ISSUES IN VOIP AND NEW INFORMATION HIDING PROTOCOL TO DEAL WITH THESE ISSUES

Satish Hadap*, Pooja Singh and Samruddhi Naik****

*Subject Matter Expert on High Performance Computing, BNP Paribas India Solutions Private Limited, Mumbai
**Support Analyst, BNP Paribas India Solutions Private Limited, Mumbai

Abstract **-**This paper provides a comprehensive understanding of Voice over Internet Protocol (VoIP), issues with VOIP and deals with the study of new information hiding protocol which addresses issue with VoIP service. The new protocol described in this paper is an alternative for the IETF's (Internet Engineering Task Force) RTCP (Real Time Control Protocol) for real-time application's traffic. Additionally, this solution offers authentication and integrity, which is capable of exchanging and verifying QoS and security parameters. It is based on digital watermarking and steganography that is why it does not consume additional bandwidth and the data transmitted is inseparably bound to the voice content.

Index Terms **-** VOIP, issues in VOIP, steganography, digital watermarking, new information hiding protocol for VOIP.

I. INTRODUCTION

VoIP stands for Voice over Internet Protocol. It is also referred to as IP Telephony or Internet Telephony. It is another way of making phone calls, with the difference of making the calls cheaper or completely free. The 'phone' part is not always present anymore, as you can communicate without a telephone set. VoIP has a lot of advantages over the traditional phone system. The main reason for which people are so massively turning to VoIP technology is the cost. VoIP is said to be cheap, but most people use it for free. Yes, if you have a computer with a microphone and speakers, and a good Internet connection, you can communicate using VoIP for free. This can also be possible with your mobile and home phone. However, the two most important fields in which Voice over Internet Protocol (VoIP) is lacking are providing certain Quality of Service (QoS) parameters and security considerations, which we will describe in this paper. Also In this paper, we study a new protocol that covers both those fields simultaneously with the use of Steganography and Digital Watermarking. It provides information that is vital to control the network conditions and to verify authentication of the source and data integrity.

II. DATA HANDLING IN VOIP

Before any voice can be sent, a call must be placed. In an ordinary phone system, this process involves dialing the digits of the called number, which are then processed by the telephone company's system to ring the called number. With VOIP, the user must enter the dialed number, which can take the form of a number dialed on a telephone keypad or the selection of a Universal Resource Indicator (URI), but after that a complex series of packet exchanges must occur, based on a VOIP signaling protocol. The problem is that computer systems are addressed using their IP address, but the user enters an ordinary telephone number or URI to place the call. The telephone number or URI must be linked with an IP address to reach the called party, much as an alphabetic web address, such as "www.nist.gov" must be linked to the IP address of the NIST web server. A number of protocols are involved in determining the IP address that corresponds to the called party's telephone number.

Figure 1 below illustrates the basic flow of voice data in a VOIP system. Once the called party answers, voice must be transmitted by converting the voice into digitized form, then segmenting the voice signal into a stream of packets. The first step in this process is converting analog voice signals to digital, using an analog-digital converter. Since digitized voice requires a large number of bits, a compression algorithm can be used to reduce the volume of data to be transmitted. Next, voice samples are inserted into data packets to be carried on the Internet. The protocol for the voice packets is typically the Real-time Transport Protocol, RTP (RFC 3550). RTP packets have special header fields that hold data needed to correctly re-assemble the packets into a voice signal on the other end. But voice packets will be carried as payload by UDP protocols that are also used for ordinary data transmission. In other words, the RTP packets are carried as data by the UDP datagrams, which can then be processed by ordinary network nodes throughout the Internet. At the other end, the process is reversed: the packets are disassembled and put into the proper order, digitized voice data extracted from the packets and uncompressed, then the digitized voice is processed by a digital-to-analog converter to render it into analog signals for the called party's handset speaker

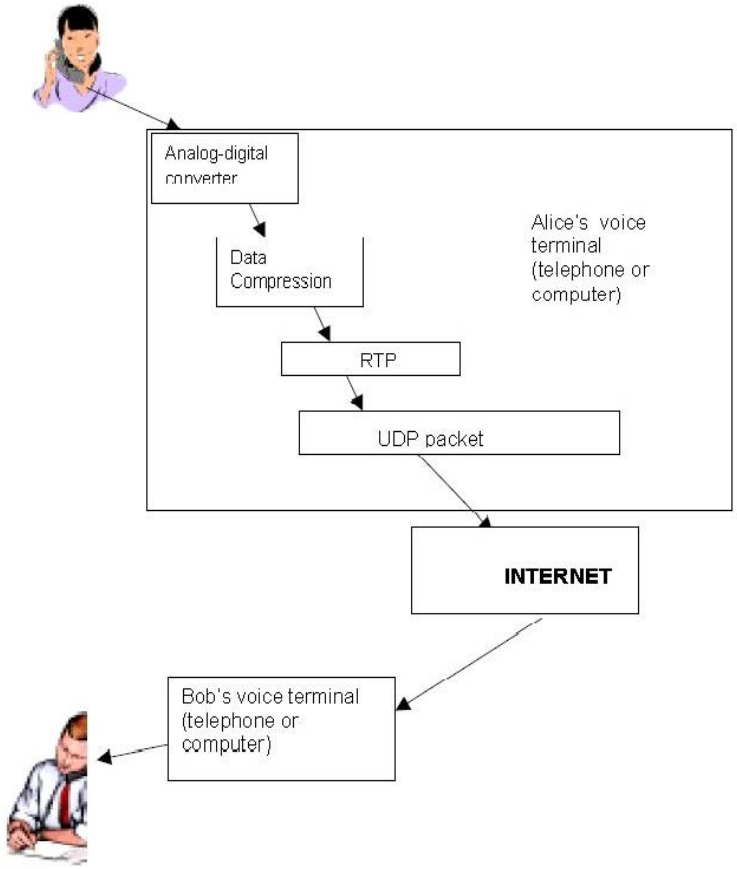

Figure 1: Voice Data Processing in VOIP system

.

III. VOIP ISSUES

Vulnerable or inadequate security and need to maintain high Quality of service (QoS) are two main challenges with VOIP.

A. VOIP Security Issues

With the introduction of VOIP, the need for security is compounded because now we must protect two invaluable assets, our data and our voice. Federal government agencies are required by law to protect a great deal of information, even if it is unclassified. Both privacy-sensitive and financial data must be protected, as well as other government information that is categorized as sensitive but unclassified. Protecting the security of conversations is thus required. In a conventional office telephone system, security is a more valid assumption. Intercepting conversations requires physical access to telephone lines or compromise of the office private branch exchange (PBX). Only particularly security-sensitive organizations bother to encrypt voice traffic over traditional telephone lines. The same cannot be said for Internet-based connections. For example, while ordering merchandise over the phone, most people will read their credit card number to the person on the other end. The numbers are transmitted without encryption to the seller. In contrast, the risk of sending unencrypted data across the Internet is more significant. Packets sent from a user's home computer to an online retailer may pass through 15-20 systems that are not under the control of the user's ISP or the retailer. Because digits are transmitted using a standard for transmitting digits out of band as special messages, anyone with access to these systems could install software that scans packets for credit card information. For this reason, online retailers use encryption software to protect a user's information and credit card number. So it stands to reason that if we are to transmit voice over the Internet Protocol, and specifically across the Internet, similar security measures must be applied.

The current Internet architecture does not provide the same physical wire security as the phone lines. The key to securing VOIP is to use the security mechanisms like those deployed in data networks (firewalls, encryption, etc.) to emulate the security level currently enjoyed by PSTN network users. This publication investigates the attacks and defenses relevant to VOIP and explores ways to provide appropriate levels of security for VOIP networks at reasonable cost.

B. Quality of service issues

Quality of Service (QoS) is fundamental to the operation of a VOIP network. Despite all the money VOIP can save users and the network elegance it provides, if it cannot deliver at least the same quality of call setup and voice relay functionality and voice quality as a traditional telephone network, then it will provide little added value. The implementation of various security measures can degrade QoS. These complications range from delaying or blocking of call setups by firewalls to encryption-produced latency and delay variation (jitter). QoS issues are central to VOIP security. If QoS was assured, then most of the same security measures currently implemented in today's data networks could be used in VOIP networks. But because of the time-critical nature of VOIP, and its low tolerance for disruption and packet loss, many security measures implemented in traditional data networks just aren't applicable to VOIP in their current form.

IV. STEGANOGRAPHY AND DIGITAL WATERMARKING

Steganography and Digital Watermarking are two information hiding techniques. The general difference between those two techniques is that steganography's primary objective is to keep the existence of the information secret and digital watermarking's primary purpose is making it imperceptible.

A. Steganography

Steganography is a process of hiding one chunk of data inside another chunk of data, normally transmitted data. Usually it means hiding a secret message within an ordinary message and the extraction of this message at its destination. In ideal situation, anyone scanning data will fail to know it contains covert data. In modern digital steganography, data is inserted into redundant (provided but often unneeded) data, e.g. fields in communication protocols, graphic image, etc. TCP/IP steganography utilize the fact that few headers in packet are changed during transit. We will exploit here a covert channel, which is a method of communication that is not a part of an actual computer system design, but can be used to transfer information to users or system processes that normally would not be allowed access to the information. In TCP/IP stack, there is a number of methods available, whereby covert channels can be established and data can be exchanged secretly between hosts. An analysis of the headers of typical TCP/IP protocols e.g. IP, UDP, TCP, HTTP, ICMP results in fields that are either unused or optional. This reveals many possibilities where data can be stored and transmitted. IP header consists of few fields that are available to be used as a covert channel. Those fields are marked in Figure 2 with italics. The total capacity of those fields exceeds 60 bits per packet. And there are UDP and RTP protocols fields left to be used.

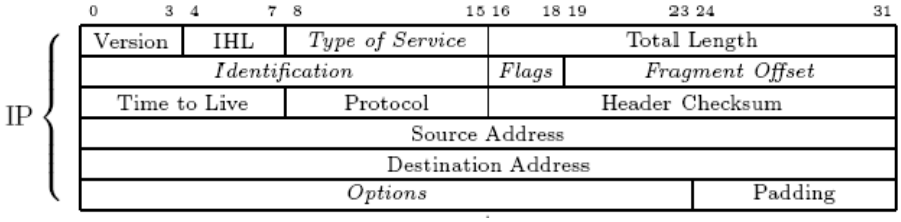

Figure 2: IP header with marked fields (italics) available for steganography

For VoIP and this solution we will exploit unused/optional fields in IP/UDP/RTP packets because those protocols are used in almost all IP telephony implementations. We do not limit this solution to using only IP/UDP/RTP protocol. Lower layers of TCP/IP stack also offer steganography possibilities. Furthermore we can distribute those control bits among those fields in a predetermined fashion (this pattern can be exchanged during a signaling phase of conversation). In those chosen fields we will transmit only header (control bits) of this protocol with the use of steganography technique. Header consists of 6 bits per packet, so such a type of transmission is potentially hard to discover.

B. Digital Watermarking

Digital watermarking is a multidisciplinary methodology widely developed in the last two decades. It covers a large field of various aspects, from cryptography to signal processing, and is generally used for marking the digital data (images, video, audio or text). There are several applications for digital watermarks that include: Fingerprinting (embedding a distinct watermark into every copy of the author's data), Annotation Watermark/Content labeling (embedding information, which describes the digital work that can be later extracted), sage control/Copy control (authors can insert a watermark that indicates the number of copies permitted for each set). However, the most important applications for our purposes are: the possibility of embedding the authentication and integrity watermark and exchanging additional information (for the controlling RTP packets purposes).

The watermark that will be used in, proposed here, authentication and integrity solution must possess certain parameters, like: robustness, security, transparency, complexity, capacity, verification and invertibility. Their optimization for real-time audio system is crucial, which is often mutually competitive. However, there is always a compromise necessary. That is why the embedded watermark, that we will use, must be characterized by high robustness, high security and must be non-perceptual.

Not every watermarking technique is applicable for this solution. IP Telephony is the demanding, real-time service. That is why we need the watermarking schemes that really work for the real-time conversations. Generally, audio watermarking algorithm is based on two functions: embedding of the watermark into voice and its extraction. As soon as the conversation begins, certain information is embed into the voice samples and sent through the communication channel. Then, the watermark is extracted from those samples before they reach the callee and the information retrieved is verified. If the watermark's data sent is correct, the conversation can
be continued.

Most digital watermarking algorithms for the real-time communication are designed to survive typical non-malicious operations like: low bit rate audio compression, codec changes, DA/AD conversion or packet loss. For example, the watermarking scheme developed at the Fraunhofer IPSI (Institut Integrierte Publikations und Informations systeme) and the Fraunhofer IIS (Institut Integrierte Schaltungen) were tested for different compression methods. Those results revealed that the large simultaneous capacity and robustness depend on the scale of the codec compression. When the compression rate is high (1:53), the watermark is robust only when we embed about 1 bit/s. With a lower compression rate we can obtain about 30 bit/s, whereas the highest data rate was 48 bit/s with good robust, transparent and complexity parameters. For the monophonic audio signal, which is a default type for IP Telephony the watermark embedding algorithm appeared around 14 times faster and the watermark detector almost 6 times faster than the real-time.

The next important thing for this scheme is how much information we can embed into the original voice data. This will influence the speed of the authentication and integrity process throughout the conversation. This parameter, in this solution, is expected to be high but it is not crucial. With low compression rates, we propose to add a pre-conversation stage. In this stage there will be few seconds of the RTP packets exchange without the conversation. It will delay the setup of the call but then, during the conversation, the time of verification will be shorter. However, the lowest payload watermarks (about 1 bit/s) cannot be accepted in our scheme because, in this case, the conversation would have to last enormously long to work correctly.

V. NEW PROTOCOL FOR INFORMATION HIDING IN VOIP

The most important security services to secure IP Telephony system are: authentication, integrity and confidentiality. The first two can be provided with the use of this protocol. The third should be guaranteed in a different manner, e.g., with the use of the security mechanisms from a classical security model (the cryptographic mechanisms).

As we described earlier we will use two information hiding techniques: steganography to create covert channel that will be used to transmit header (control bits) and digital watermarking to bind the parameters of the protocol to voice send into the network (watermark). We assume that this solution is for using in IP protocol version 4 networks.

The protocol we are proposing here should possess PDU (Protocol Data Unit), of which size must be kept to minimum. It is important because as we said in the Section 2 the capacity capability of watermarks is limited if we want also watermark other parameters like robustness or security. Every PDU consists of header (control bits) and a certain number of data bits that are embedded into sender/receiver voice. Because the capacity of the watermark depends greatly on the codec's compression rate that is used, so it is possible that the lot of parameters can be distributed into a number of packets. The size (number of bits) of each parameter that will be transmitted with this protocol should be low. For all parameters it should not exceed 32 bits. This value is taken from RTCP protocol size of the parameters. Only one parameter (NTP timestamp) is greater than given value. Limited size of every parameter results in shorter time for the parameter to be transmitted and verified. However we do not dictate this value. It should depend on network bandwidth, status and codec's compression rate.

A. Protocol data unit description

The PDU consists of two parts: the header (control bits) and the watermark data. The header/control fields are transmitted in a covert channel in unused/optional fields of IP/UDP/RTP protocol's headers. Actual value of the parameter is embedded into voice as a watermark. Moreover, the PDU can have one of two payload types: security or informational. Security payload means that PDU contains certain authentication and/or integrity information that should be verified after its extraction. Two kinds of security payloads are available, first is used to provide authentication and integrity of the voice and its source. Second's role is to authenticate protocol parameters that were send earlier (both security and informational).

Table 1: Header fields and their function

Type of field	Number of bits	Function
P (Parameter)	4	Describes parameter that is transmitted in the watermark
S (Side)	1	Describes the side of the communication (1 - sender, 0 - receiver report)
C (Continuity)	1	Describes if a packet contains the beginning or continuation of the parameter indicated in the field P (1 – beginning of new parameter, 0 – continuation of the last parameter)

The header (control bits) is organized in fields as shown in the Table 1
Exemplary values of the field P are shown below:

0001 – authentication or integrity parameter (32 bits)
0010 – parameter: LSR – Last sender report (32 bits)
0011 – parameter: DLSR – Delay of last sender report (32 bits)
0100 – parameter: Interarrival jitter (32 bits)
0101 – parameter: Extend highest sequence number received (32 bits)
0111 – parameter: Cumulative number of packet lost (24 bits)
1000 – parameter: Fraction lost (8 bits)
1001 – parameter: Sender's packet count (32 bits)
1011 – parameter: NTP timestamp (64 bits)
1010 – parameter: RTP timestamp (32 bits)

Another payload type is informational. Each PDU carries one of the parameters that are used to monitor the quality of service and the network conditions.

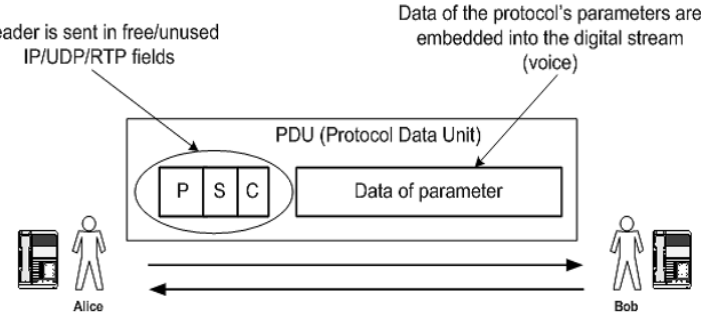

Figure 3: General protocol operation

Usually in one IP/UDP/RTP packet there is about 20-30 milliseconds of voice, which is about 20-30 bytes, depending on type of codec used. Let's say that we are able to embed in average about 10 bits/s of watermark into the voice stream. With that assumption we must send about 3-4 packets to achieve those 10 bits. In this protocol we set parameter's value to 32 bits, so this parameter will be transmitted in about 9-12 packets in more than 3 seconds of the voice. In the example scenario in Figure 4, we see how the exemplary parameter: Interarrival jitter (32 bits) is transmitted for assumption: 10bits/packet.

As we can see in Figure 4 below, parameter characterized by code 0100 (Interarrival jitter) was sent in four IP/UDP/RTP packets. In the first packet, both fields S and C were set to 1. In the next packet field C changed its value to 0 because it is a continuation of the parameter's data that was sent in last packet. At the destination there must be a buffer to extract all data from each packet. After transmitting all packets for one parameter data is available to be used (for QoS monitoring) or to be verified (for security reasons).

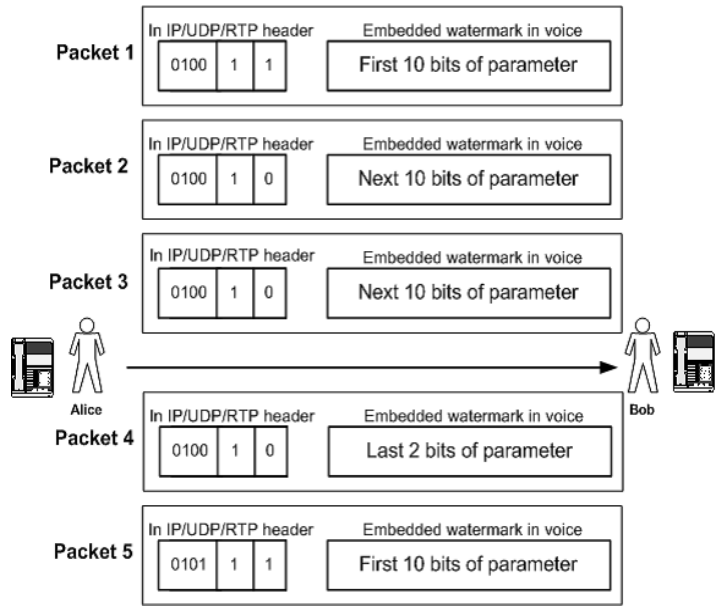

Figure 4: Example of transmission of the Interarrival jitter parameter

B. Authentication and integrity parameter calculation and security payload

Authentication and integrity calculation will be performed with watermark specific considerations. Earlier we mentioned that two security payloads are available:

- one is used to provide authentication and integrity of the voice and its source
- second is to authenticate protocol parameters (both security and informational) that were send earlier

First security parameter is a combination of user global identification and features that were extracted from the voice stream. It is expected that this parameter will have 32 bits. So if the concatenation of those two values exceeds this number of bits, there will be a hash function (marked as H) performed. Then only predetermined bits will be transmitted as a security parameter.

Second security payload is a special parameter that will be used to provide greater security of the whole digital stream and transmission. The general idea of its calculation is presented in Figure 5.

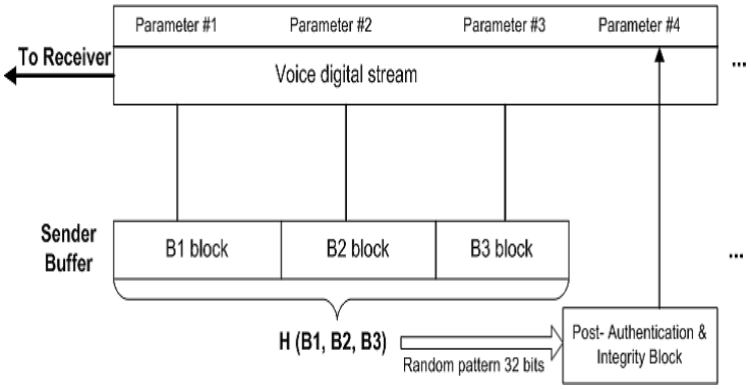

Figure 5: Example of authentication and integrity mechanism for transmitted parameters

C. Level of Trust (LoT) mechanism

We can still imagine a situation, in which the attacker disrupts transmission of the header/controls bits. In this situation the receiver is unable to retrieve any parameters that were transmitted by the sender. Here we will describe a mechanism that will prevent such a situation. Both parties of the conversation will update special parameter named LoT (Level of Trust), during a conversation. If a parameter (security or informational) is received and verified, LoT value increases. In any other situation its value decreases. Additionally parameters that are exchanged during conversation influence LoT value differently. Informational parameters (QoS) add/subtract to LoT's value 1, first kind security parameters 2 and the second kind security parameter 5. If A sends to B a parameter, the algorithm of handling the LoT parameter (on B side) works, as described below in a pseudo-code:

```
START /* CL - Critical Level, LoT - Level of Trust, T - timer */
CL = a; LoTA = x; TA = 0; /* Initiating values */
StartTimer(TA);
FOR (i = 0; i++; i< End of Transmission) /* i - Time slot */
{
IF (ParameterA correct) THEN
{
LoTA + {1 or 2 or 5}; *
ResetTimer(TA);
}
ELSE (LoTA – {1 or 2 or 5}); *
IF (LoTA <= CL) OR (TA > k) THEN STOP; (1)
IF (LoTA = a*x) THEN LoT = x; (2)
}
* value depends on the type of parameter (QoS, security)
```

As we can see, the breakage of the call (or notification to the calling parties) will take place if the value of the LoT parameter is equal or below the given threshold (CL value) or if the timer TA expires (1). The LoT value changes during the conversation time. If every signaling message is successfully verified, the LoT value rises. To prevent its increase from reaching the infinity, we lower it, as soon as it reaches the value of the critical level multiplied by the start value of LoT (2).

This way of decreasing the LoT value has one serious disadvantage: it allows an attacker to wait until LoT= (a*x)-1. But we must assume that he is able to possess information about its value and then safely spoof ((a*x) - 1 – (CL + 1)) audio packets without LoT's falling below the threshold (CL). To prevent it, one must choose the initiating values (a and x) carefully. Their values should depend on network's parameters: the packet loss and possible delays. If the network does not suffer heavily from the packet loss, those values must be low. In the other case, they must be set to a higher level. For example, the network administrator or service provider can circumscribe those parameters for a certain network/user.

VI. COMPARISON

Earlier security protocols in VOIP made use of H.323 and SIP to provide authentication and integrity using various profile levels. However, it failed in providing the required level of confidentiality for IP Telephony or voice transmission. New protocol for information hiding was aimed at providing high level of authentication and integrity for VOIP. Moreover, it also provides considerate confidentiality for voice transmission. Apart from improving various security issues, this protocols also resolves various QoS issues like Jitter, Latency, etc. which was not provided by earlier security protocols for VOIP.

VII. CONCLUSIONS

A comprehensive explanation on VOIP has been provided. VOIP implementation has two major issues which impacts its security and Quality of Service. A new protocol for information hiding for VoIP service was presented. It uses two information hiding techniques: steganography to create covert channel in which the header (control bits) are passed and digital watermarking to transmit the actual data (parameter's value) in voice stream. The most important advantages of this solution are no consumption of available bandwidth, providing security, parameters to monitor QoS and network status in one protocol. We would emphasize that the process of sending information for this protocol is continuous in time and although the bit rate per second offered by watermarking is usually not very high, when we consider a whole conversation we can observe that we are able to exchange quite an amount of data. The different kind of parameters that can be implemented using this solution is not limited to security/monitoring status of the network. Hence, this solution can be further used to data types other than VOIP.

VIII. REFERENCES

[1] J. Dittmann, A. Mukherjee, M. Steinebach, "Media-independent Watermarking Classification and the need for combining digital video and audio watermarking for media authentication", Proceedings of the International Conference on Information Technology: Coding and Computing, IEEE Computer Science Society, Las Vegas, Nevada, USA (2000) pp. 62-67.

[2] M. Steinebach, F. Siebenhaar, C. Neubauer, R. Ackermann, U. Roedig, J. Dittmann, "Intrusion Detection Systems for IP Telephony Networks", Real time intrusion detection symposium, Estoril, Portugal (2002) pp. (17)1-9

[3] D. Richard Kuhn, Thomas J. Walsh, Steffen Fries: Security Considerations for Voice Over IP Systems, Computer Security Division, Information Technology Laboratory, National Institute of Standards and Technology (2004).

[4] B. Goode, "Voice Over Internet Protocol (VOIP)". Proceedings of the IEEE, VOL. 90, NO. 9, Sept. 2002.

[5] H. Schulzrinne, S. Casner, R. Frederick, V. Jacobson: RTP: A Transport Protocol for Real-Time Applications, IETF, RFC 3550, July 2003.

[6] S. Miner and J. Staddon. Graph-based authentication of digital streams. In Proceedings of the IEEE Symposium on Research in Security and Privacy, pages 232-246, May 2001.
[7] S. J. Murdoch, S. Lewis: Embedding Covert Channels into TCP/IP. Information Hiding 2005: 247-26.
[8] K. Ahsan, D. Kundur: Practical Data Hiding in TCP/IP. In: Proceedings of Workshop on Multimedia Security at ACM Multimedia '02, Juan-les-Pins (on the French Riviera), December 2002.
[9] WojciechMazurczyk and Zbigniew Kotulski: New security and control protocol for VoIP based on steganography and digital watermarking.
[10] R. Anderson, (Ed.): Proceedings of: Information Hiding .First International Workshop, Cambridge, U.K., May 30, June 1, 1996, vol. 1174 of Lecture Notes in Computer Science, Springer-Verlag Inc.
[11] K. Szczypiorski: HICCUPS: Hidden Communication System for Corrupted Networks. In Proc. of: The Tenth International Multi-Conference on Advanced Computer Systems ACS'2003, October 22-24, 2003 Międzyzdroje, Poland, pp.31-40, ISBN 83-87362-61-1
[12] T. Friedman, R. Caceres, A. Clark: RTP Control Protocol Extended Reports (RTCP XR), IETF, RFC 3611, November 2003.
[13] V. Korjik, G. Morales-Luna: Information Hiding through Noisy Channels, Proceedings of 4th International Information Hiding Workshop, pp. 42-50, Pittsburgh, PA, USA, April 2001
[14] M. K. Mihcak and R. Venkatesan: A Perceptual Audio Hashing Algorithm: A Tool For Robust Audio Identification and Information Hiding, Proceedings of 4th International Information Hiding Workshop, Pittsburgh, PA, April 2001.

IX. AUTHORS

First Author – Satish Hadap, completed Bachelor in Computer Engineering from Mumbai University. Currently working as a Subject Matter Expert on High Performance Computing in Risk Department, BNP Paribas India Solutions Private Limited, Mumbai. Email Id :Satish.hadap@yahoo.com

Second Author – Pooja Singh, completed Bachelor in Electronics and Telecommunications Engineering from Mumbai University. Currently working as a Support Analyst in Risk Department, BNP Paribas India Solutions Private Limited, Mumbai. Email Id :poojass91@gmail.com

Third Author – Samruddhi Naik, completed Bachelor in Electronics Engineering from Mumbai University. Currently working as a Support Analyst in Global Markets Solutions Department, BNP Paribas India Solutions Private Limited, Mumbai. Email Id :samruddhi.04@gmail.com