RESEARCH ARTICLE OPEN ACCESS

# A Review of Peer to Peer Traffic Identification Method Using K-Means Clustering, SVM & Genetic Algorithm

Puja Raghav [1], Dr. Vivek Jaglan [2], Mr. Akshat Aggarwal [3]
Amity school of Engineering & Technology
Amity University Haryana

## ABSTRACT

The need of shared (P2P) applications is growing significantly, which brings major concerns, for example, the system blockage and traffic hindrance. Consequently P2P traffic identification is the most blazing theme of P2P traffic identification. SVM is basically dependent on its parameters execution and selection. In this paper we have proposed a combined approach using SVM, K-means & genetic algorithm for the purpose of P2P traffic identification. The novelty or peculiarity of the proposed methodology lies in the extent of packets traded between IPs inside within seconds. The supported parameters of the proposed technique lie in that quick calculation, high identification precision, and observations that had high accuracy. At last, experiment results demonstrate the reasonable performance of the proposed method.

***Keywords:-*** Peer to Peer networking (P2P), support vector machine (SVM) genetic algorithm & k-means clustering algorithm.

## I. INTRODUCTION

### *Overview*

A distributed (P2P) system is a bundle of PCs where every PC go about as a hub for sharing documents inside the gathering. Regardless of having a focal server which goes about as a mutual drive, every PC go about as a server and customer for the documents put away upon it. In a p2p arrange the assignment or assets are shared or designated similarly and the system is built up over the Internet.

Essentially, a p2p system is basic system where every hub has given similarly advantaged to play out the assignment, records and assets on a server are partaken in like manner way. P2P systems are said to be similarly as home system or workplaces arrange. The measure of the system and the documents accessible on the system enable gigantic measure of information to be shared at whatever point the P2P systems are built up on the system. Prior P2P systems like Napster utilized the customer programming and a focal server. Later-on, Kazaa and BitTorrent began sharing obligations between numerous hubs to free up the transfer speed. P2P system are typically connected with web assaulting, unapproved utilize and illicit record sharing. The starting utilization of P2P systems was conveyed in the mid-1980s in business of unattached or autonomous PCs. In variety to the day to the scaled down casings of the day, for example, the VS framework from Wang Laboratories Inc., which served up word preparing and different applications to moronic terminals from a focal PC and put away records on a focal hard drive, then the new PCs has independent hard drives and worked in CPUs. The savvy boxes likewise had on-board applications, which implied they could be conveyed to desktops and be valuable without an umbilical string connecting them to a centralized server. In the most straightforward way, a distributed system is framed at whatever point at least two PCs are connected & share their resources without the need of a central server computer. A P2P network can be an ad hoc connection—a couple of computers connected via a Universal Serial Bus to transfer files. A P2P network also can be a permanent infrastructure that links half-dozen computers in a small office over copper wires. Or a P2P network can be a network of much larger scale in which special protocols and applications set up direct relationships among users over the Internet.

In a P2P arrange, the "companions" are PC frameworks which are related with each other by methods for the Internet. Documents can be shared particularly between frameworks on the system without the need of a focal server. Toward the day's end, each PC on a P2P arrange transfo rms into a record server and what's more a customer. The primary requirements for a PC to join a conveyed framework are an Internet affiliation and P2P programming. Ordinary P2P programming programs fuse Kazaa, Limewire, BearShare, Morpheus, and Acquisition. These ventures connect with a P2P framework, for instance, "Gnutella," which empowers the PC to get to a colossal number of various structures on the framework. At the point when related with the framework, P2P programming empowers you to search for records on other people's PCs. At that point, diverse customers on the framework can filter for records on your PC, yet normally simply inside a singular envelope that you have relegated to share. While P2P sorting out makes record sharing straightforward and favorable, is also has provoked a lot of programming burglary and unlawful music downloads. In this way, it is best to be erring on the side of caution of alert and simply download programming and music from true blue locales.

Shared (P2P) is a decentralized correspondence demonstrate in which each gathering hosts comparative capacities and either get-together can begin a correspondence session. Not in the slightest degree like the customer show, in which the customer makes an administration ask for and the server fulfills the request, the P2P organize exhibit empowers each hub to act as both a customer and server. P2P frameworks can be used to give anonymized directing of activity development, enormous parallel processing conditions, and conveyed limit and distinctive capacities. Most P2P projects are focused on media sharing and P2P is thusly consistently associated with programming and copyright encroachment. Usually, shared applications empower customers to control various parameters of operation: what number of part relationship with search for or allow at one time; whose frameworks to interface with or stay away from; what administrations to offer; and what number of framework resources for provide for the system. Some simply connect with some subset of dynamic hubs in the system with little customer control, regardless.

## II.LITERATURE REVIEW

An easy way to comply with the conference paper formatting requirements is to use this document as a template and simply type your text into it. The role of P2P has proven to be positive and significant in real time communication, file sharing, business, web search engines and content distribution. It is important have an efficient more enhanced of building IP network traffic. In the next sub heading we will be discussing about the related work done by different authors in past.

### *A RELATED WORK:*

**Jie Cao, Zhiyi Fang, Dan Zhang, and Guannan Qu [01],**

Organize movement characterization is the foundation of many system investigate works. As of late, the investigation on movement grouping and distinguishing proof in light of machine learning strategy is another examination course. Bolster Vector Machine (SVM) is the one of the machine learning procedure which performs incredible accuracy and dependability. Be that as it may, the conventional grouping execution of SVM is not impeccable, this paper proposed an upgraded strategy which can improve the execution of SVM essentially. The paper proposed subset with wrapper approach and figured the perfect working parameters thus in light of matrix inquiry calculation. We associated this methodology to two-class SVM classifier. The reenactment result happens affirmed that most of the streams' ordinary precision accomplishes 99.64%, typical component estimation decreases 20% than one of a kind estimation and ordinary snuck past time is shorter 98.88% than conventional SVM.

**Prof S. R. Patil, Suraj Sanjay Dangat [02],** The P2P (Peer-to-Peer) system is dynamic, self-composed and has some different components. Along these lines, P2P traffic has turned out to be a standout amongst the most significant portions of the network traffic. But, it has also caused network blockage issues as a result of asset occupation (for the most part data transmission). Accurate identification of traffic makes great sense for proficient network management and effective utility of network resources. In this paper we address traffic identification technology, for example, traffic identification by using network characteristics for solving the issue with various parameters, enhanced exactness rate and productivity. Experiments on different P2P applications exhibit that the strategy is generic and it can be connected to a large portion of P2P applications. Exploratory outcome demonstrates that the algorithm can identify P2P application precisely. In this paper we first quickly present P2P innovation and afterward we made a short review on the general advance in P2P traffic recognizable proof advances. At long last we examine the proposed strategy and its outcome investigation.

**Satoshi Ohzahata, Yoichi Hagiwara, Matsuaki Terada, and Konosuke Kawashima [03],** Pure P2P applications are comprehensively used nowadays as a file sharing structure. In the overlay frameworks, music and video files are the essential things exchanged, and it is understood that the traffic volume is significantly greater than that of customary customer/server applications. Regardless, the present status of the P2P application traffic is not remarkable an immediate consequence of their unknown correspondence designs. In particular, in circumstances where the application does not use the default benefit port, and the correspondence course and the mutual file are in like manner mixed, the identification traffic has not been practical. To deal with this issue, we have developed an identification method for Pure Peer-to-Peer correspondence applications, especially for traffic for Winny, the most surely understood Peer-to-Peer application in Japan, by using server/customer associations among the companions.

**Joseph Stephen Bassi, Loo Hui Ru, Khammas, Muhammad, Nadzir Marsono [04],** Shared (P2P) applications are bandwidth-overwhelming and prompt network blockage. The disguising way of P2P traffic makes conventional strategies for its Identification futile. So as to manage and control P2P traffic effectively ideally in the network, it is important to identify such traffic on the web and precisely. This paper proposes a technique for online P2P identification based in view of traffic events signatures.

**Jinghua Yan, Zhigang Wu, Hao Luo, Shuzhuang Zhang [05],** Distributed (P2P) systems have been broadly connected in document sharing, streaming media, texting and different fields, which have pulled in substantial consideration. In the meantime P2P network traffic intensifies the blockage of a network fundamentally. So as to manage and control P2P traffic, it is essential to recognize P2P traffic precisely. In this paper we propose a novel P2P recognizable scheme plan, in view of the host and stream behavior qualities of P2P traffic. In the first place we decide whether a host takes part in a P2P application by coordinating its conduct with some predefined host level behavior rules. Along these lines, we refine the traffic identification by looking at the factual elements of each flow in the host with a few stream feature profiles. The experiments on genuine network information demonstrate that this strategy is very proficient to recognize P2P traffic.

**Marcell Perényi, Trang Dinh Dang, András Gefferth and Sándor Molnár [06],** recent estimation inspects report that a basic portion of Internet activity is dark. It is likely that the vast majority of the unidentified activity starts from conveyed (P2P) applications. Regardless, conventional frameworks to distinguish P2P movement seem to crash and burn since these applications generally cover their world by using self-decisive ports. In adition to the ID of genuine P2P movement, the attributes of that kind of movement are correspondingly hardly known.

The chief clarification behind this paper is twofold. Regardless, the paper proposes novel unmistakable confirmation framework to uncover P2P action from development storing up. This framework does not depend on upon package payload so we stay away from the burdens ascending out of authentic, security related, budgetary and specific tangles. Or, then again perhaps, our strategy depends endless supply of heuristics gotten from the solid properties of P2P movement. We demonstrate our strategy with current development information gotten from one of the best Internet suppliers in Hungary. We correspondingly demonstrate the high accuracy of the proposed estimation by methods for endorsement consider. Second, two or three consequences of an expansive development examination study are spoken to in the paper. We display the well-ordered lead of P2P clients differentiated and the non-P2P clients. We demonstrate our essential finding about the in every practical sense solid degree of the P2P and aggregate number of clients. Stream sizes and holding times are in like manner separated and eventual outcomes of a stunning tail examination are portrayed. At last, we investigate the unmistakable quality dispersing properties of P2P applications. Our outcomes show that the special properties of P2P application development appear to cloud away in the midst of aggregate and characteristics of the action will resemble that of other non-P2P action amassing.

**Dedinski, H. De Meer, L. Han, L. Mathy, D. P. Pezaros, J. S. Sventek, Z. Xiaoying [07],** P2P applications seem to rise as extreme executioner applications because of their capacity to build exceptionally unique overlay topologies with quickly changing and eccentric development stream, which can constitute a certifiable test for fundamentally over-provisioned IP frameworks. Along these lines, ISPs are confronting new, phenomenal framework organization issues that are not ensured to be tended to by statically passed on orchestrate building instruments. As a fundamental walk to these issues, this paper proposes a P2P estimation, recognizing verification and change configuration, wanted to acclimate to the dynamicity and remarkableness of existing, plainly fathomed and future, obscure P2P structures. The motivation driving this is configuration is to give the ISPs a persuading and versatile way to deal with oversee control and upgrade the action proposed by P2P applications in their frameworks. This can be master through a blend of various application and framework level programmable strategies, affecting a cross-layer ID and streamlining process. These frameworks can be related utilizing Active Networking stages, which can rapidly and enough send building parts on request. This flexibility of the redesign configuration is vital to address the energetic change of new P2P traditions and the arrangement of known traditions.

**Bela Hullar, Sandor Laki, Andras Gyorgy [08],** To manage and monitor the networks appropriately, network operators are regularly inspired by identifying the applications creating the traffic going through their networks, and doing it as quick (i.e., from a couple of packets) as could be expected under the circumstances. State-of-the-art packet-based traffic classification strategies are either based on the expensive inspection of the payload of a few packets of each flow or on essential flow measurements that don't consider the packet content. In this paper we consider the middle of analyzing only the first couple of bytes of the first (or first couple of) packets of each flow. This paper proposed programmed, machine-learning-based strategies accomplishing very great early classification execution on genuine traffic traces created from an assorted set of applications (including a few adaptations of P2P TV and file sharing), while at the same time requiring just constrained computational and memory assets.

## III. CONCLUSION

The conclusions introduced in this paper of network identification gives essential advantages to IP network building, organization and control and other key spaces. A large number of the acclaimed methods, for example, port based and payload based, have given a couple disservices. This paper comes with the technique of machine learning which a potential one is. The traffic is instructed by the payload autonomous truthful characters. The paper displays the assortment of levels in network traffic investigation and the critical data in machine learning space, perception of the issues of port-based and payload-based methodologies in traffic classification. Considering the need of the machine learning-based procedure, we have attempted distinctive things with K-means, SVM and GA to evaluate the productivity and execution. The trial happens on traffic datasets pass on that the precision gained by our approach is moved forward.

Subsequently, the execution of P2P network traffic is enhanced in proficient way and with more accurate outcomes.

## REFRENCES

[1] Jie Cao, Zhiyi Fang, Dan Zhang, and Guannan Qu, **"Network Traffic Classification Using Feature Selection and Parameter Optimization"**, Journal of Communications Vol. 10, No. 10, October 2015, doi:10.12720/jcm.v.n.p-p doi:10.12720/jcm.10.10.828-835.

[2] Prof S. R. Patil, Suraj Sanjay Dangat, **"Identifying Peer-to-Peer Traffic Based on Traffic Characteristics",** Recent Advances in Computer Science, ISBN: 978-1-61804-320-7.

[3] Satoshi Ohzahata, Yoichi Hagiwara, Matsuaki Terada, and Konosuke Kawashima, **"A Traffic Identification Method and Evaluations for a Pure P2P Application"**.

[4] Joseph Stephen Bassi, Loo Hui Ru, Khammas, Muhammad, Nadzir Marsono, **"Online Peer-To-Peer Traffic Identification Based On Complex Events Processing Of Traffic Event Signatures"**, Jurnal Teknologi (Sciences & Engineering) 78:7 (2016) 9–16, eISSN 2180–3722.

[5] Jinghua Yan, Zhigang Wu, Hao Luo, Shuzhuang Zhang, **"P2P Traffic Identification Based on Host and Flow Behaviour Characteristics"**, CYBERNETICS AND INFORMATION TECHNOLOGIES • Volume 13, No 3, Sofia • 2013, ISSN: 1314-4081, DOI: 10.2478/cait-2013-0026.

[6] Marcell Perényi, Trang Dinh Dang, András Gefferth and Sándor Molnár, **"Identification and Analysis of Peer-to-Peer Traffic"**, Proceedings of 5th International IFIP-TC6 Networking Conference, Coimbra, Portugal, May, 2006. © 2006 IFIP.

[7] I. Dedinski, H. De Meer, L. Han, L. Mathy, D. P. Pezaros, J. S. Sventek, Z. Xiaoying, **"Cross-Layer Peer-to-Peer Traffic Identification and Optimization Based on Active Networking".**

[8] Bela Hullar, Sandor Laki, Andras Gyorgy, "**Early Identification of Peer-To-Peer Traffic"**, IEEE Communications Society, 978-1-61284-233-2/11/$26.00 ©2011 IEEE.[09]. Baji Shaheed Baba Shaik, G Chinna Babu, Uppe Nanaji, **"Implementation of P2P reputation management scheme based on distributed identities and decentralized recommendation chains",** International Journal of Engineering Research and Applications (IJERA), ISSN: 2248-9622, Vol. 1, Issue 4, pp.1803-1810.