King Saud University

Journal of King Saud University – Computer and Information Sciences

www.ksu.edu.sa
www.sciencedirect.com

Threshold-based generic scheme for encrypted and tunneled Voice Flows Detection over IP Networks

M. Mazhar U. Rathore

School of Electrical Engineering and Computer Science, National University of Sciences and Technology, Islamabad, Pakistan

Received 14 February 2013; revised 24 May 2014; accepted 4 June 2014
Available online 18 June 2015

KEYWORDS

Voice over IP (VoIP);
Deep packet inspection;
Internet service provider (ISP);
Telecommunication authority;
Tunnel

Abstract VoIP usage is rapidly growing due to its cost effectiveness, dramatic functionality over the traditional telephone network and its compatibility with public switched telephone network (PSTN). In some countries, like Pakistan, the commercial usage of VoIP is prohibited. Internet service providers (ISPs) and telecommunication authorities are interested in detecting VoIP calls to either block or prioritize them. So detection of VoIP calls is important for both types of authorities. Signature-based, port-based, and pattern-based VoIP detection techniques are inefficient due to complex and confidential security and tunneling mechanisms used by VoIP. In this paper, we propose a generic, robust, efficient, and practically implementable statistical analysis-based solution to identify encrypted, non-encrypted, or tunneled VoIP media (voice) flows using threshold values of flow statistical parameters. We have made a comparison with existing techniques and evaluated our system with respect to accuracy and efficiency. Our system has 97.54% direct rate and .00015% false positive rate.

1. Introduction

Voice over Internet protocol (VoIP) is a mechanism that sends voice over the IP based network. Use of VoIP for commercial purpose is growing day by day due to the advantages of cost effectiveness, functionality over the traditional telephone network and its compatibility with PSTN. The main steps that are involved in VoIP call setup are signaling and media channel setup. The signaling is used to setup connection between two communicating parties. The media channel setup is the actual voice transmission channel between two parties after a successful signaling; it includes digitization of voice, encoding, packetization and transmission of the voice packets over the packet switched network. SIP and H.323 are mostly used signaling protocols and RTP is mostly used media transmission protocol. Some detection methods are applied on signaling traffic and others on media traffic. There are also some methods that examine both signaling and media traffic. VoIP signaling and media transmission both may be encrypted or any one may only be encrypted. The media session may be encrypted by SRTP, SSL/TLS, IPSec, or propriety protocols. The signaling may be encrypted by SIPS, SSL/TLS, SMIME, IPSec, or propriety protocols. Port-based, signature-based, pattern-based VoIP detection techniques are inefficient due to complex, confidential and secure privacy protocols used

E-mail address: 09msccsmrathore@seecs.edu.pk

Peer review under responsibility of King Saud University.

ELSEVIER Production and hosting by Elsevier

http://dx.doi.org/10.1016/j.jksuci.2014.06.016
1319-1578

by VoIP. Moreover these techniques are specific to VoIP applications or protocols. Statistical analysis-based techniques are generic and can detect VoIP hidden in secure tunnels. So we are proposing a statistical analysis based solution for VoIP call detection.

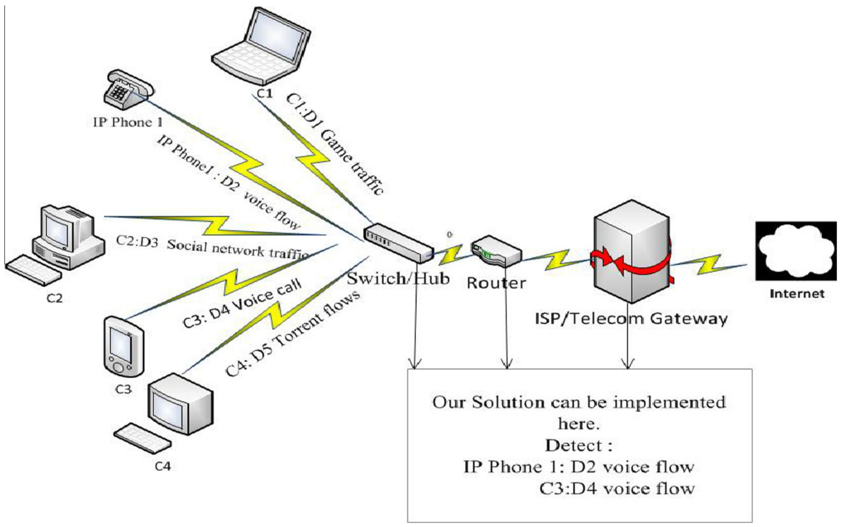
Detection of VoIP traffic is important by two aspects; one for blocking or restricting commercial usage of VoIP, other for prioritizing it. In some countries, like Pakistan, use of VoIP for commercial purposes is prohibited as it incurs loss of a large amount of money to the national telecom operator. In spite of this, a significant amount of data traffic traveling through the internet today is commercial VoIP. Pakistan Telecom Authority (PTA), being the regulatory telecom authority of Pakistan, is interested in detecting the commercial use of VoIP and punish illegal operators. On the other hand, ISPs or other service providers may also want to prioritize VoIP for the paying customers. Multiple solutions (Li et al., 2010; Li et al., 2010; Idrees and Aslam Khan, 2008; Freire et al., 2008; Yildirim and Radcliffe, 2010; Maiolini et al., 2009; Nguyen and Armitage, 2008; Yildirim and Radcliffe, 2010; Lin et al., 2009; Dusi et al., 2009; Alshammari and Zincir-Heywood, 2011; Li et al., 2007; Rossi et al., 2008; Bonfiglio et al., 2008; Alshammari and Zincir-Heywood, 2010) exist for detecting encrypted VoIP but they are neither suitable for telecom authorities and ISPs to detect VoIP calls irrespective of VoIP application nor have good results in case of encrypted and tunneled VoIP detection. Some of them are not generic, others could not provide real time detection. In this paper, we propose a statistical analysis-based solution using threshold values of flow statistical parameters to identify the VoIP media (voice) flows. The solution is generic, efficient, accurate and real time (to some extent) and detects encrypted, non-encrypted, and tunneled VoIP. It is independent from any VoIP application, protocol, security mechanism, or tunneling mechanism and practically implementable at telecommunication authority or ISP gateway to either block or prioritize VoIP traffic. Fig. 1 shows the network model of our proposed solution. Our solution can be implemented on any network device such as router, telecommunication/ISP gateways, servers etc. as in Fig. 1. (IP Phone1, D2) and (C3, D4) are IP pairs that communicate voice and rest of the terminals sent non-voice traffic. Our solution will detect IP Phone1:D2 and C3:D4 flows as voice flows.

The rest of the paper is organized as follows. Section 2 discusses the work that has been done in VoIP detection. Section 3 represents the datasets that are analyzed and tested. Section 4 presents the statistical analysis. The proposed system is discussed in Section 5. Evaluation and results are shown and comparison is made in Section 6. Section 7 concludes the work.

2. Background and related work

There are basically 4 types of VoIP detection techniques as mentioned by Rathore and Mehmood (2012), i.e. Port-based, signature-based, pattern-based, and statistical analysis-based techniques. By port-based analysis techniques, the traffic is classified by examining port number at the transport layer. IANA specified some standard ports to specific applications such as VoIP uses 5060, 5061 ports for SIP signaling, 1718 to 1720 for H.323 signaling, and port 2427, 2944 for media gateway control protocol (MGCP), H.248 and Megaco protocols. In Leung and Chan (2007), Baset and Schulzrinne (2006), port-based analysis are used as helping information to detect VoIP. By Renals and Jacoby (2009), Skype VoIP traffic is detected by matching distinct Skype keywords, ports, and content. Usage of non standard and dynamical allocation of ports makes port-based detection inefficient.

Signature-based techniques detect VoIP using deep packet inspection by matching specific strings within packet payload. SIP packet has string "sip" within packet payload. RTP header mostly starts with 0x80, 0x81. ZRTP packet contains "1000xxxx5a525450" at the start of pay load (ZRTP header). In Renals and Jacoby (2009), Skype VoIP traffic is detected by matching distinct contents as well. By Renals and Jacoby (2009) Skype packets sometimes contain the keywords "/getla testversion?ver=" or "/getnewestversion" combined with "/ui/" string. By Renals and Jacoby (2009), the outgoing data packets of Skype contain content "16 03 01 00 00", the incoming packets have content "17 03 01 00 00" and if the packet of these contents is blocked, Skype tries to send a new packet that



Our Solution can be implemented here.
Detect :
IP Phone 1: D2 voice flow
C3:D4 voice flow

Fig. 1 Network model of threshold-based generic scheme for encrypted and tunneled Voice Detection over IP Networks.

contains “03 01 00 cd 41 03 00 09 80 40 04 08 c0 00” and “00 0c 01 17 03 01 00”. Detection of VoIP in Adami et al. (2009), Lu et al. (2010), Birke et al. (2010) also used signature-based technique. Use of encrypted and tunneling mechanism in VoIP makes signature-based techniques inaccurate. Moreover signatures vary from application to application and protocol to protocol.

By pattern-based analysis, the particular pattern of signaling communication of different VoIP applications is to be identified. Many researches (Leung and Chan, 2007; Baset and Schulzrinne, 2006; Lu et al., 2010; Peranyi and Molnar, 2007) have been done for detecting Skype traffic by pattern analysis. By investigating Skype, the paper (Leung and Chan, 2007) discuss 15 basic stages of Skype communication from start to end such as start up, registration, etc. and also reveals all the entities and nodes that are participating in conversation such as Skype client, super node (SN), etc. Pattern-based techniques depend on signaling mechanism that is varied from application to application and protocol to protocol.

Statistical analysis-based techniques are generic and can be used for encrypted and tunneled traffic. Statistical analysis are mostly performed on voice data by taking flow features as input such as packet size, arrival time, etc. The techniques (Li et al., 2010; Li et al., 2010; Idrees and Aslam Khan, 2008; Freire et al., 2008; Yildirim and Radcliffe, 2010; Maiolini et al., 2009; Nguyen and Armitage, 2008; Yildirim and Radcliffe, 2010; Lin et al., 2009; Dusi et al., 2009; Alshammari and Zincir-Heywood, 2011; Li et al., 2007; Rossi et al., 2008; Bonfiglio et al., 2008; Alshammari and Zincir-Heywood, 2010) are statistical approaches to detect VoIP. In Li et al. (2010), the IP addresses and ports are examined. In host behavioral analysis, the difference D between source ports and destination ports for a particular flow must be less than threshold. Moreover the inter-arrival packet time measures are used in detecting VoIP flows.The main flaw in this approach is that it could not provide a real time solution for VoIP detection; as first you have to calculate the number of source and destination ports used for a particular source destination pair. In some cases it had false positive ratio more than 10% which is still large. Moreover it could not handle VoIP into IPSec tunnel. Fauzia and Uzma (Idrees and Aslam Khan, 2008) separate out VoIP traffic by using traffic feature, that are difficult to alter, such as packet interval time, packet sizes, rate of exchange. This technique only considered UDP traffic. Some VoIP applications may use TCP for voice transmission when UDP is blocked and SSL/TLS VoIP also uses TCP. Moreover there are many VoIP applications that transmit voice packets of size less than 100 bytes, so in this case this statistical technique could not detect that VoIP flows. It could also not handle IPSec VoIP. Freire and Ziviani describe a scheme (Freire et al., 2008) that detects VoIP calls hidden in web traffic, on port 80 and 443, by using web request size, web response size, inter arrival time between requests, no. of requests per page, page retrieval time. They use goodness-of-fitness test, Kolmogorov–Smirnov (KS) distance and chi-square values and obtain metrics to identify VoIP in web traffic. This technique has good results but It is specific to VoIP hidden in the web traffic using port 80,443. It only supports the http version 1.1. Moreover it is also not real time detection and need more prior data for analysis and detection.VoIP hidden in IPSec tunnel could not be detected by Freire et al. (2008). Yildirim and Radcliffe proposed a statistical technique (Yildirim and Radcliffe, 2010) to identify the protocol such as VoIP within encrypted tunnel by using probabilistic information and packet size distribution (PSD) of flows. This technique only considered one VoIP application (Skype) for analysis and testing. Only 3 voice codec schemes are analyzed by this technique. Moreover, only packet size is used as a basic parameter to identify the VoIP traffic so more false positives. Similar to other techniques, this technique could also not used for IPSec tunneled VoIP. Ying-Dar and Chun-NanLu (Lin et al., 2009) also proposed a generic technique to classify the network traffic into different application types by using packet size distribution (PSD) and port association. In the case of VoIP classification it only analyzed two VoIP applications i.e. MSN and Skype. similar to previous technique this technique also only depends on packet sizes, so results might not be more accurate. Results show that in the case of MSN VoIP detection, there is 9% false positive and in the case of Skype VoIP detection, there is 18% false negative. Riyad (Alshammari and Zincir-Heywood, 2010) detects VoIP traffic by using flow features such as size and time and evaluates three different machine learning (ML) techniques namely C4.5, AdaBoost, and SBB-GP. This technique only analyzed Gtalk and Skype, other important application such as Yahoo, MSN, Zfone etc are given no importance. All results are specific to these two applications. Toshiya Okabe proposed flow level behavior (FLB) VoIP detection technique in Okabe et al. (2006) that uses packet size and inter arrival time to measure average, median, and distribution for VoIP detection. Yildirim and Radcliffe (2010) proposed a simplest statistical technique for VoIP identification hidden in IPSec tunnel by considering packet size only. This technique identifies VoIP packets whose packet size lies within a certain limit. It can not be used for detection purpose to block VoIP calls as it has more false positives and false negatives. The latest work done in this field using statistical analysis by Branch and But (2012). They described the construction and performance of classifiers able to identify variable rate VoIP flows. They use machine learning techniques to classify VoIP flows by constructing C4.5 decision tree using J48 algorithm as does by Li et al. (2007) and Alshammari and Zincir-Heywood (2010). They use flow statistical parameters such as minimum mean of packet length at each direction, normalized ratio of the number of bytes in both direction, and absolute one packet difference as input to the classifier. The technique (Branch and But, 2012) although needs just part of the flow for voice classification but it still needs to train the classifier. Moreover this technique could only be implemented at two way interface only as it needs both directions traffic for VoIP flows detection. The authors of the paper analyzed and tested only two VoIP applications traces which are not enough. Table 1 shows the work done by using statistical methods, features and parameters, and techniques used and VoIP applications for which the system is tested. Table 2 shows comparative study among these techniques by considering limitations in terms of whether they are generic (nor dependent on any VoIP application/protocol), whether they can be implemented at one-way or two-way interface, whether they can detect IPSec hidden VoIP, and whether they are specific to VoIP detection. We have seen, these statistical techniques neither meet the requirements to be generic, efficient, more accurate, independent from VoIP applications/protocols and security mechanisms nor they are

Table 1 Modern statistical techniques.

Ref.	Year	Parameters used	Techniques used	VoIP Applications tested
RGIPVTF (Branch and But, 2012)	2012	$\overline{X}$ pkt-size (each direction), Normalized ratio (both direction), packet diff.	C4.5 Decision tree, J48 algorithm	Skype, Gtalk
HFBA (Li et al., 2010)	2010	No.of ports, packet time	Difference of no. of ports, ratio of small and large inter-packet arrival times	Skype
PDF-PSD (Yildirim and Radcliffe, 2010)	2010	Packet size, packet time	Prob. density function, Packet size distribution (PSD)	Skype
IPSec VoIP detection (Yildirim and Radcliffe, 2010)	2010	Packet size	packet size rang only	Own VoIP setup
C4.5, AdaBoost, SBB-GB (Alshammari and Zincir-Heywood, 2010)	2010	Packet size, time, mean, S.D, max-time, etc.	C4.5, adaBoost, SBB-GB classifiers	Gtalk, Skype
PSD-PA (Lin et al., 2009)	2009	Packet size, ports	PSD, ports	Skype, MSN
K-means classifier (Maiolini et al., 2009)	2009	Packet size, time, direction	K-means classifier (only 1st few pkts)	Nil
Statistical thresholds (Idrees and Aslam Khan, 2008)	2008	Packet size, exchange rate	Mean and S.D by threshold	Skype, MSN, Yahoo, Gtalk
VoIP hidden in web traffic (Freire et al., 2008)	2008	Request and respond size, time, no. of requests	Goodness-to-fitness test, KS distance, chi-square	Skype, Gtalk
J48, REP tree (Li et al., 2007)	2007	Packet size, time, flow duration	J48, REP tree	MSN, Skype
FLB (Okabe et al., 2006)	2006	Packet size, time	Flow level behavior (FLB)	SIPSoftphone, Netmeeting, Skype, Kaza

Table 2 Modern statistical techniques.

Technique	Generic (w.r.t VoIP application or protocol)?	Supported interface (one-way, two-way)	IPSec VoIP detection?	Specific to VoIP detection?
HFBA (Li et al., 2010)	Yes	Two-way	No	Yes
Statistical thresholds (Idrees and Aslam Khan, 2008)	Yes	Both	No	Yes
VoIP hidden in web traffic (Freire et al., 2008)	No	Two-way	No	Yes
IPSec VoIP detection (Yildirim and Radcliffe, 2010)	Yes	Both	Yes	Yes
K-means classifier (Maiolini et al., 2009)	No	Two-way	No	No
PDF-PSD (Yildirim and Radcliffe, 2010)	Yes	Both	Yes	Yes
PSD-PA (Lin et al., 2009)	Yes	Two-way	No	No
J48, REP tree (Li et al., 2007)	Yes	Both	No	No
C4.5, AdaBoost, SBB-GB (Alshammari and Zincir-Heywood, 2010)	No	Two-way	Yes	Yes
FLB (Okabe et al., 2006)	No. (specific to some VoIP applications)	Both	No	Yes
RGIPVTF (Branch and But, 2012)	yes	Two-way	yes (but not tested)	Yes

practically implementable at telecommunication authorities' gateways to either block or prioritize VoIP calls. Here in this paper, we are going to address these limitations.

3. Datasets collection

Datasets are collected at different time from different environments and locations for analysis and testing. The datasets are collected from (1) NUST SEECS WISNET lab (2) home users' computers (3) Pakistan Telecommunication Authority (PTA) and Pakistan Telecommunication Limited (PTCL) gateways, that include traffic from wired as well as wireless networks such as EDGE, GPRS, etc. (4) sample traces downloaded from Wireshark site (http://wiki.wireshark.org/SampleCaptures) (5) Skype voice traces downloaded from tstat site (http://tstat.tlc.polito.it/traces-skype.shtml) (6) by making own simple encrypted and non encrypted VoIP setups using Asterisk as VoIP server and Zfone, X-lite, Eyebeam, Blink as client. Table 3 presents the information of our own VoIP setup traces such as minimum size, number of captured traces, and minimum duration of trace. e.g. "C-SRTP-RTP" in Table 3 shows four conversations of two VoIP clients in which one VoIP client traffic is encrypted by SRTP and other client traffic is un-encrypted and minimum size of these conversations is .5MB and duration is 82 s. Moreover traces of VoIP that use SSL/TLS and IPSec tunnels are also captured. Voice

Table 3 VoIP setup traces.

Trace	Size (MB)	No. of files	Duration (sec)
A-RTP-RTP	1.5	4	478
B-RTP-SRTP	.5	4	66
C-SRTP-RTP	.5	4	82
D-SRTP-SRTP	1.5	4	151
Zfone-X-lite	.25	1	32
Asterisk voice	4.5	1	151

Table 4 VoIP testing traces.

VoIP application	Versions	Min-size (MB)	Min-duration (sec)
Gtalk	1.0.0.104 beta, Gmail voice	3	504
Skype	4.0.0.215, 5.5.0.119, 5.5.0.124	3	664
MSN	7.5,8.0,15.4	1	88
Yahoo (SSL tunnel)	9.0, 10.1, beta, 11.0	2	332
Oovoo	3.5.9.4	4	123
QQ messenger	1.6	1	57
Trillian IM	3.1.12.0	2	133
Mix VoIP		40	10,714

traffic traces from different VoIP applications such as Gtalk beta version, Skype 4.0.0.215, Skype 5.5.0.119 and Skype 5.5.59.124, Yahoo 9.0.0.2152, Yahoo 10.0, Yahoo beta, Yahoo 11.0, MSN 7.0, MSN 8.5, MSN 15.4.3538.0513 and Windows Live messenger are taken for analysis. These traffic traces are described in Table 4. non-VoIP applications traces are also captured such as YouTube, torrents, antivirus updates, videos, online live TVs, audio songs, FTP downloads, online games, web mails such as Gmail, Yahoo mail, Hotmail, Bluetooth, chatting, DNS traffic, document retrieval, frame relay, remote access, SMTP, SSH, and telnet-remote access traces. Moreover mixture of VoIP and non-VoIP traffic traces are also collected for testing. Main non-VoIP applications traces that are analyzed and tested are described in Table 5. The details of tstat Skype traces (http://tstat.tlc.polito.it/traces-skype.shtml) are presented in Table 6 with codec, transport protocols used, size of trace, and time duration of the traffic. These traces contain both UDP and TCP voice conversations. Moreover large size dumps are collected from PTA gateway of 2-4 GB, and hundreds of dumps of 1GB from PTCL gateway. These large traces include traffic from wired as well as wireless networks.

Table 5 non-VoIP traces.

Trace	Min-size (MB)	No. of files	Min-duration (sec)
Gmail-Yahoomail	3	5	156
Hotmail	3	3	101
Mix (VoIP-NonVoIP)	65	4	1023
NonVoIP-mix	112	6	1331
Torrent-YouTube-Gmail	1	1	431
YouTube	9	6	97
Online TV	2	1	88
Bittorent	150	5	2043

Table 6 Skype tstat traces.

Trace	Codec	Transport protocol	Size (MB)	Duration (sec)
E2E-140606-1	G729	UDP	8	905
E2E-140606-2	iLBC	UDP	11	1003
E2E-140606-3	iSAC	UDP	12	1116
SkypeOut-260906-1	G729	TCP	9	919
SkypeOut-260906-2	G729	UDP	7	910
Internet-E2X		TCP	212	343,562
Internet-E2O		UDP	264	343,562
Internet-E2E		UDP	3.5GB	344,700
Internet-SIG		UDP	6GB	344,700

4. Statistical analysis

Our aim of statistical analysis is to find distinct threshold values of flow statistical parameters for voice flows. Wireshark is used for simple analysis and capturing traffic. C language is used for complex analysis. The code is written in C using Winpcap 4.1.2 library to analyze the offline as well as online traffic. The proposed algorithm is also developed in C using Winpcap. The statistical analyses are performed on traces by two ways; firstly, the statistical parameters of each flow are calculated and analyzed for complete session and in the 2nd phase the statistical parameters are calculated and analyzed by taking different seconds traffic chunks for each flow. We noticed that the system could not give better results when we considered traffic chunks of less than 5 s for each flow but when we considered traffic chunks of more than 5 s, it tended to degrade the performance of system. So to maintain the equilibrium between accuracy and efficiency we chose 5,5 s traffic chunks for analyzing each flow for VoIP call detection. All VoIP and non-VoIP applications are statistically analyzed in this way. Moreover the RTP, SRTP, ZRTP, SSL, TLS, IPSec voice traffic is deeply analyzed. The traffic of these protocols is detected by signatures and then analyzed from PTA and PTCL dumps.

The flow is distinguished by 4 tuples source-IP, destination-IP, source-port, and destination-port (S-IP, D-IP, S-Port, D-Port), as every packet has these properties. In the case of IPSec, the transport layer information is encrypted so the flow is distinguished by 3 tuples source-IP, destination-IP, and security parameters index (S-IP, D-IP, SPI). Main statistical parameters that are used to analyze each flow are packet rate (pkt-rate) in packets/sec, mean ($\overline{X}$) and standard deviation (S.D) of packets' sizes in bytes, and maximum difference between the current and previous packets' time for all packets (max-diff-time), mean and standard deviation of the difference between the current and previous packet time ($\overline{X}$ (diff-time), S.D (diff-time)), all in seconds. IP layer packet size is considered for all measurements.

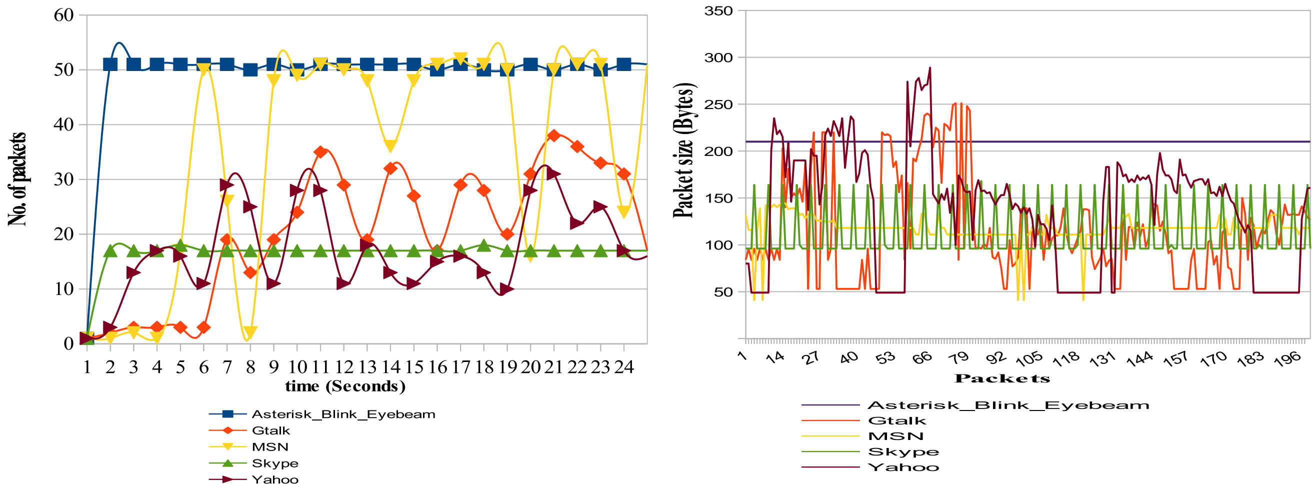
VoIP is not tolerant to delay, latency, jitter, and packet loss which affect the quality of voice. Voice packet total delay consists of packet creation time plus network transmission timeout plus receiving buffering and decoding time. Latency is the delay in packet delivery. Variation in delays is called jitter. More latency, jitter and packet loss degrade the quality of voice. The total delay of a voice packet is increasing function of packet size. If the packet size is larger, more voice is encoded in a packet; it will take more packet creation time, transmission time, and decoding time, resultantly the total packet delay is increased. To maintain the quality of voice, the delay should be bearable and the packet size should also be within a limit. Moreover in the case of voice, the loss of large size packet means the loss of more voice which is not tolerant. Due to these facts the voice packet length must lie within a certain limit to maintain the quality. Jitter also affects the quality of voice as larger variation in packet delays does not produce clear voice at the receiver side. There should be bearable variation in delays in the case of voice packets to maintain voice quality. By considering these facts, we use IP layer packet size as the basic parameters for statistical analysis. We analyzed the flows for packet size distribution (PSD) and packet rates. The packet size distribution and packet rate of different VoIP applications are shown in Fig. 2. We observed that the voice flows has higher packet rates, so we considered the non-VoIP applications that have higher packet rates and then analyzed them by considering PSD, shown in Fig. 3. We noticed that PSD graph for both VoIP flows and non-VoIP flows, that have higher packet rates, is quite different. Hence on the basis of these analyses we chose first three parameters i.e. packet rate, $\overline{X}$ (size) and S.D (size) as statistical parameters for VoIP traffic analysis to distinguish voice flows. Moreover the ITU-T recommends to capsulate 20-30 ms voice in a packet for better performance and quality assurance. It shortens packet size and increases packet rate i.e. more packets per second as compared to other applications. This is also another reason to choose packet rate as parameter for statistical analysis. It is also a fact that voice has a continuous behavior as the voice packets are continuously sent when a person speaks on VoIP phone. Very short time is elapsed between the current and previous voice packets. We consider this fact and take max-diff-time, $\overline{X}$ (diff-time), and S.D (diff-time) as statistical measures for time-based analysis to distinguish voice flows. The parameters and corresponding values ranges for voice flows on different VoIP application are shown in Table 7 by considering 5,5 s chunks of traffic for each flow.

We have also observed that some VoIP applications such as Gtalk, MSN send small number of packets at the start of the media session. The range of these packets is 2-15 packets in first 10–15 s. Sometime max-diff-time is quite high i.e. greater than 1 for first or last packet of the flow. In such cases, the absolute value of difference ($\overline{X}$(diff-time)-S.D(diff-time)) also exceeds from normal range. Other non-VoIP traffic traces such as YouTube, torrent, antivirus updates, FTP downloads, online live TVs, mail servers traffic (Gmail, Yahoo mail, Hotmail), online games traces are also analyzed on these parameters. The values of statistical parameters are quite distinctive for both VoIP and non-VoIP flows. So on the basis of these parameters we can identify VoIP flows efficiently and with higher accuracy.

5. Proposed System

The flow is distinguished by 3 tuples i.e. source IP, destination IP, security parameters index (S-IP,D-IP, SPI) in the case of IPsec and 4 tuples i.e. source IP, destination IP, and source port, destination port (S-IP, D-IP, S-port, D-port) in the case of all other traffic. On the basis of parameters values' ranges as in Table 7, detailed statistical analysis of different VoIP applications presented in Section 4, and study of voice codecs, standards, and facts about voice transmission, we defined 8 rules for VoIP detection. These rules are:

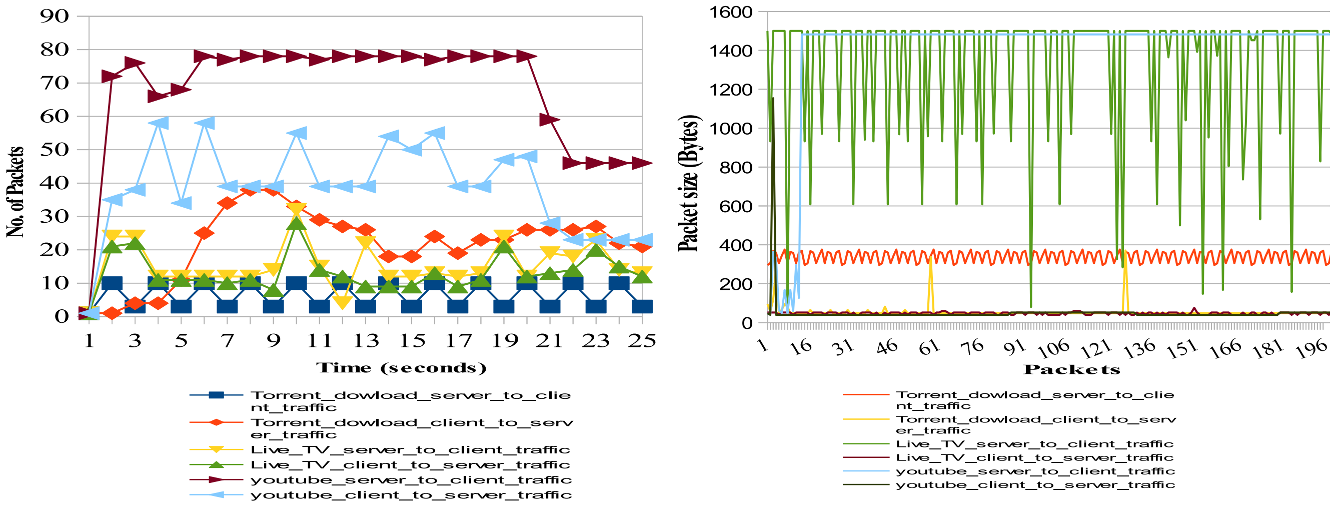
1. pkt-rate $>$ 13 packets/sec
2. $56 \leqslant \overline{X}$ (size) $\leqslant 210$ bytes
3. $0 \leqslant$ S.D (size) $\leqslant 75$ bytes
4. $\overline{X}$ (size) $\geqslant$ S.D (size)
5. $0 <$ max-diff-time $\leqslant .8$ s
6. $0 < \overline{X}$ (diff-time) $\leqslant .09$ s



(a) Packet rate (packet/seconds)

(b) Packet size distribution

Fig. 2 Packet rate and packet size distribution of different VoIP applications.



(a) Packet rate (packet/seconds)

(b) Packet size distribution

Fig. 3 Packet rate and packet size distribution of non-VoIP applications.

Table 7 Statistical parameters values ranges for VoIP application considering 5,5 s traffic for each flow.

Trace	Pkt-rate	$\overline{X}$ (size)	S.D (size)	Max-diff-time	$\overline{X}$ (diff-time)	S.D. (diff-time)	$\lvert\overline{X}$-S.D.$\rvert$ (diff-time)
Skype	16–50	60–140	.38–27	.075–.393	.019–.061	.008–.12	0–.07
Gtalk	17–37	90–170	5–65	.101–.426	.027–.056	.011–.578	0–.02
Yahoo	12–37	64–170	1–75	.065–.49	.026–.086	.010–.073	0–.03
MSN	17–50	120–140	05–20	.06–.74	.020–.058	.005–.055	0–.02
Asterisk traces with Zfone, X-lie, Eyebeam clients	17–30	190–210	0–40	.02–.41	.010–.046	.05–.49	0–.032

7. $0 <$ S.D (diff-time) $\leqslant .25$ s
8. $0 < |\overline{X}$-S.D (diff-time) $|\leqslant .1$ s

A flow is a VoIP flow if and only if the first 4 rules are true and at least 3 rules from the last 4 rules are satisfied, because of the finding presented in the last paragraph of Section 4. A flow is confirmed non-VoIP if the rate rule is true but any one rule from the rules 2, 3, 4 is false. If the flow is neither VoIP nor non-VoIP then it is either a suspected flow or not to be decided yet. If the flow is suspected for first 5 s traffic then it is reinvestigated for next 5 s traffic and if it remains suspected 3 times then it is detected as non-VoIP flow. The algorithm pseudo code is:

```
1. For each packet determine the flow to which it belongs. If no flow is found then register it as a new flow uniquely distinguished by (S-IP,D-IP,SPI) for IPSec or by (S-IP,D-IP,S-port,D-port) for other traffic and calculate parameters.
2. Capture the first 80 packets or all the packets within 5 s for each flow.
3. Investigate the flow for VoIP only if the flow total no. of packets > 65 within 5 s.
    (a) If (all first 4 rules == true) and if (at least 3 rule from rule 5, 6, 7, 8 == true) then the flow is VoIP flow.
    (b) If (all first 4 rules == true) and if (less than 3 rule from rule 5, 6, 7, 8 == true) then the flow is suspected. Suspected flows are reinvestigated for next phase (next 5 s traffic).
    (c) If (rate rule == true) and if (any rule from rule 2, 3, 4 == false) then the flow is confirmed non-VoIP.
    (d) If (flow suspected ⩾3 time) then it is non-VoIP flow.
```

6. Evaluation and comparison

We evaluate our system with respect to accuracy and efficiency and by comparing results with the existing systems. We use typical parameters such as true positive (TP), false negative (FN), true negative (TN), false positive (FP), direct rate (DR), and false positive rate (FPR) that are mostly used for measuring accuracy of a system. TP is the measure of flows that are correctly identified as VoIP flows. FN is the measure of flows that are incorrectly identified as non-VoIP flows. TN is the measure of flows that are correctly identified as non-VoIP flows. FP is the measure of flows that are incorrectly identified as VoIP flows. DR reflects how much VoIP flows are correctly identified as VoIP flows and calculated by Eq. 1. FPR reflects how much non-VoIP flows incorrectly identified as VoIP flows and calculated by Eq. 2. The ideal solution is that which has 100% DR and 0% FPR.

Table 8 Overall accuracy results.

Traffic	DR = TP/ (TP + FN)%	FPR = FP/ (FP + TN)%
Real-time traffic	92.5	.0002
Off-line traffic	96.56	.00013
Own-VoIP-setup traces	100	–
Tstat Skype traces	97.78	–
Overall	97.54	.00015

$$DR = TP/(TP + FN) \quad (1)$$

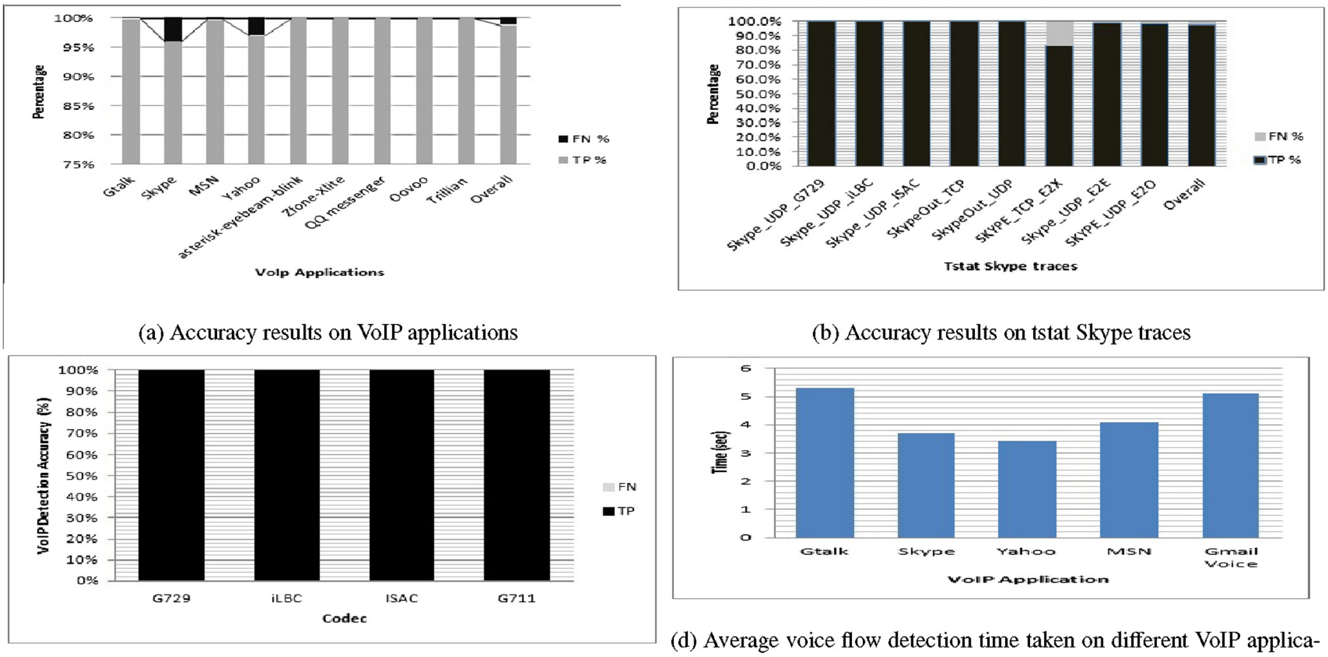
$$FPR = FP/(FP + TN) \quad (2)$$

Table 8 shows overall accuracy results on all datasets presented in Section 3 such as real time traffic, offline captured traces, sample traffic (http://wiki.wireshark.org/SampleCaptures), and tstat Skype traces (http://tstat.tlc.polito.it/traces-skype.shtml) as well as on mixture of VoIP and non-VoIP traffic traces. Our system has 97.54 % DR which is quite higher and .00015% FPR which is quite lower. Accuracy results on VoIP applications traces are shown in Fig. 4(a). Gtalk, MSN, Zfone, X-lite, and Asterisk with Eyebeam and Blink as clients give 100% TP and 0% FN. Only Skype and Yahoo has TP bit lower than 100%. Accuracy results on tstat Skype traces are shown in Fig. 4(b). Only Skype voice trace "SKYPE-TCP-E2X", that uses TCP, has quite higher FN nearer to 17% but overall accuracy performance on these traces is better. The results on different voice codecs are shown in Fig. 4(c). All these codecs have 100% TP.

The efficiency is measured in terms of VoIP flow detection time. The detection time of voice calls is less than 6 s; as we only consider the small part of the flow traffic (i.e. first 60 packets or packets within 5 s for each flow) for VoIP detection. Fig. 4(d) shows the average detection time of voice flows on different VoIP applications. These results are taken from real time implementation by communicating voice using different VoIP applications. MSN, Yahoo, and Skype voice flows are detected within 5 s and Gtalk, Gmail voice flows take more than 5 s to be identified as they send less numbers of packets within 5 s.

We compare our technique with host and flow behavior analysis (HFBA) technique (Li et al., 2010), threshold-based detection (Idrees and Aslam Khan, 2008), and IPSec tunneled VoIP detection (Yildirim and Radcliffe, 2010), flow level behavior (FLB) technique (Okabe et al., 2006) in terms of TP, FP, and FN. HFBA (Li et al., 2010), threshold-based detection (Idrees and Aslam Khan, 2008), and FLB (Okabe et al., 2006) techniques gave more importance to Skype voice traffic in analysis and testing, so we consider larger size 3.5 GB tstat Skype trace "Internet-Skype-UDP-E2E" (http://tstat.tlc.polito.it/traces-skype.shtml) for comparing these VoIP detection techniques with our technique. Table 9 shows results and comparison on tstat 3.5GB Skype trace. Our technique is a threshold-based statistical analysis-based technique which can also detect IPSec tunneled voice calls, so we compare our technique with threshold-based VoIP detection technique presented in Idrees and Aslam Khan (2008) and IPSec VoIP detection technique(Yildirim and Radcliffe, 2010) on all datasets collected and presented in Section 3. Table 10 shows overall comparison.

We have observed that our system performance is better than existing techniques with respect to accuracy and efficiency. Moreover our system is generic that can detect VoIP traffic regardless of the VoIP application, protocol, codec, and security mechanism. The system can detect tunneled VoIP such as SSL/TLS and IPSec VoIP. The system can be implemented at one-way or two way network interface. It meets the need of telecommunication authorities and ISPs



(a) Accuracy results on VoIP applications

(b) Accuracy results on tstat Skype traces

(c) Accuracy results on different codecs

(d) Average voice flow detection time taken on different VoIP applications voice flows

Fig. 4 Testing results.

Table 9 Comparison between our technique and existing techniques w.r.t accuracy on 3.5 GB tstat Skype traces.

Technique	TP%	FN%
Host and flow behavior analysis (HFBA) (Li et al., 2010)	90.28	9.72
Threshold-based detection (Idrees and Aslam Khan, 2008)	79.2	20.8
Flow level behavior (FLB) (Okabe et al., 2006)	55.6	44.4
Our technique	98.86	1.14

Table 10 Comparison between our technique and existing techniques w.r.t accuracy on different captured traces.

Technique	Offline traffic	
	TP %	FP
Threshold based detection (Idrees and Aslam Khan, 2008)	35.7	.0001
IPSec VoIP detection (Yildirim and Radcliffe, 2010)	73	25
Our technique	96.56	.00013

for detecting VoIP flows to either prioritize or block them. We test our solution on many traces of more than 10 VoIP applications which are more than others techniques tested applications. Moreover our solution is designed after detailed statistical analysis of VoIP applications, voice codecs, and voice standards. So we assume that the accuracy and efficiency results will be quite similar on other VoIP applications as on the tested applications.

7. Conclusion

In this paper we have proposed a generic, robust, efficient, and practically implementable statistical analysis-based solution to identify encrypted, non-encrypted, or tunneled VoIP media flows using threshold values of flow statistical parameters by giving one-way or two-way traffic. So it is the best choice for telecommunication authorities and ISPs to detect VoIP calls. The system is tested on large datasets of different VoIP and non-VoIP traffic. The comparisons and results show that our technique is the best among all the existing techniques. This technique has 97.54% TP and .00015% FP.

References

Adami, D., Callegari, C., Giordano, S., Pagano, M., Pepe, T., 2009. A real-time algorithm for skype traffic detection and classification. In: 9th International Conference on Smart Spaces and Next Generation Wired/Wireless Networking and 2nd Conference on Smart Spaces, Postersburg, Russia.

Alshammari, Riyad, Zincir-Heywood, A. Nur, 2010. An investigation on the identification of VoIP traffic: case study on Gtalk and Skype. In: 2010 International Conference on Network and Service Management (CNSM), Niagara Falls, Canada, pp. 310–313.

Alshammari, Riyad, Zincir-Heywood, Nur, 2011. Can encrypted traffic be identified without port numbers, IP addresses and payload inspection? Elsevier Comput. Netw. 55, 1326–1350.

Baset, Salman A., Schulzrinne, Henning, 2006. An analysis of the skype peer-to-peer internet telephony protocol, INFOCOM 2006. In: 25th International Conference on Computer Communication, Barcelona, Spain, pp. 1–11.

Birke, Robert, Mellia, Marco, Petracca, Michele, Rossi, Dario, 2010. Experiences of VoIP traffic monitoring in a commercial ISP. Int. J. Netw. Manage. 20, 339–359.

Bonfiglio, Dario, Mellia, Marco, Meo, Michela, Ritacca, Nicol'o, Rossi, Dario, 2008. Tracking down skype traffic, INFOCOM 2008. In: 27th IEEE International Conference on Computer Communication, Phoenix, Arizona, USA, pp. 261–265.

Branch, P., But, J., 2012. Rapid and generalized identification of packetized voice traffic flows. In: 37th IEEE Conference on Local Computer Networks (LCN12), Clearwater, Florida.

Dusi, M., Crotti, M., Gringoli, F., Salgarelli, L., 2009. Tunnel hunter: detecting application-layer tunnels with statistical fingerprinting. Elsevier Comput. Netw. 53, 81–97.

Freire, Emanuel P., Ziviani, Artur, Salles, Ronaldo M., 2008. Detecting VoIP calls hidden in web traffic. In: IEEE Transaction on Network and Service Management, vol no. 5, pp. 210–214.

<http://tstat.tlc.polito.it/traces-skype.shtml>.

<http://wiki.wireshark.org/SampleCaptures>.

Idrees, Fauzia, Aslam Khan, Uzma, 2008. A generic technique for Voice over Internet Protocol (VoIP) traffic detection. In: IJCSNS International Journal of Computer Science and Network Security, vol.8 no. 2, pp. 52–59.

Leung, Chun-Ming, Chan, Yuen-Yan, 2007. Network forensic on encrypted peer to-peer VoIP traffics and the detection, blocking, and prioritization of skype traffics. In: 16th IEEE International Workshops on Enabling Technologies: Infrastructures for Collaborative Enterprises, Paris, France, pp. 401–408.

Li, Jun, Zhang, Shunyi, Liu, Shidong, Xuan, Ye, 2007. Active P2P traffic identification technique. In: 2007 International Conference on Computational Intelligence and Security, Harbin, China, pp. 37–41.

Li, Bing, Jin, Shigang, Ma, Moade, 2010. VoIP traffic identification based on host and flow behavior analysis. J. Netw. Syst. Manage. 19.

Li, Bing, Jin, Shigang, Ma, Moade, 2010. VoIP traffic identification based on host and flow behavior analysis. In: 2010 International Conference on Wireless Communication Networking and Mobile Computing (WICOM), Chengdu, China, pp 1–4.

Lin, Ying-Dar, Lu, Chun-Nan, Lai, Yuan-Cheng, Peng, Wei-Hao, Lin, Po-Ching, 2009. Application classification using packet size distribution and port association. Elsevier: J. Netw. Comput. Appl. vol. 32, 1023–1030.

Lu, Feng, Liu, Xiao-Lei, Zhi-Nan, M.A., 2010. Research on the characteristics and blocking realization of Skype protocol. In: 2010 IEEE International Conference on Electrical and Control Engineering (ICECE), Wuhan, China, pp. 2964–2967.

Maiolini, Gianluca, Molina, Giacomo, Baiocchi, Andrea, Rizzi, Antonello, 2009. On the fly application flows identification by exploiting K-means based classifiers. J. Inf. Assur. Security 4, 142–150.

Nguyen, Thuy T.T., Armitage, Grenville, 2008. Clustering to assist supervised machine learning for real-time IP traffic classification. In: IEEE International Conference on Communication, Beijing, China, pp. 5857–5862.

Okabe, T., Kitamura, T., Shizuno, T., 2006. Statistical traffic identification method based on flow-level behavior for fair VoIP service. In: Proceedings of IEEE Workshop on VoIP Management and Security, pp. 35–40.

Peranyi, M., Molnar, S., 2007. Enhanced Skype traffic identification. In: ValueTools, Proceedings of the 2nd International Conference on Performance Evaluation Methodologies and Tools, ICST (Institute for Computer Sciences, Social-Informatics and

Telecommunications Engineering), ICST, Brussels, Belgium, Belgium, pp. 1–9.

Rathore, M.M.U., Mehmood, T., 2012. Research on VoIP traffic detection. In: 2012 International Symposium on Performance Evaluation of Computer and Telecommunication Systems, SPECTS 2012, Genoa, Italy.

Renals, P., Jacoby, G.A., 2009. Blocking Skype through deep packet inspection. In: The 42nd Annual Hawaii International Conference on System Sciences, HICSS'09, Big Island, HI, pp. 2009.

Rossi, Dario, Mellia, Marco, Meo, Michela, 2008. A detailed measurement of skype network traffic. In: Proceedings of the 7th International Conference on Peer-to-Peer Systems, USENIX Association.

Yildirim, Taner, Radcliffe, P.J., 2010. VoIP traffic classification in IPSec tunnels. In: 2010 International Conference on Electronics and Information Engineering (ICEIE), Koyoto, Japan, vol. 1, pp. VI-151-VI-157.

Yildirim, Taner, Radcliffe, P.J., 2010. A framework for tunneled traffic analysis. In: 12th International Conference on Advance Communication Technology (ICACT), Phoenix Park, Korea, vol 2, pp. 1029–1034.