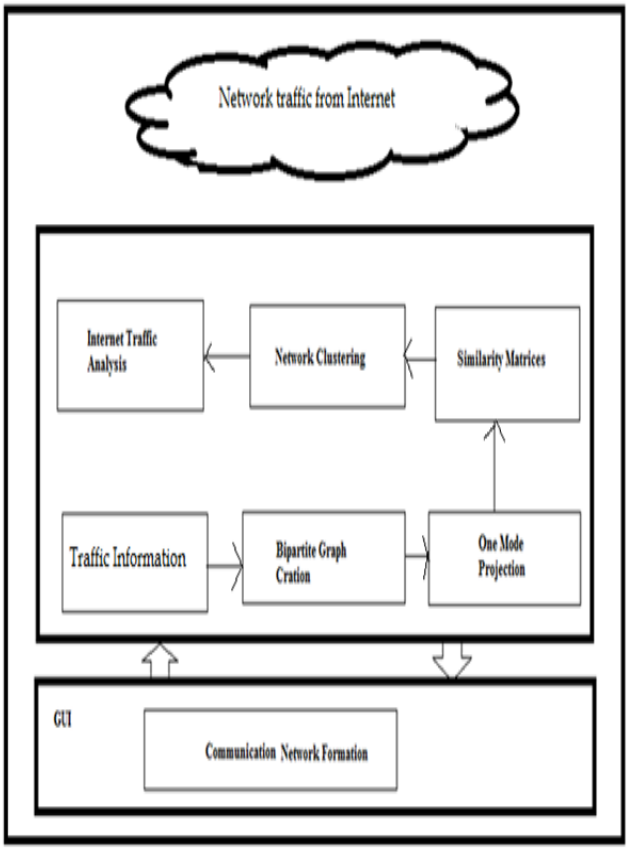

Fig. System Architecture

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 standard one-mode projection of GB, in the form 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. Note too that nodes 7

and 9, which we identified as exhibiting an example of antisocial behavior, stand out clearly as being both members of their respective cliques and, at the same time, connected by a single edge to each other's cliques.

2.2ALGORITHM:-

Input:-

1. Construct bipartite graph of host communication from flow traces;
2. Generate one mode projection of bipartite graph and its weighted adjacency matrix for end host in prefix and the obtain similarity matrix in prefix.
3. Construct diagonal matrix for adjacency matrix.
4. Compute laplacian matrix with adjacency matrix-degree of host matrix.
5. Normalize the row of matrix for clustering.
6. Perform clustering for source and destination end host for communication.
7. Assign the original IP address for clustering for row according to laplacian matrix.
8. Clustered IP address from available source and destination addresses.

Output:

Group of similar source and destination hosts.

2.3 RESULT AND DISCUSSION

We consider bipartite graphs to represent host communications from network traffic and build one-mode projections of bipartite graphs for understanding social-behavior similarity of end-hosts. By applying simple and

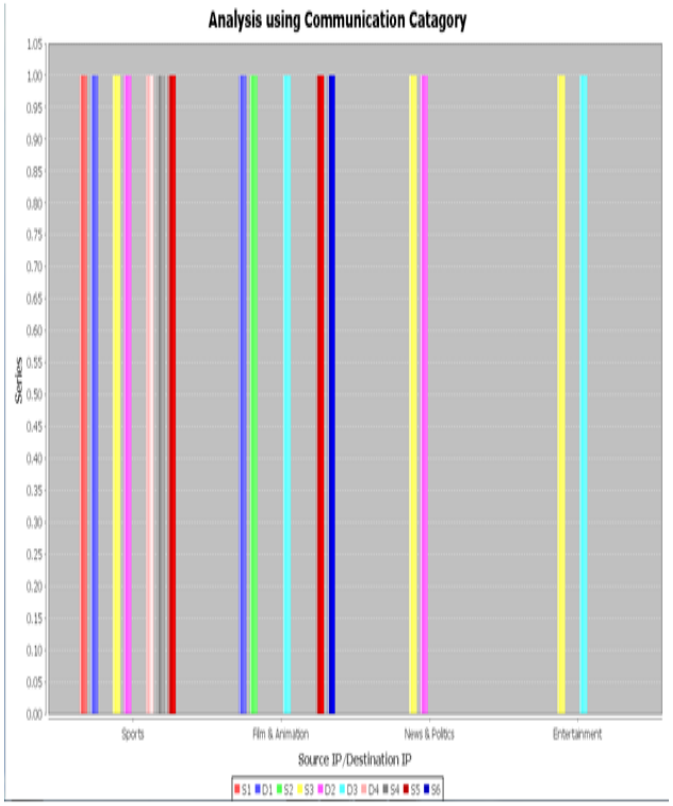

Fig: Analysis of communication pattern in internet traffic.

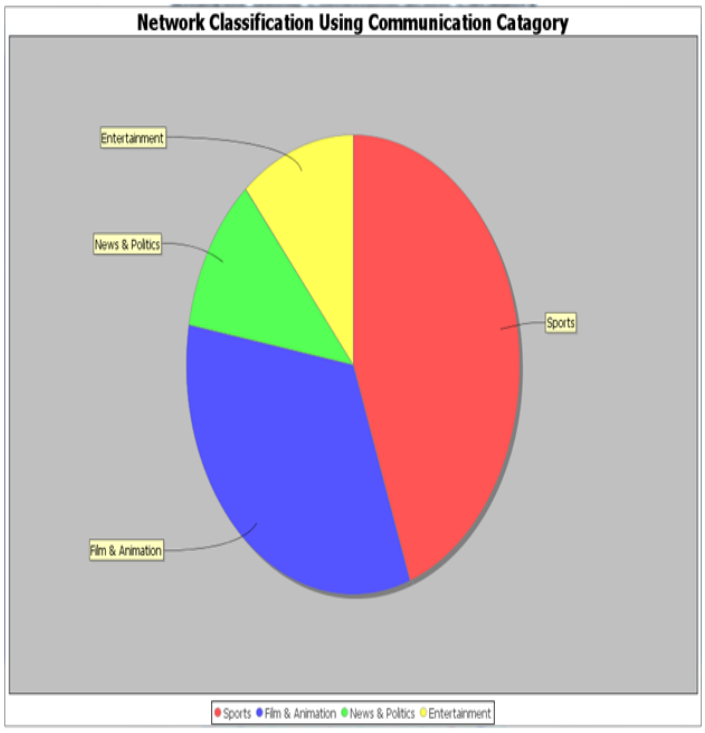

Fig. Classification of Internet traffic according destination services

efficient clustering algorithms on the similarity matrices and clustering coefficient of one-mode projection graphs, we perform network based clustering of end-user in the same network region with different end-user behavior clusters and discover inherent clustered groups of Internet end user for different application.

3. CONCLUSIONS

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 clustering of 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.

[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, and Z.-L. Zhang, "Unveiling core network-wide communication patterns through application traffic activity graph decomposition," in Proc. ACM SIGMETRICS, Jun. 2009, pp. 49–60.