International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2013): 6.14 | Impact Factor (2014): 5.611

# Behavior Analysis of Internet Traffic Using Graph and Similarity Matrix

**Tejashree A. Deshmukh[1], Shubhangi Suryawanshi[2]**

[1]ME Computer (Networks), Savitribai Phule Pune University, G.H Raisoni College of Engg. & Technology, Wagholi, Pune,

[2]Assistant Prof, Department Of Computer Engineering, Savitribai Phule Pune University, G. H .Raisoni Collage of Engg. & Technology, Wagholi, Pune

**Abstract:** ***As system activity of utilizations, for example, video spilling and distributed applications proceeds to grow it is difficult task to understand behavior patterns of end host and other network application. Keeping in mind the end goal to comprehend conduct closeness of Internet end has in the same system prefixes bipartite diagrams are utilized to model system activity, and after that one-mode projection diagrams are built for catching social-conduct closeness of end has. At that point by applying a basic and effective algorithm and similarity matrices and cluster coefficient of one mode projection graph, inherent clustered groups of internet application are discovered. Through a vector diagram including coefficient bunching that catch social practices of end hosts, the clusters are found that not just display comparable social conduct of end hosts, additionally have comparable attributes in the accumulated activity.***

**Keywords:** Bipartite graph, Network Security, clusters, One Mode Projection, Profiling

## 1. Introduction

A diagram G is a couple of sets (V, E), where V is a limited nonempty set of articles called vertices (or hubs) and E is a 2-component subsets of V called edges (or connections). V is the vertex set, V(G) = {v1 , v2 , … , vn}, where n = |V| is the quantity of vertices, which frequently called the request of diagram G. E is the edge set, E(G) = {e1 , e2 , … , em}, where m = |E| is number of edges, which regularly called the measure of chart G. The edge, e is composed as 2-component set {u, v}. The vertices u and v are alluded to as neighbors of one another. For this situation, the vertex u (or v) and the edge e = uv are said to be episode with one another. Two vertices u and v is said to be contiguous if there is an edge between vertex u and vertex v. A chart is known as a coordinated diagram (digraph) is a diagram when the edges are requested sets. Else, it is called undirected chart. Bunching is unsupervised method for gathering related unlabeled information (articles) such that, information with most-comparative qualities have a place with the same group and information with most disparate attributes will be in different bunches. Accordingly, a group is an accumulation of information which have high similitude in the middle of them and not at all like information fitting in with different bunches [5]. Chart bunching is the undertaking of collection the diagram's vertices into groups thinking seriously about of the chart's edge structure such that, there ought to be numerous edges inside of every bunch and moderately few between the bunches.

Concentrating on grouped activity conduct in system prefixes not just lessens the quantity of conduct profiles for examination contrasted and host-level activity profiling, additionally uncovers some examples for a gathering of end hosts having comparative practices analyzed with system level movement profiling. we utilize bipartite diagrams to show social-conduct of Internet end hosts and we determine one-mode projection diagrams to catch the social-conduct closeness of host interchanges through edges between source (or destination) has that discussion to the same destinations (or sources). Bipartite diagrams are utilized for displaying information correspondence as a part of system activity and the one-mode projection for catching social conduct likeness of end hosts in the same prefixes [1]. The errand of discovering great groups has been centre for machine learning and example acknowledgment. However ghastly systems are utilized for clustering.[6]

### A.Bipartite Graphs of Host Communications

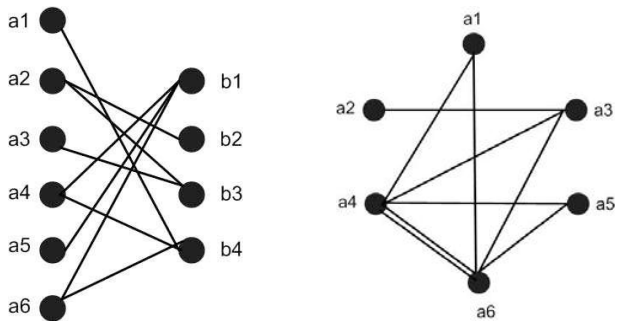


**Figure:** (a) Bipartite graph (b) One-mode projection

Fig. 1. (a): A sample of bipartite charts in light of host correspondences between the source has (a1 - a6) and the destination has (b1 - b4); (b):The one-mode projection on the vertex set of the left-side hubs, i.e., the source has (a1 - a6) The host interchanges saw in system movement of Web spine connections or Internet-confronting connections of fringe switches for big business systems could be actually displayed with a bipartite diagram G = (A, B, E), where An and B are two disjoint vertex sets, and E ⊆ A×B is the edge set [6].

In particular, all the source IP locations frame the vertex set A, while the vertex set B comprises of all the destination addresses. Each of the edges, ek in G unites one vertex ai ∈ An and another vertex bj ∈ B. To break down the activity

conduct for system prefixes which incorporate end has with the same system bits in their IP locations, then decay the bipartite chart of all the activity into an arrangement of littler disjoint bipartite sub diagrams such that each bipartite sub chart catches the host interchanges for a solitary source or destination IP prefix, e.g., GP = (AP, B, EP) and GQ = (A, BQ, EQ) speaking to the bipartite sub charts of host correspondences for the source IP prefix P and the destination IP prefix Q, respectively[1].

**B. One-mode projection of bipartite graphs:**

To contemplate the conduct comparability of end hosts in the same system prefixes, it is important to influence one-mode projection diagrams of bipartite diagrams that are utilized to concentrate concealed data or connections between hubs inside of the same vertex sets [6]. Figure 1[a] demonstrates a straightforward bipartite diagram that is created in light of host interchanges between the six source has (a1 - a6) and the four destination has (b1 - b4), while Figure 1[b] outlines the one-mode projection of the bipartite chart on the vertex set of the six left-side hubs, i.e., the source has (a1 - a6). An edge unites two hubs in the one mode projection if and just if both hubs have associations to no less than one same hub in the bipartite diagram. One could effectively draw the one-mode projection diagram for the vertex set of the right-side hubs utilizing the same procedure. The one-mode projection of the bipartite diagrams uses edges between end has in the same prefixes to evaluate the closeness of their system association patterns [1].

**C. Discovering behavior clusters using clustering algorithm:**

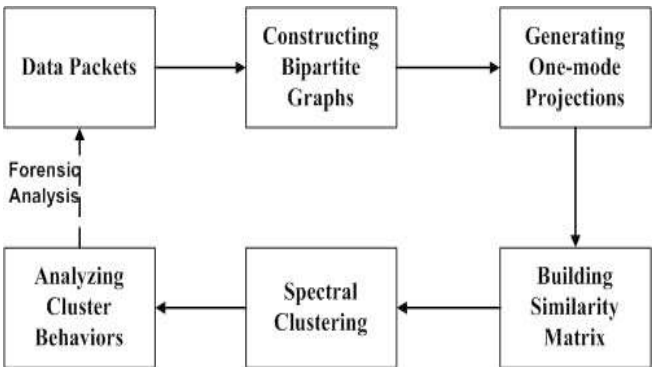


Figure illustrates the schematic process of the algorithm from constructing bipartite graphs based on IP packets to analyzing behavior clusters of network prefixes. We focus on the social-behavior of end hosts in data communications through bipartite graphs and one-mode projection graphs, and are interested in exploring the social-behavior similarity of end hosts to discover inherent traffic clusters in the same network prefixes. An important starting point of a clustering algorithm is to define the appropriate similarity metrics between data points. In this paper we use the weighted edge between two hosts u and v of the same prefix in the one-mode projection graph as the similarity measure su,v between u and v, because the weighted edges capture and quantify the social-behavior similarity of host communications in network traffic.

## 2. Literature Survey

**1. Network aware behavior clustering of Internet end hosts**

They used bipartite graphs to model network traffic, and then construct one-mode projection graphs for capturing social-behavior similarity of end hosts. By applying a simple and efficient spectral clustering algorithm, he performed network-aware clustering of end hosts in the same prefixes into different behavior clusters. Based on information-theoretical measures, found that the clusters exhibit distinct traffic characteristics which provides improved interpretations of the separated traffic compared with the aggregated traffic of the prefixes. Finally, demonstrated the applications of exploring behavior similarity in profiling network behaviors and detecting anomalous behaviors through synthetic traffic that combines Internet backbone traffic and packet traces from real scenarios of worm propagations and denial of service attacks. This paper explores the behavior similarity of Internet end hosts in the same network prefixes. They used bipartite graphs to model network traffic, and then construct one-mode projection graphs for capturing social-behavior similarity of end hosts. By applying a simple and efficient spectral clustering algorithm, performed network-aware clustering of end hosts in the same prefixes into different behavior clusters. Based on information-theoretical measures, they found that the clusters exhibit distinct traffic characteristics which provide improved interpretations of the separated traffic compared with the aggregated traffic of the prefixes. Finally, demonstrated the applications of exploring behavior similarity in profiling network behaviors and detecting anomalous behaviors through synthetic traffic that combines Internet backbone traffic and packet traces from real scenarios of worm propagations and denial of service attacks.[1]

**2. Behavioral graph analysis of internet applications**

Recent years have witnessed rapid growth of innovative and disruptive Internet services such as video streaming and peer-to-peer applications. As network traffic of these applications continues to grow, it has become a challenging task to understand their communication patterns and traffic behavior of end hosts engaging in these applications. This paper presents a novel approach based on behavioral graph analysis to study social behavior of Internet applications based on bipartite graphs and one-mode projection graphs. Through a vector of graph properties including coefficient clustering that capture social behaviors of end hosts, we discover the inherent clustered groups of Internet applications that not only exhibit similar social behavior of end hosts, but also have similar characteristics in the aggregated traffic[7].

**3. Profiling and clustering internet hosts**

The objective of this research is to study the behavior of IP Network nodes (IP hosts) from the prospective of their communication behavior patterns to setup hosts' behavior profiles of the observed IP nodes by clustering hosts into clusters of similar communication behaviors. The problem of IP address behavior analysis and profile establishment is the one that not fully discussed and the results achieved are not good enough, there is no complete solution yet. There are

many potential applications of this work, the results of this research will be useful to the network management and Network security situation awareness in addition to the applications in studying the network user behavior. This paper includes: 1) Discussion about the features or host behavior communication patterns to be utilized in hosts clustering to characterize accurately and efficiently groups of host behavior traffic. 2) We presented an algorithm to extract most significant IP nodes to be analyzed instead of analyzing the complete list of millions of IP nodes that exist in the trace. 3) We analyzed IP nodes traffic behavior on relatively long periods of traces, which help to extract a more stable host's behavior[8].

**4. Profiling internet backbone traffic: Behavior models and applications**

Recent spates of cyber-attacks and frequent emergence of applications affecting Internet traffic dynamics have made it imperative to develop effective techniques that can extract, and make sense of, significant communication patterns from Internet traffic data for use in network operations and security management. In this paper, there is a general 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 information-theoretic techniques, the methodology consists of significant cluster extraction, automatic behavior classification and structural modeling for in depth analyses. We validate the methodology using data sets from the core of the Internet. The results demonstrate that it indeed can identify common traffic profiles as well as anomalous behavior patterns that are of interest to network operators and security analysts[13].

**5. Bipartite graphs as models of complex networks**

It appeared recently that the classical random graph model used to represent real-world complex networks does not capture their main properties. Since then, various attempts have been made to provide accurate models. We study here the model which the following challenges: it produces graphs which have the three main wanted properties (clustering, degree distribution, average distance), it is based on some real-world observations, and it is simple to make it possible to prove its main properties. This model consists in sampling a random bipartite graph with prescribed degree distribution[4].

**6. Internet traffic behavior profiling for network security monitoring**

Recent spates of cyber-attacks and frequent emergence of applications affecting Internet traffic dynamics have made it imperative to develop effective techniques that can extract, and make sense of, significant communication patterns from Internet traffic data for use in network operations and security management. In this paper, they present a general 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[9].

## 3. Conclusion

Bipartite graphs and one-mode projection metod investigate social conduct of end hosts taking part in the same Internet applications. Through clusters and other graph properties, there is comparability of social conduct among distinctive applications. Web applications in the same groups likewise have comparative attributes in the movement. We exhibit similarity of social behavior among different applications. Internet applications in the same clusters also have similar characteristics in the aggregated traffic. Utilizing bipartite charts and one-mode projection diagrams and by applying unearthly bunching calculations on the one-mode projection, we locate the clustered behavior of end hosts in the same system prefixes.

## 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] A. Ng, M. Jordan, and Y.Weiss, "On spectral clustering: Analysis and an algorithm," in *Proc. NIPS*, 2001, pp. 849–856.
[3] Yessica Nataliani,Theophilus Wellem " HTTP Traffic Graph Clustering using Markov Clustering Algorithm ", I*nternational Journal of Computer Applications (0975 – 8887) Volume 90 – No 2, March 2014*
[4] J.-L.Guillaume and M. Latapy, "Bipartite graphs as models of complex networks," *Physica A, Stat. Theor. Phys.*, vol. 371, no. 2, pp. 795–813,2006.
[5] Y. Dong and Y. Zhuang, "Fuzzy hierarchical clustering algorithm facing large databases," in Fifth World Congress on Intelligent Control and Automation, 2004. WCICA 2004, vol. 5, pp. 4282-4286, June 2004.
[6] K. Xu,Z.-L. Zhang, andS.Bhattacharyya, "Profiling internet backbone traffic: Behavior models and applications," in *Proc. ACM SIGCOMM*, Aug. 2005, pp. 169–180.
[7] K. Xu and F. Wang, "Behavioral graph analysis of internet applica- tions," in Proc. IEEE GLOBECOM, Dec. 2011, pp. 1–5.
[8] S. Wei, J. Mirkovic, and E. Kissel, "Pro filing and clustering internet hosts," in Proc. Int. Conf. Data Mining, 2006, pp. 269–275.
[9] K. Xu, Z.-L. Zhang, and S. Bhattacharyya, "Internet traffic behavior profiling for network security monitoring," IEEE/ACM Trans. Netw., vol. 16, no. 6, pp. 1241–1252, Dec. 2008.
[10] A. Ng, M. Jordan, and Y.Weiss, "On spectral clustering: Analysis and an algorithm," in *Proc. NIPS*, 2001, pp. 849–856.
[11] J. Ramasco, S. Dorogovtsev, and P. Romualdo, "Self-organization of collaboration networks," *Phys. Rev.*, vol. 70, no. 3, p. 036106, 2004.
[12] M. Kaiser, "Mean clustering coefficients: The role of isolated nodes and leafs on clustering measures for small-world networks," *New J.Phys.*, vol. 10, pp. 083042–083052, Aug. 2008.
[13] K. Xu, Z.-L. Zhang, and S. Bhattacharyya, "Profiling Internet Backbone Traffic: Behavior Models and

Applications," in *Proceedings of ACM SIGCOMM*, August 2005.
[14] H. Jiang, Z. Ge, S. Jin, and J. Wang, "Network Prefix-level Traffic Profiling:Characterizing, Modeling, and Evaluation," *Computer Networks*,2010.
[15] Y. Jin, E. Sharafuddin, and Z.-L. Zhang, "Unveiling core networkwide communication patterns through application traffic activity graph decomposition," in *Proceedings of ACM SIGMETRICS*, June 2009.