Most Wanted Internet Applications: a framework for P2P Identification

Guthemberg Silvestre
Univ. Pierre et Marie Curie
Paris, France
guthemberg.silvestre@lip6.fr

Stênio Fernandes
University of Ottawa
Ottawa, Canada
IF-AL (On leave)
Maceió, Brasil
stenio.fernandes@ieee.org

Carlos Kamienski
Univ. Federal do ABC
Santo André, SP, Brasil
cak@ufabc.edu.br

Djamel Sadok
Univ. Federal de Pernambuco
Recife, PE, Brasil
jamel@gprt.ufpe.br

Abstract – **For almost a decade, Peer-to-Peer (P2P) traffic have been putting pressure on network operators. Taking actions to control P2P traffic is a daunting task. In this paper, we present a comprehensive framework for the identification of P2P traffic based on information-theoretic techniques. Despite the inherent difficulty to single out such applications, our methodology is able to successfully identify P2P traffic using a set of communication patterns or profiles. We show that profiles built on the observation of traffic volume are more accurate than those using the number of flows.**

I. INTRODUCTION

P2P traffic has increased dramatically over the last few years. Currently the volume of bytes carried by P2P applications, such as file sharing, IPTV, etc., has been significantly impacting the Internet traffic profile [1]. As a result, many recent research studies have been investigating the development of efficient classification mechanisms capable of a better understanding of P2P traffic behavior and its characteristics [2][3][4]. Such volume increase has raised many concerns for traffic engineers, causing bandwidth misuse and clogging, copyright infringement problems and contributed to malware propagation. Hence many ISPs and corporate managers consider P2P file sharing traffic as the most wanted for identification purposes and want to be able to control it. However, identifying P2P traffic is by no means an easy task. Conventional identification approaches, such as port number identification and application level signature identification, have become deprecated, mainly because there are a large number of P2P applications good at hiding behind legitimate services, a constant release of new P2P protocols, the common use of cryptographic and tunneling procedures.

Prior works have used probabilistic approaches for traffic classification, making it possible to identify general profiles based on communication patterns between hosts and networking services [5]. The work in [6] presents a general traffic profiling methodology to automatically identify traffic profiles at backbone networks, providing a plausible explanation to canonical behaviors. Such methodology presented satisfactory results in identifying several unwanted traffic, including malware. However, the paper did not address the effectiveness of such profiling approach in detecting P2P traffic, which poses a considerable threat on ISP network management and operation. This work attempts to overcome this limitation by introducing a new methodology capable of dealing with the specific case of P2P traffic identification. This paper addresses the challenges of identifying P2P traffic, with the focus on file sharing applications. Our main contributions are threefold: i) we evaluate the existent limitation of the methodology proposed in [6] when it comes to dealing with P2P traffic, ii) we specify profiles to ensure the accuracy and completeness of the P2P traffic identification process, and iii) we also define simple rule set and procedures to classify P2P traffic through communication patterns.

The remainder of the paper is organized as follows. In Section II we briefly present technical background, related work and we describe our evaluation framework and traces collected in a ISP backbone. We explain our experiments in Section III, and in Section IV we show the preliminary results obtained during the training stage. In Section V we present the final results of our evaluation. Finally, Section VI concludes this paper.

II. TECHNICAL BACKGROUND

A. Fundamentals

With the increased networking complexity, the workload to operate and manage network services has followed suit. In terms of traffic measurement, many traffic capture and aggregation approaches have been proposed, in particular those which are flow based for the sake of scalability. These solutions have been characterized by the aggregation of packet information (e.g., volume and duration of flow) established by a well-known key composed of the five-tuple: the source IP address (srcIP), destination IP address (dstIP), source

978-0-7695-4041-2/10 $26.00 © 2010 IEEE
DOI 10.1109/CNSR.2010.64

port (srcPrt), destination port (dstPrt) and transport layer protocol field.

A great deal of recent works has been investigating methods that classify application traffic based on flows (e.g., [5][6][9], where the most relevant studies come from [5] and [6]. In [5] a methodology that defines communication patterns for identifying applications using network and transport layer information was proposed. A more daring classification approach that identifies traffic profiles using information-theoretic techniques is presented [6]. The results from both research studies suggest that using flow-based techniques for traffic profiling can be highly effective in also identifying malicious traffic. In this work, we dwell on such experience to build a profiling strategy for P2P traffic.

B. *Evaluation framework*

In order to evaluate the identification of P2P traffic based on communication patterns, we have introduced some necessary changes to the traffic profiling strategy initially defined in [6] in addition to the development of a framework for it. The proposed evaluation framework consists of the following two successive steps: training and testing as shown in Figure 1. First, our framework processes a set of training dataset traffic to enable it to recognize and build the most representative P2P communication patterns. This is important to give the framework the ability to adapt to changes in P2P protocols and the emergence of new ones. In a second stage, the framework starts identifying P2P flows based on the communication patterns characterized during the first step.

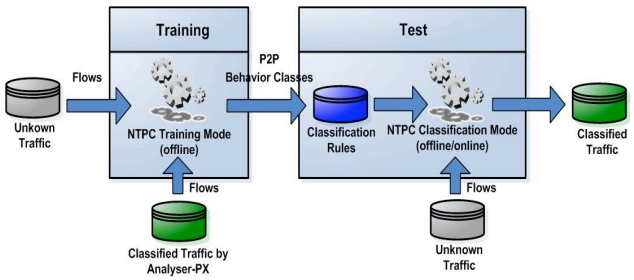

Fig. 1. Evaluation framework

For undertaking both steps of our framework, we developed the Network Traffic Profile Classifier (NTPC) tool. It implements the components necessary to operate properly during each step. In fact, in order to evaluate our proposed methodology, we have intentionally processed different datasets for each learning step of the framework.

C. *Datasets*

The datasets used in this study were collected from several sources, including broadband and academic networks, where we have been collecting packets for five years. We present only the results from one of the Points of Presence (PoPs) of the Brazilian Academic Network (RNP) due to lack of space and similarity of results to the broadband operator. The traces were captured on a high-speed link and stored as flow records, using as record keys the fields from the IP header: source and destination IP addresses, source and destination ports and protocol identification. P2P flows were identified through an accurate deep packet inspection (DPI) tool, which performs pattern matching to search for specific P2P applications signatures. Table I shows shortly the traces processed by the two steps that make up our framework. That traffic was collected for two consecutive weeks: the first one for training and the second one for testing.

Table I – Traces description

Stage	Total (TB)	P2P traffic (TB)	Number of Flows (Millions)	P2P % (for bytes)
Training	1.9	1.2	299	66,09
Test	2.1	1.3	281	65,19

D. *Evaluation metrics*

We based our evaluation procedures on two important metrics: completeness and accuracy. As described in [5], the values of these metrics measure how effective a classification or identification scheme is. Completeness measures the percentage of the traffic identified by our approach. That is to say, completeness is defined as a ratio of number of identified measurement unit by our traffic profile classification tool (NTPC) over the total number of measurement unit indicated by payload analysis (DPI). Our second metric, accuracy, measures the percentage of the classified traffic by NTPC that is correctly labeled. In more detail, accuracy measures the number of identified flows really belongs to a given class. In this paper, these metrics were computed for two measurement units: bytes and flows.

III. METHODOLOGY FOR P2P TRAFFIC IDENTIFICATION

Our methodology of P2P traffic identification relies on information-theoretic techniques to extract and classify flows. In particular, this approach places great emphasis on the entropy concept. Entropy can be seen as the measurement of information of a given dataset, which essentially quantifies "the amount of uncertainty" contained in that dataset [7]. In this paper, we only

present the traffic volume (in bytes) as the main metric to calculate the entropy, instead of number of flows. The reason for such a choice is related to the power-law characteristic of P2P flows, in which few flows carry large amounts of bytes [8]. Consequently, our concern is to focus on the P2P traffic volume as a better indicator to identify and ultimately prevent P2P traffic.

A. Behavioral classes identification

Our approach, NTPC, implements an identification methodology, which consist of four related basic stages:

1. *Preprocessing*: Packets are captured at the measurement network interface, aggregated and exported as flow records;
2. *Extracting significant clusters*: determines the clusters for four features or *dimensions*. This procedure aims to reduce and facilitate dataset behavior inspection through the identification of its most significant or principal elements.
3. *Clusters classification*: classifies each cluster's element into behavioral classes based on similarities and dissimilarities of communication patterns.
4. *Communication patterns interpretation*: defines a set of behavior classes capable of better describing given applications and services.

Figure 2 shows the relationship between these four stages. The overall process offers both automated and supervised adaptation of parameters.

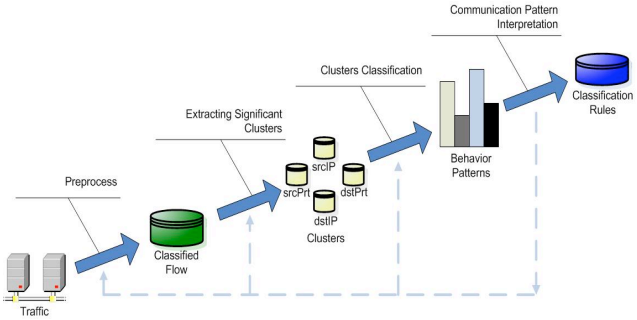

Fig. 2. Behavior class identification process

B. Significant clusters extraction algorithm

This sub-section begins with an explanation of the profiling methodology used in this paper, then we introduce the analysis of two key parameters for improving P2P identification.

The extraction of significant clusters deals with a four-dimensional feature space composed by the four attributes srcIP, dstIP, srcPrt and dstPrt. Considering these elements, we can identify two relevant types of network communication behavior. First, there is a relationship between IP addresses (srcIP and dstIP), one that determines the communication pattern between hosts. Second, there is also the behavior built from port/service (srcPrt and dstPrt) usage patterns.

The significant clusters extraction algorithm adopted by the NTCP tool is fed by the DPI flow information output and is responsible for processing each dimension separately. The extraction of significant clusters essentially relies on two parameters: amount of observations of a feature dimension (m), and uniformity degree (β).

The first key parameter is m. As described in [6], for each one of the four dimensional attributes, we calculate during a given time interval T, the parameter m representing the amount of flows with a given attribute value. For example, considering the feature space srcIP, for each source IP address, m stores the amount of flows whose components have this same source IP. According to the value of m for each IP address, NTPC computes the likelihood P to have m related to (divided by) the total amount of flows on a given time interval T. Then the NTPC tool uses the obtained likelihood values from each element in order to compute the cluster's relative uncertainty (RU) information. In this paper, the RU value provides an index of variety or uniformity regardless of the supported sample size. In other words, the RU of a set of data is obtained by the division of the entropy of this set by the maximum entropy observed on a sample. In particular cases, a null RU (RU=0) indicates that all flows have the same srcIP, otherwise RU=1 expresses that each flow set has a unique IP source address srcIP. That methodology was efficient for malicious traffic [6] where the number of flows (m) is essential information. However the number of bytes is more meaningful for P2P file share applications than number of flows.

The analysis of parameter m aims at evaluating the maximum completeness from significant clusters. We evaluated two different implementations of the algorithm responsible for clusters extraction with different possible values for m. Both of them took into account the completeness metric. In the first one (I1), as originally proposed in [6], the extraction algorithm considered m as a variable that stores the number of flows. In the second implementation (I2) however, our proposed extraction algorithm considers m as a parameter to store the amount of bytes. In addition, we also evaluated the union and intersection operations between these two strategies. Consequently, we developed a total of four approaches to evaluate the most significant way of extracting significant clusters for P2P traffic: I1, I2, $I1 \cap I2$ and $I1 \cup I2$. Figure 3 presents the results we obtained by applying our four approaches using the test traces. This figure shows the results for the completeness metric.

Therefore, they present the maximum percentage of bytes of the overall significant clusters. Figure 3 suggests that the approaches based on I2 examining the number of bytes and consequently also I1∪I2 for both numbers of flows and bytes reached the highest completeness rates. For instance, on January the 10th, the maximum completeness values achieved by the approaches I1, I2, I1∩I2 and I1∪I2 were 59.9%, 98.6%, 59.8% and 98.6% respectively. Otherwise, we can also observe that the values of the intersection and union implementations are almost the same of those from implementations I1 and I2. These results mean that practically all extracted flows from I1 are included in the set of I2 extracted flows. Hence the I2 strategy is more efficient to extract significant flow information.

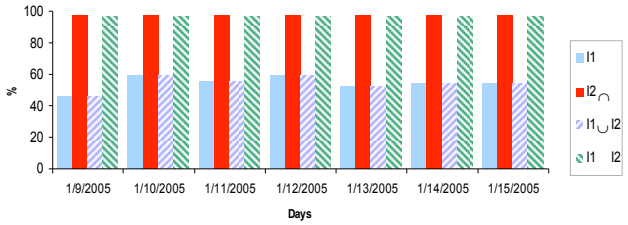

Fig. 3. Maximum Completeness of I1 and I2

The second key parameter β controls the amount of uniformity of a given cluster. In general terms, it means that the number of selected elements of a dimension is proportional to β. Diminishing the number of elements of a group is essential for analyzing huge amounts of traffic. But it also decreases the completeness of our profiling method. In order to find a better value for β than that proposed in [6], taking into account this tradeoff, different values of β were evaluated. The results are shown in Table II.

β	Average group size	Completeness
0.95	3066	98.03%
0.85	178	96.42%
0.75	**66**	**95.16%**
0.65	28	93.40%
0.55	20	91.28%
0.45	15	89.27%

Table II - β parameter evaluation

The Table II presents the average group size and computed completeness per β value, where β varies from 0.95 (as proposed in [6]) to 0.45. Considering those results, 0.75 was empirically chosen for presenting a significant decrease of the group size (66) and an acceptable maximum value of completeness for our experiments.

We can conclude that in entropy-based P2P traffic classification the traffic volume (amount of bytes) is a parameter more relevant than the number of flows. In fact, among the P2P flows identified by the DPI, most of them carry huge amounts of bytes. And also that a simple verification of β allows us to evaluate our data in a more efficient way. Thus, the following results in this paper, will only consider flow classification based on the byte quantity information metric and β equal to 0.75. We emphasize that the previous methodology from [6], is not sufficient on its own to identify P2P traffic.

IV. SIGNIFICANT CLUSTERS EVALUATION

In this section, we show an information-theoretic approach for characterizing the behavior of the significant clusters extracted using the algorithm described in the previous section. We also show a natural behavior classification scheme which establishes a suitable way to put together clusters in distinct behavior classes (BC). Afterwards, we introduce a supplementary refinement approach in order to evaluate and select the best BCs which represents the P2P traffic.

A. Behavior class definition

We define an adaptive algorithm that extracts significant clusters of flows from feature dimension, such as srcIP, during a given time interval. Each selected srcIP is connected to the other three "free" feature dimensions, namely srcPrt, dstIP and dstPrt, referred to as X, Y and Z respectively (cf. Table III). For each set of flows sharing the same key, e.g. a srcIP value, we can calculate the relative uncertainty (RU) to the other three free dimensions, characterized through a vector of values of RU[RUx, RUy, RUz].

Table III – Identifiers' convention

Group Key	Free dimensions		
	X	Y	Z
`srcIP`	`srcPrt`	`dstPrt`	`dstIP`
`dstIP`	`srcPrt`	`dstPrt`	`srcIP`
`srcPrt`	`dstPrt`	`srcIP`	`dstPrt`
`dstPrt`	`srcPrt`	`srcIP`	`dstPrt`

As RU can assume values between 0 and 1, three levels of relationship were defined for use representing the association between a significant cluster and each one of its free feature-dimensions [6]. These levels are defined by L(RU) = {0,1,2} according to an established threshold ε, as follows:

$$L(RU) = \begin{cases} 0 \text{ (low)}, \text{ if } 0 \le RU \le \varepsilon, \\ 1 \text{ (intermediate)}, \text{ if } \varepsilon < RU < 1-\varepsilon, \\ 2 \text{ (high)}, \text{ if } 1-\varepsilon \le RU \le 1 \end{cases} \quad (1)$$

For instance if the L(RU) from the free feature-dimension dstIP results in 0, this means that a given srcIP has communicated with very few destination IP addresses. On the other hand, if L(RU) of dstPrt takes the value 2, this is to say then that this particular srcIP talks to many destination ports. To establish an adequate behavior classification the L(RU) limits were set using ε = 0.2 as threshold for port feature-dimensions, and ε = 0.3 for IP addresses. For example if an RU value is between 0 and 0.2 then its L(RU) is set to null. These values were evaluated in [6].

The definition of these levels allows the establishment of weighted ranges of identifiers according to the formula: $id = L(Rux)*3^2 + L(Ruy)*3^1 + L(Ruz)*3^0 \in \{0,1,2, \ldots, 26\}$. Using these identifiers, the behavior classes can be characterized as BCid. For example, a srcIP classified as part of a behavior BC2 = [0,0,2], then its identifier is given as: $id = 0*3^2 + 0*3^1 + 2*3^0$. This class refers to the behavior class where given source hosts talk to many destination hosts using only few destination ports and one or few source ports.

B. Significant clusters evaluation: training stage

From the classified significant clusters, a selection of P2P clusters of behavior, BCs, is required. In order to select out the best representative P2P BCs, some graphics were examined. Figure 4 shows the total of P2P bytes in each BC in decreasing order. We can observe that most P2P traffic is within BC17 = [1,2,2], i.e. hosts that are using a medium number of srcPrt to talk to a large number of destination IP addresses on many destination ports.

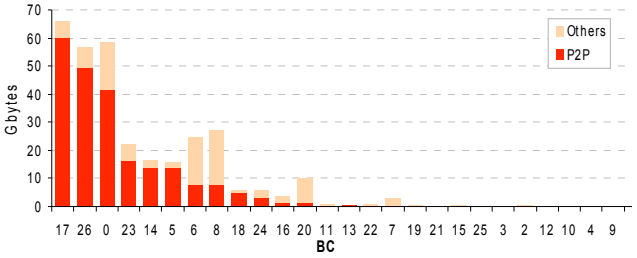

Fig. 4. Bytes distribution of dimension-feature srcIP

Figure 5 shows the values of maximum completeness and accuracy to different combinations of BCs. The first combination is composed by the BC with the higher rate of bytes P2P (BC17). To subsequent combinations, we proceed with the unitary addition of BCs, obeying the decreasing order established in Figure 4. The results in Figure 5 demonstrate that there is a somehow intuitive way for choosing P2P BCs by watching when the two curves representing completeness and accuracy cross each other. According to this criterion, and when considering the feature srcIP, the number of significant BCs is 6, namely BCs 17, 26, 0, 23, 14 and 5 (Figure 4). From this point on completeness increments are small whereas accuracy decreases is significant. The results obtained when considering the feature-dimension dstIP were omitted here for space restrictions; we observed a set with the same number of BCs, composed by the behavior classes 0, 23, 18, 6, 26 and 17, also in decreasing order.

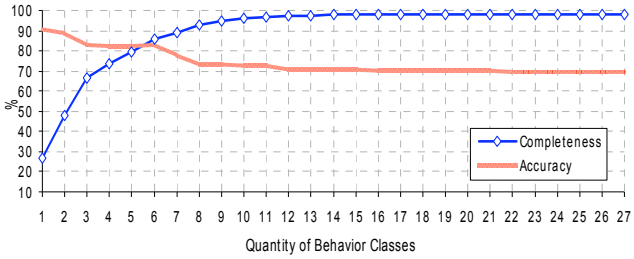

Fig. 5. Accuracy and completeness values for srcIP

Furthermore, for both feature-dimensions, we verified that after six behavior classes there is a tendency for a steep distance increase between the two curves. Then, when there is a large number of BCs, the curves became asymptotic. We have observed this characteristic for BC combinations bigger than sixteen for srcPrt and nineteen for dstPrt.

C. Heuristic for behavior class selection

We now introduce a BC selection method based on thresholds of both accuracy and completeness metrics, which is composed of two selective procedures: selection from completeness rates and categorization based on accuracy. First, we study the completeness values for each behavior class. Our study aimed at identifying a method to select automatically BCs that concentrated the P2P traffic. For this purpose, we built a procedure with the following steps:

1. Maximum completeness calculation: consists of obtaining the completeness sum for all BCs;
2. Completeness threshold definition: this threshold determines the smallest acceptable percentage between the completeness value of a given BC and the maximum completeness.
3. Behavior class selection from completeness: here, we divided the completeness of each BC by the maximum completeness. Only the BCs with rates greater than the threshold are selected.

Next, we present a method to categorize the selected BCs based on the accuracy metric. This procedure takes the following steps:

1. Sorting BCs according to their accuracy values: BCs are arranged in decreasing order of accuracy.
2. Defining the number of categories: this step consists

of defining ranges for accuracy. For our evaluations during the training step, we considered three ranges of accuracy. These were in decreasing order of accuracy: 1 (high), 2 (intermediate) and 3 (low).

3. Defining threshold of categorization: These thresholds establish the borders among the categories. In our methodology, we used the standard deviation value.
4. Categorization of selected BCs: Using the obtained sequence of sorted BCs in step 1, we assigned to each BC a category. In order to do so, we go through the sorted sequence of BCs, summing the accuracy on each related BC, calculating the standard deviation according to the most accurate BC and comparing the resulting value against the thresholds.

Thus, each BC will represent an accuracy category. Finally, from these categories, we can define classification rules for flows that consider both completeness and accuracy metrics.

D. Final results of the training stage

Adding these last two procedures to our approach, we introduced a new methodology to efficiently extract and to categorize behavior classes of P2P traffic. The final results of the testing phase are based on the same Table I traces and are described in Table IV.

V. P2P IDENTIFICATION: TESTING PHASE

Previously, the training stage has pointed us to the best BCs capable of identifying P2P traffic generated by file sharing applications. In this section, we evaluate the NTPC testing stage using real network traffic measurements from different sources, (Points of Presence PoP, Universitiy, ISPs, etc). Due to similarity of results and lack of space, the traffic data presented in this paper are the outcome of measurements taken at a countrywide academic network backbone. In order to make the collected data more representative of the traffic diversity, we sniffed the network for several days, in order to improve the traffic behavior capture. We selected a representative sample from this collection. It is worth stressing that experiments using different traffic data from other sources also provided very similar results.

Table IV. Selected BCs

Category (Precision)	BCs by feature-dimension			
	srcIP	srcPrt	dstIP	dstPrt
High (1)	5	20, 17, 0, 11, 23	18,0, 11, 23	18, 24
Medium (2)	17, 14, 26, 18, 23, 0	-	26	25
Low (3)	24, 8, 6	26,25	6, 14, 24, 17	0, 26, 17, 23

NTPC testing selects and classifies the significant clusters from the input trace. Then, the tool compares each flow's feature-dimension to the set of P2P BCs built by the training stage.

A. Flows Identification algorithm

In this paper, we introduce an algorithm which we will use to identify P2P flows based on two information criteria: classified significant clusters and P2P BCs that were identified and categorized by the training phase. The input trace behavior is designed for the information theoretic classification of each feature-dimension (srcIP, srcPrt, dstIP and dstPrt). In order to extract significant clusters, the NTPC during the testing stage is set exactly using the same procedures and parameters as those applied during the training stage. To identify the target flows, the classified feature-dimensions are compared to the categorized BC categories. This comparison consists in verifying pre-established rules of the identification algorithm. That identification method is called the test module, shown in Fig. 7.

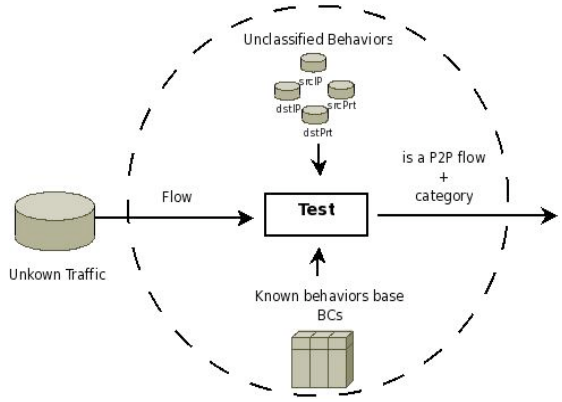

Fig. 7 – Test unit for flows

This module classifies a specific feature-dimension of a given flow and forwards it with a non-zero category number or zero. When the number zero is returned, it indicates that the respective feature-dimension was not clustered or it has no P2P behavior. Figure 8 shows the tree algorithm composed of a couple of test nodes combinations. Our rule set now considers the flow's significant components and the previous category number, except for the first test module. Each test module is represented by the letter T and an identification number.

This rule set concerns the hierarchical feature-dimension evaluation, whose components are checked for obeying the following order: srcIP, dstIP and ports. Therefore, the IP address has a higher priority than ports in the rule set. Hence the ports information is rather used

to refine the flow identification of low level IP address categories.

Consequently, a flow shall be identified as P2P if a sequence of conditions had been matched. According to Figure 8, a given flow would be identified as element of P2P if the evaluation of test node sequence T1, T4 and T10 were successful. For instance, T1 has to assign the value 3 to srcIP's category and forward it to T4. T4 receives the flow and the result of T1 categorization as input information, and it evaluates if the feature-dimension dstIP belongs to category 1. Finally, T10 verifies if any port feature-dimension has a category value between 1 and 3. Once these conditions apply, the flow will be classified as a P2P flow.

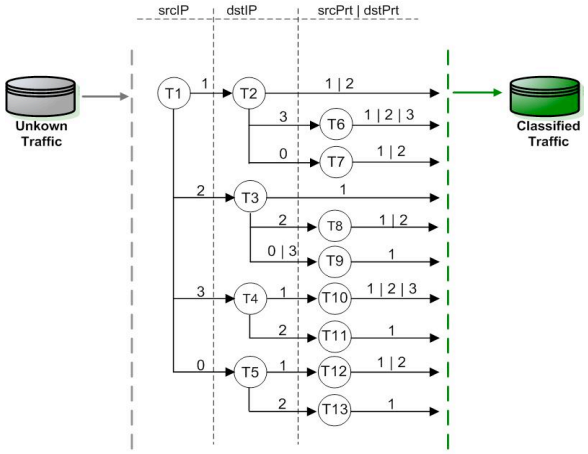

Fig. 8. Classification scheme

B. Results evaluation

Here, we present the final NTPC results during the last test stage for the processing of the collected traces. Table V shows the comparative results between the percentages of identified bytes and flows. NTPC tool identified 77.50% of P2P bytes with an accuracy rate of 93.12%, a whole of 1025.46GB of data representing 9,074,263 flows. This amount of flows corresponds to 18.68% of P2P flows in the test trace, which are insignificant in terms of traffic volume, since most of them are P2P control signals and hence carry small quantities of traffic. The accuracy from identified flows was 39.43%. We emphasize that our focus is to identify such heavy hitters.

Table V. NTPC results

	Completeness (%)	Accuracy (%)
Bytes	77.50	93.12
Flows	18.68	39.43

Despite the low rate of completeness in terms of flows, our approach could successfully recognize the majority of P2P volume. It means that our tool can efficiently identify P2P flows that concentrate huge amounts of bytes. Consequently, this identification process allows any policy for traffic shaping could be accurately enforced to such P2P flows. More precisely, it has been shown that one single technique it is not enough to provide accurate traffic classification [10]. Therefore, we argue that our methodology can be seen as complementary to a broader real-time classification tool.

VI. CONCLUSIONS AND FUTURE WORK

P2P traffic has evoked many concerns for managing and engineering networks, causing bandwidth misuse and copyright problems. In this work, we have presented a strategy that identifies P2P traffic without resorting to deep packet inspection and have shown its efficiency. Our file sharing P2P traffic identification methodology is capable of classifying the majority of P2P traffic volume with high level of accuracy. Our P2P identification framework relies on network and transport layer network information, identifies communication patterns automatically and establishes simple rules to traffic classification.

We envisage that our methodology could also be fine-tuned to identify applications that also transfer large amount of traffic volume, as IPTV. As future work, we intend to evaluate our methodology in the scope of streaming video identification, since it has been showing to grow in popularity.

REFERENCES

[1] M. Crovela, B. Krishnamurthy. "Internet measurement infrastructure, traffic & applications", JW & Sons Ltd. England. 2006.

[2] W. B. Norton. "The evolution of the U.S. Internet peering ecosystem,"2003

[3] T. Karagiannis, A. Broido, N. Brownlee, Kc Claffy. "Transport layer identification of P2P traffic," ACM IMC, 2004.

[4] K. Xu, et al, "Reducing unwanted traffic in a backbone Network," in Proc. of SRUTI Workshop, July 2005.

[5] T. Karagiannis, et al., "BLINC: Multilevel traffic classification in the dark", In ACM SIGCOMM, August,2005.

[6] K. Xu, et al, "Profiling internet backbone traffic: behavior models and applications," ACM SIGCOMM, 2005.

[7] C. E. Shannon and Weaver. "The Mathematic Theory of Communication," University of Illinois Press, 1949.

[8] Kun-Chan Lan, J. Heidemann. "A measurement study of correlation of Internet flow characteristics." Computer Networks, vol 1, num 50, 2006.

[9] Simon, G. J., Kumar, V., and Zhang, Z. "Semi-supervised approach to rapid and reliable labeling of large data sets". In 14th ACM SIGKDD KDD '08. ACM, New York, NY.

[10] Szabó G, Orincsay D, Malomsoky S, Szabó I. On the Validation of Traffic Classification Algorithms. Passive and Active Measurements Workshop 2008 (PAM 2008), April 2008.