A Realistic Comprehensive Study on Various Classification Models in Data Mining

Dr.K.Rajasekhararao[1], M.Jayaram[2]

[1]*Professor of Computer Science and Engineering, Director, Usha Rama College of Engineering and Technology, Telaprolu, A.P.India*

[2]*Research Scholar, Department of Computer Science, Rayalaseema University, Kurnool, A.P, India*

Abstract: In the Internet's beginning, just a couple of uses were predominant. These included electronic mail and basic record exchanges. In any case, as the Internet has kept on advancing, the quantity of uses pervasive has become considerably. The recognizable proof of these applications and application-layer conventions conveying and planning them is the significant assignment. According to the type of data it's mandatory to classify the packet. Once classified packets must be scheduled for further processes. There are many algorithms playing vital role in classification and scheduling of packets. In this paper i would like discuss some of the core algorithms used for both classifications as well as scheduling.

Keywords: Network Classifications, TCP/IP,UDP

I.INTRODUCTION

A PC system or information system is an information transfers system which permits PCs to trade information. In PC systems, organized registering gadgets trade information with one another along system joins (information associations). The associations between hubs are set up utilizing either link media or remote media. The best-known PC system is the Internet.

System PC gadgets that begin, course and end the information are called system nodes.[1] Nodes can incorporate has, for example, PCs, telephones, servers and additionally organizing equipment. Two such gadgets can be said to be organized together when one gadget can trade data with the other gadget, regardless of whether they have an immediate association with one another.

PC systems vary in the transmission media used to convey their signs, the interchanges conventions to sort out system movement, the system's size, topology and hierarchical plan. By and large, interchanges conventions are layered on (i.e. work utilizing) other more particular or more broad correspondences conventions, aside from the physical layer that specifically manages the transmission media.

PC systems bolster a tremendous number of utilizations, for example, access to the World Wide Web, video, computerized sound, shared utilization of use and capacity servers, printers, and fax machines, and utilization of email and texting applications and in addition numerous others.

II. PORT BASED CLASSIFCATION

In the past, traffic classification techniques used well-known port numbers to identify the packets communicated on the Internet. This type of detection is the oldest methods which ease the analysis of data. This was easy and provided good results because many traditional applications used fixed port numbers assigned by or registered with the Internet Assigned Numbers Authority (IANA). After the birth of the Internet most of the applications used only one default port number. They communicated with the server using only those port numbers. Jacobson (1998) detected, TCP, UDP packet headers and analyzed them by comparing port numbers with the official list of default port numbers assigned by IANA. For example, sending and receiving Email we use the Simple Mail Transfer Protocol (SMTP) on port 25 to send email and the Post Office Protocol version 3 (POP3) on port 110.

In recent days Port-based classification is ineffective because the latest applications do not communicate with standardized ports allocated to them.

The recent versions of P2P applications, intentionally try to conceal their traffic, by using ephemeral ports or by using the port numbers of well-known applications to make the traffic indistinguishable to port-based classification and filtering. They compared port-based classification with a classification technique that relies on a set of transport layer heuristics.

Their trace only had SYN, FIN, and RST packets due to the longitudinal nature of their trace, and thus, validation of their classification results using payload-based techniques for example was not feasible. They found that 30% to 70% of the traffic is classified as unknown with port-based analysis. In addition, they found that the amount of unknown traffic was typically from 10% to 30% in the September 2003 to April 2004 portion of their trace.

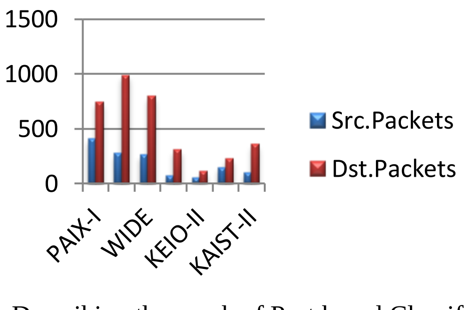


Fig. 1: Describing the graph of Port based Classification

Set	Duration	Link Type	Src.Packet	Dst.Packet	Packets Dropped	Bytes	Avg.Util	Avg.Flows	Payload
PAIX-1	2h	Backbone	410 K	745 K	250 M	91 G	104 Mb/S	1055 K	16 Bytes
PAIX-II	2h 2m	Backbone	275 K	984 K	1529 M	891 G	997 Mb/S	4651 K	16 Bytes
WIDE	55m	Backbone	263 K	794 K	32 M	14 G	35 Mb/S	312 K	40 Bytes
Keio-I	30m	Edge	73 K	310 K	27 M	16 G	75 Mb/S	158 K	40 Bytes
Keio-II	30m	Edge	54 K	110 K	25 M	16 G	75 Mb/S	91 K	40 Bytes
KAIST-I	48h 12m	Edge	148 K	227 K	711 M	506 G	24 Mb/S	19 K	40 Bytes
KAIST-II	21h 16m	Edge	86 K	101 K	357 M	259 G	28 Mb/S	21 K	40 Bytes

Table 1: Classification of Packets Rate in Port Based Classification

It has since increased from 30% to 70% by the spring of 2005. They provide strong circumstantial evidence that this increase in unknown traffic is highly correlated to the increase in P2P traffic found with their transport-layer heuristic. The difference between port based and payload based protocol detection is that the first method analyzed only the packet headers whereas the other method examines the whole content of the packet. Because the payload area contains more information than that of a port number.

III. PAYLOAD –BASED CLASSIFICATION

Similar Approach for Internet traffic classification that avoids port based identification is an analysis of packet payloads and also commonly referred as "Deep Packet Inspection". Here, the packet payloads are analyzed to see whether or not they contain characteristic signatures of known applications. This approach works very well for the Internet traffic that includes P2P traffic also. However, these techniques also have drawbacks. First, they require increased processing and storage capacity of the machines. Secondly, they are unable to detect encrypted transmissions. Finally, they can only identify traffic for which signatures are available, and are unable to classify previously unknown traffic. The payload-based approach has been well analyzed and the matter presented here indicates the latest classifying process available for traffic classification packets.

One example of a study integrating payload-based analysis into a classification approach is a content-based methodology to classify network traffic. The first step of their classification methodology uses IANA assigned port numbers to create an initial classification. Then, using an iterative procedure, they use increasingly more information at later steps. This approach allows the traffic to, be classified with increased confidence. The last step concludes the process by relying on manual analysis of the traffic for any remaining unclassified traffic.

To measure the effectiveness of port-based classification the trace of the traffic generated from approximately 1,000 users. This comparison found that approximately 30% of the bytes in the trace are either misclassified or unclassified when using just the IANA port assignments. However, with the content-based approach 99.9% of the traffic was identified confidently.

The difference between port based and payload based protocol detection is that the first method analyzed only the packet headers whereas the other method examines the whole content of the packet. Because the payload area contains more information than that of a port number. This makes the classification more accurate. Port numbers need not be considered in this case.

As well, some research conducted to address the aforementioned concerns such as the automatic detection of signatures and decreasing the processing requirements of deep packet inspection will also be outlined.

Set	Timeout	Src.Pkt	Dst.Pkt	Pkt Dropped
Edonkey	1min	1479	359	1120
Kazaa	2min	1479	792	687
Gnutella	4min	1479	770	709
HTTP	5min	1479	756	723
FTP	6min	1479	696	783
RTP	8min	1479	676	803
SIP+RTP	10min	1479	674	805

Table 2: Classification of Packets Rate based on Payload Based Classification

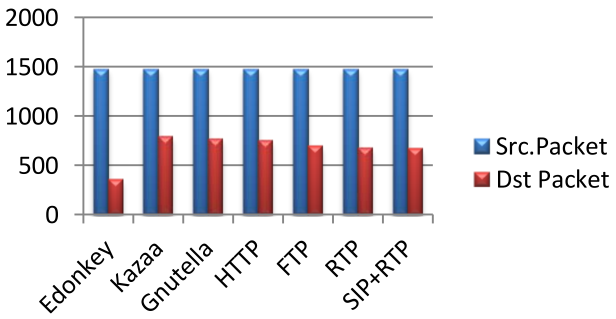


Fig. 2: Describing the Graph of Payload Based Classification

A. Identifying Payload Signatures:

An approach to accurately identify P2P applications is based on application-level payload signatures. The focus of their research is to identify signatures that are highly accurate, that are scalable for analysis of large volumes of traffic, and that is robust to variable network dynamics such as packet loss, asymmetric routing, and packets arriving out of order. Their work focused on the five most predominant P2P applications: Gnutella, eDonkey, Direct Connect, BitTorrent, and KaZaA.

It's found that signatures with a fixed-offset are trivial to implement and have a low computational overhead; while, variable-offset signatures are much more computationally expensive1. The method is validated on two full packet traces both collected in November 2003 that contain 120 Gigabytes and 1.8 Terabytes of data, respectively. They found that by examining a few packets in each flow over 99% of the P2P traffic could be identified. The authors also analyzed port-based identification and found that 30% to 70% of the traffic for KaZaA and Gnutella use non-standard port numbers, whereas only 1% to 4% of the traffic for BitTorrent and eDonkey use a non-standard port. The payload classification should create a pattern database and it is important to update the database periodically to match with the invention of new protocols. The problem with this analysis is analyzing the packet one by one. It requires a lot of computational power and consumes more time when compared with other methods.

Voice over IP (VoIP) protocols prefers UDP rather than TCP because it has no error correction methods. UDP does not check the packet for error and resending is not a part of UDP. If the datagram carrying VoIP is lost, then there is no need to resend the lost packet because the voice flow is continuous and cannot wait for a longer period.

Some algorithms use the payload analysis to provide "base truth" to compare new behavioral-based traffic classification methods that they propose. In earlier the network traffic dynamics of Internet Chat Systems the authors focus on IRC and web-based chat systems. Their paper describes a port and payload-based methodology for identifying the chat flows and filtering out non-chat traffic. Their approach uses well-known port numbers to filter out traffic that is most likely non-chat such as Gnutella traffic on port 6346. After this filtering has taken place they use payload signatures to separate the web-based chat flows from the regular non-chat traffic.

B. Automated detection of payload signatures

One of the concerns of payload-based analysis of network traffic is the identification of characteristic signatures for use in deep packet inspection. This problem addressed by attempting to automatically learn the application signatures using three machine learning algorithms. The algorithms studied include Naive Bayes, AdaBoost, and Regularized Maximum Entropy. The approach uses a binary feature vector to train the algorithms, which is obtained from the first n-bytes of a flow's payload. The flow's payload is encoded into binary vectors so that for each of the n bytes of payload, the binary vector has 256 elements corresponding to this byte. Each of these elements is initialized to 0 first and then the element whose number corresponds to the value contained in this byte is set to 1. Many algorithms are tested upon FTP, SMTP, POP3, IMAP2 In supervised machine learning training data is used to learn a function that can be used to predict the class labels of test data.

Haffner et al (2005) relied on training the classifiers with each specific application, it wanted the classifiers to identify. Recently, Ma et al (2006) has extended this work by proposing an unsupervised approach to the detection of application signatures. This allows similar flows (most likely from the same protocol) to be grouped together. These groups (clusters) are then labelled in a later step to create a classification of the current and future flows placed into that group. The authors achieve this by using a generic classification framework and compare the use of three different methods: product distributions of byte offsets, Markov models of byte transitions, and common substring graphs. The authors evaluate methods to determine if flows from the same protocols are grouped together, and that a new protocol is placed in a separate group when it is introduced. The misclassification rate varied between 2% to 10% with their various methods.

C. Speeding up Deep Packet Inspection

It's found that payload analysis is much more computationally expensive when the payload signatures use a variable-length offset instead of a signature based on a fixed-length offset. It was addressed by proposing algorithms to increase the speed of deep packet inspection of regular expressions The authors propose a new method of representing regular expressions that condenses the transition state space and reduces the previously large memory requirements for regular expression matching. The method is evaluated using regular expressions obtained from several popular Intrusion Detection Systems such as Snort and Bro. The evaluations show that, with a careful implementation, regular expression matching of full-packet payloads can be successfully achieved at Gbps link speeds.

IV. BEHAVIOURAL BASED CLASSIFICATION

Karagiannis et al (2004) and Karagiannis et al (2005) classify P2P traffic and report on trends in the usage of P2P file sharing. The authors analyze data from a tier 1 ISP; however, they are limited by having only 16 bytes of payload data available and only 4 bytes in some of their older traces. This would limit the effectiveness of an analysis and evaluation using only payload-based classification. Instead, the authors develop a non-payload based method, specifically two transport layer heuristics to classify P2P.

One of the heuristics looks for IP addresses that are concurrently using both TCP and UDP. This heuristic works on the basis that most P2P applications typically send control information by UDP and transfer data by TCP. Flows using port numbers of well-known UDP applications such as DNS on port

53 are excluded to reduce false positives. The second heuristic looks at the ratio of the number of unique IP addresses with unique port numbers to which a host is connected. If this ratio is roughly equal, then the flows from this host are classified as P2P. A higher ratio would tend to indicate a non-P2P type of flow, such as HTTP because multiple concurrent flows are generally spawned from a web server to decrease the response times when a web page with multiple objects is requested.

Karagiannis et al (2004) validate these heuristics by creating a "base truth" using well-known port numbers of P2P applications, payload signatures, and a heuristic where if an IP address and port number pair had previously been used in a P2P flow in the last five minutes then future unlabeled IP/port pairs would also be classified as P2P. The transport layer heuristics were shown to be able to identify 90% of the total P2P bytes and 99% of the P2P flows. In addition, the transport layer heuristics were able to identify P2P traffic that was previously unidentified with the payload analysis method used to establish the "base truth".

Recently a classification approach based on the analysis of communication patterns of hosts, leverages information obtained from the social, the function, and the application layers to identify the application classes of particular flows from a host. The social level information is information such as the popularity of a host and the communities with which the host communicates. The functional level attempts to determine if the host's communication paradigm is a client / server or collaborative (e.g., P2P). The application layer uses the communication patterns of application protocols referred to by the authors as "graphlets" to identify the applications. Constantinou & Mavrommatis (2006) proposes a similar technique that looks at the connection graph of hosts.

A methodology introduced based on data mining and information-theoretic techniques, to discover functional and application behavioral patterns of hosts and the services used by the hosts. They subsequently use these patterns to build general traffic profiles, for example, "servers or services", "heavy hitter hosts", and "scans or exploits".

Set	Src.Pkt	Dst.Pkt	Packets Dropped
P2P	1500	689	811
Non P2P	1500	751	749

Table 3: Classification of Packets rate based on Behavioral Based Classification

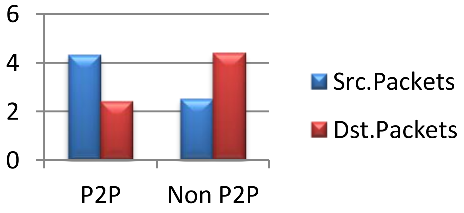


Fig. 3: Describing the Graph of Behavior Based Classification

V. MACHINE LEARNING BASED APPROACHES

Another promising approach to traffic classification is the use of machine learning. This approach relies on the premise that a set of features for objects would be similar when objects are of the same class. In general, a feature can be any attribute that is relevant to the prediction of the target set of classes. In the case of traffic classification, the objects dealt with are flows and the classes are the different applications or traffic types the flow is attempted to be classified as.

Generally, in machine learning there are two stages when developing a classifier. The first stage "learns" a mapping between the objects and the desired classes. This mapping is done using a labelled training data set. Subsequently, in the second stage this learned mapping is used by the classifier to label new objects. A new classification is proposed in which they suggested that sub-flows of 25 packets has increased timeliness, Precision, and Recall for both ET and VoIP traffic. They have proposed an optimal size for a classification window which balances the tradeoff between the classifier's Precision and Recall, classification timeliness, classification speed, and processing overhead. Their work results in 99% of precision and 95% of re-class with in 1s. This approach used Naive Bayes and C4.5 machine learning algorithms.

A rapid packet classification mechanism realized by HaRP able to not only exhibit high scalability in terms of both the classification time and the SRAM size involved, but also effectively handle incremental updates to the filter data sets. Based on a single set-associative LuHa hash table (obtained by lumping a set of hash table units together) to support two-staged search, HaRP promises to enjoy better classification performance than its known software oriented counterpart, because the LuHa table narrows the search scope effectively based on the source and the destination IP addresses of an arrival packet during the first stage, leading to fast search in the second stage. With its required SRAM size lowered considerably, HaRP makes it possible to hold entire search data structures in the local cache of each core within a contemporary processor, further elevating its classification performance. Evaluation results have shown that HaRP_ with the set associative degree of 4, generally experiences very rare set overflow instances.

A algorithm which combines tree based algorithm and a decision tree based algorithm. It basically builds a tree and conditionally constructs a decision tree if the number of rules included in a tree node is greater than a threshold value. Hence, the number of rule comparisons is reduced, and the number of rule replications is controlled by a limited set of rules. TSS categorizes rules according to their prefix length combinations; therefore, the search procedure of packet classification involves producing all matching entries from a set of hash tables.

VI. ANALYSIS AND SELECTION OF FEATURE SET

Obtaining a set of relevant features is a difficult problem in machine learning, the focus of much of the prior work using machine learning techniques has been on demonstrating the ability of algorithms to group together flows according to application type and not on classifying traffic. These techniques generally use only features obtained from a single flow, such as packet sizes, inter arrival times, or aggregate statistics. These approaches do not consider the application labels of the flows when forming the groups. An abstract model represents the amount of data travel from both sender. The idle time taken between the message exchanges. The feature vectors for a flow are extracted from these times. Then the hierarchical clustering used to group the flows based on similarity. Around 5,000 flows were clustered that many of the clusters corresponded roughly to a single application. For example, one of their clusters contained web flows and another contained flows from mail protocols. Traffic classes were categorized into four pre-determined traffic classes (inter-active, bulk data transfer, streaming, and transactional) using the Nearest Neighbor and the Linear Discriminate Analysis classification techniques. It was proved that in 2004 that it is possible to successfully separate the flows of different traffic classes using only flow statistics and give explanations to why their chosen flow statistics (aver- age packet size, and flow duration) would work for the different traffic classes.

It was analyzed that the packet sizes and inter arrival times of different application types to determine whether different applications exhibit different packet size and inter arrival characteristics. In analyzing plots of packet sizes and inter arrival times, they found that while there were some distinguishing characteristics between applications, it would be difficult to do rich traffic classification. A new method was proposed to use Expectation Maximization (EM) clustering that will group flows using flow statistics including byte counts, connection durations, and packet size statistics. A preliminary analysis was conducted using cluster visualization to examine the clusters and find that many of the clusters correspond to a single type of traffic class such as bulk data transfers and DNS traffic.

Internet backbone traffic is classified on transport layer according to network applications. Classification is done by a set of heuristics inspired by two previous articles and refined in order to better reflect a rough and highly aggregated backbone environment. Obvious misclassified flows by the existing two approaches are revealed and updated heuristics are presented, excluding the revealed false positives, but including missed P2P streams. The proposed set of heuristics is intended to provide researchers and network operators with a relatively simple and fast method to get insight into the type of data carried by their links.

A new mechanism was proposed for securing network boundaries. The mechanism, called Tunnel Hunter, relies on the statistical characterization at the IP-layer of the traffic that is allowed by a given security policy, such as HTTP or SSH. The statistical profiles of the allowed usages of those protocols can then be dynamically checked against traffic flows crossing the network boundaries, identifying with great accuracy when a flow is being used to tunnel another protocol. Real-time traffic classification is a fundamental task for many network management decisions: by timely identifying the applications that generate traffic on a specific network link, network managers can optimize the utilization of their networks; better Quality-of-Service (QoS) can be offered to connected clients while preventing the saturation of many network resources. In addition, the timely identification of malicious traffic, or of traffic, presenting anomalous patterns, can be also achieved for assuring the protection of the connected hosts and network resources. However, achieving such ability is not an easy task.

The inherent complexity of current Internet applications and services together with the existence of several privacy and legal restrictions prevent the analysis of the contents of the packets, thus preventing an accurate and timely traffic classification. This issue is addressed by analyzing captured Internet traffic over several classification windows, until an accurate identification decision is achieved. A new traffic classification proposed that the dark mechanism based on matching several empirical distributions representing computer applications with the one of the target traffic. The classifier combines two methods for performing such matching in real-time and on a packet-by-packet manner: one based on the Kolmogorov-Smirnov test, and another one based on the Chi-Squared test.

VII. CLASS DEFINITION

The Internet Traffic is divided into 4 broad application classes commonly found in the Internet world

Interactive: The applications which interact with the other remote systems generate Interactive traffic. E.g. of this class are remote login sessions, an interactive web interface, Telnet, real time gaming applications.

Bulk Data Transfer: This class contains traffic, which transfers large data over the network without any real time constraints. This type of traffic is generated by applications like FTP, updating software's, audio and video downloads.

Streaming: The streaming class contains streaming videos or audios. It includes multimedia traffic flows with real time constraints.

Transactional: This type of traffic is used in a small number of request, response pairs which are paired together to identify a transaction. Examples of transactional applications are DNS, Oracle transactions etc.

The choices were motivated because of the need to select a small number of classes that would be simple, intuitive, and still adequately represent the different QoS requirements of commonly used applications.

Offering service guarantees to existing and emerging applications in the internet have been a big challenge to internet designers. One of the most important mechanisms to provide a service guarantee is scheduling. Scheduling determines the order in which the packets from different flows are served. Packet scheduling in routers has been an active area of research in the last two decades, and it is necessary to investigate it, to find an alternative scheduling algorithm for today's internet needs.

In routers, the essential component of a queue manager is a scheduler which employs a scheduling mechanism to decide the packet to be served next. Scheduling disciplines are responsible for protecting one user's traffic from that of another because bandwidth hog may occur between the users. If the scheduling mechanism does not select the correct packet of a flow, then it will affect the performance of other traffic flows.

A scheduler should require as few simple operations as possible to make a scheduling decision to select the next packet or the service. In particular, the number of operations should be as independent of the number of flows that are to be scheduled as possible. Thus, if n is the total number of queues or traffic flows to be scheduled by a scheduler, then a scheduler that has O(1) time complexity is preferred in comparison to the one that has O(n) time complexity. This is a desired property for high-speed networks, and in routers where the number of flows can be in the thousands as in the internet core.

The priority of packets and expiry times are used by the transport layer to reorder or discard packets to optimize the use of the network. This can be used for video conferencing to prioritize important data. This algorithm is implemented as an interface to the Datagram Congestion control protocol and it gives better improvements to video conferencing using the standard UDP and TCP.

VIII. CONCLUSION

This paper gives the realistic comprehensive study on Port-Based Classification, Payload Based Classification, Behavior Based Classification. Through this complete study we can estimate efficiency of the various natures of the classification. The conspicuous verification of these applications and application-layer traditions passing on and arranging them is the noteworthy task. By sort of information it's obligatory to characterize the bundle. Once grouped parcels must be planned for further procedures. There are numerous calculations assuming crucial part in characterization and planning of bundles. In this paper it is discussed about a percentage of the center calculations utilized for both arrangements and also planning.

IX. REFERENCES

[1]. Jacobson, V 1988, 'Congestion Avoidance and Control', in Proceedings of ACM Symposium on Communications Architectures and Protocols, pp. 314-329.

[2]. Jacobson, V, Nichols, K & Poduri, K 1999, 'An Expedited Forwarding PHB', Internet Engineering Task Force (IETF), RFC 2598.

[3]. Sen, S, Spatscheck, O & Wang, D 2004, 'Accurate, Scalable In Network Identification of P2P Traffic using Application Signatures', in Proceedings of World Wide Web, pp. 512-521. 117

[4]. Roughan, M, Sen, S, Spatscheck, O & Duffield, N 2004, 'Class-of Service Mapping for QoS: A Statistical Signature-based Approach to IP Traffic Classification', Proceedings of the ACM SIGCOMM Conference on the Internet Measurement, pp. 135-14.

[5]. Madhukar. A and Williamson, C 2006, A Longitudinal Study of P2P Traffic Classification, proceedings of MASCOTS, pp. 179-188.

[6]. Haffner, P, Sen, S, Spatscheck, O & Wang, D 2005, 'ACAS: Automated Construction of Application Signatures', proceedings of SIGCOMM'05, MineNEt Workshop, pp. 197-202..

[7]. Moore, A & Zuev, D 2005, 'Internet Traffic Classification Using Bayesian Analysis Techniques', proceedings of SIGMETRICS'05, pp. 50-60.

[8]. Karagiannis, T, Broido, A & Brownlee, N 2004, 'Is P2P Dying or Just Hiding?', proceedings of GLOBECOM'04, vol. 3, pp. 1532-1538.

[9]. Karagiannis, T, Broido, A, Faloutsos, M & Claffy, KC 2005, 'Transport Layer Identification of P2P Traffic', proceedings of IMC, pp. 121-134.

[10]. Karagiannis, T, Papagiannaki, K & Faloutsos 2005, 'BLINC: Multilevel Traffic Classification in the Dark', proceedings of SIGCOMM'05, pp. 229-240.

[11]. Dews, C, Wichmann, A & Feldmann, A 2003, 'An Analysis of Internet Chat Systems', proceedings of Internet Measurement, pp. 51-64.

[12]. Kumar, S, Dharmapurikar, S, Yu, F, Crowley, P & Turner, J 2006, 'Algorithms to Accelerate Mulitple Regular Expressions Matching for Deep Packet Inspection', proceedings of SIGCOMM'06, pp. 339-350.

[13]. Xu, K, Zhang, Z & Bhattacharya, S 2005, 'Profiling Internet Backbone Traffic: Behavior Models and Applications', in Proceedings of SIGCOMM'05, pp. 169-180.

[14]. Thuy, T, T, Ngayen, Philip, B, & Sebastian, Z 2012, 'Timely and Continuous machine learning based class for Interactive IP Traffic', IEEE/ACM Transactions on Networking, vol. 20, no.6, pp. 1880-1894.

[15]. Fong, P & Niang-Feng, T 2011, "HARP: Rapid packet class via Hashing round down Prefixes", IEEE Transactions on Parallel and Distributed Systems, vol.22, no. 7, pp. 1105-1119.

[16]. Hyesook, L, Youngju, C, Miranshim,& Jungwon L 2014, "A QuadTrie conditionally merged with a Decision Tree for packet class", IEEE Communication Letter, vol. 18, no.4, pp. 676-680.

[17]. Pi-Chung, W 2014, 'Scalable packet class for Datacenter Networks', IEEE Journal on Selected Areas in Communication, vol. 32, no.1, pp.124-138.

[18]. Hsiao-weihu, Yen-Liang, C, & Kweitang 2013, "A Novel Decision Tree Method for structured continuous-label class", IEEE Transactions on Cybernetics, vol. 43, no. 6, pp.1734-1747.

[19]. Hernandez Campos, F, Smith, FD, Jeffay, K & Nobel, AB 2003, 'Statistical Clustering of Internet Communications Patterns', Journal of Computer Science and Statistics, vol. 35, no.2, pp. 111-116.

[20]. McGregor, A, Hall, M, Lorier, P & Brunskill, J 2004, 'Flow Clustering Using Machine Learning Techniques', in Proceedings of PAM'04, pp.205-214. 115

[21]. Roughan, M, Sen, S, Spatscheck, O & Duffield, N 2004, 'Class-ofService Mapping for QoS: A Statistical Signature-based Approach to IP Traffic Classification', Proceedings of the ACM SIGCOMM Conference on the Internet Measurement, pp. 135-14.

[22]. Zander, S, Nguyen, T & Armitrage, G 2005a, 'Automated Traffic Classification and Application Identification using Machine Learning', in Proceedings of LCN'05, pp. 250-257.

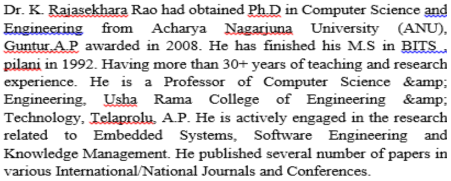
[23]. Zander, S, Nguyen, T & Armitrage, G 2005b, 'Self-Learning IP Traffic Classification Based on Statistical Flow Characteristics', in Proceedings of PAM'05, pp. 325-328.

[24]. Dusi, M, Crotti, M, Gringoli, F and Salgarelli, L 2009, 'Tunnel Hunter: Detecting application-layer tunnels with statistical fingerprinting', Elsevier Computer Networks, vol. 53, no. 1, pp. 81-97.

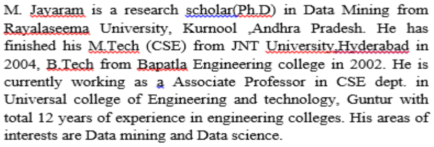
[25]. Neto, M, Gomes, JV, Freire, M & Inacio, P 2013, 'Real Time traffic class based on statistical tests for matching signature wit packet length distributions', in proceedings of LANMAN, 19th IEEE workshop, pp. 1-6.

[26]. Mathew Roughan, Subhabrata Sen, Oliver Spatscheck, Nich Duffield, 2004, 'Class-of-Service Mapping for QoS: A Statistical Signaturebased Approach to IP Traffic Classification', proceedings of the 4th ACM SIGCOMM Conference on Internet Measurment, ACM, New York, pp. 134-148.

AUTHOR PROFILE:



Dr. K. Rajasekhara Rao had obtained Ph.D in Computer Science and Engineering from Acharya Nagarjuna University (ANU), Guntur,A.P awarded in 2008. He has finished his M.S in BITS , pilani in 1992. Having more than 30+ years of teaching and research experience. He is a Professor of Computer Science &amp; Engineering, Usha Rama College of Engineering &amp; Technology, Telaprolu, A.P. He is actively engaged in the research related to Embedded Systems, Software Engineering and Knowledge Management. He published several number of papers in various International/National Journals and Conferences.



M. Jayaram is a research scholar(Ph.D) in Data Mining from Rayalaseema University, Kurnool ,Andhra Pradesh. He has finished his M.Tech (CSE) from JNT University,Hyderabad in 2004, B.Tech from Bapatla Engineering college in 2002. He is currently working as a Associate Professor in CSE dept. in Universal college of Engineering and technology, Guntur with total 12 years of experience in engineering colleges. His areas of interests are Data mining and Data science.