An Analysis of VoIP Secure Key Exchange Protocols against Man-in-the-Middle Attack

G.Aghila
Department of Computer Science
Pondicherry University, Puducherry,India.

D.Chandirasekaran
Department of Computer Science
Pondicherry University, Puducherry,India.

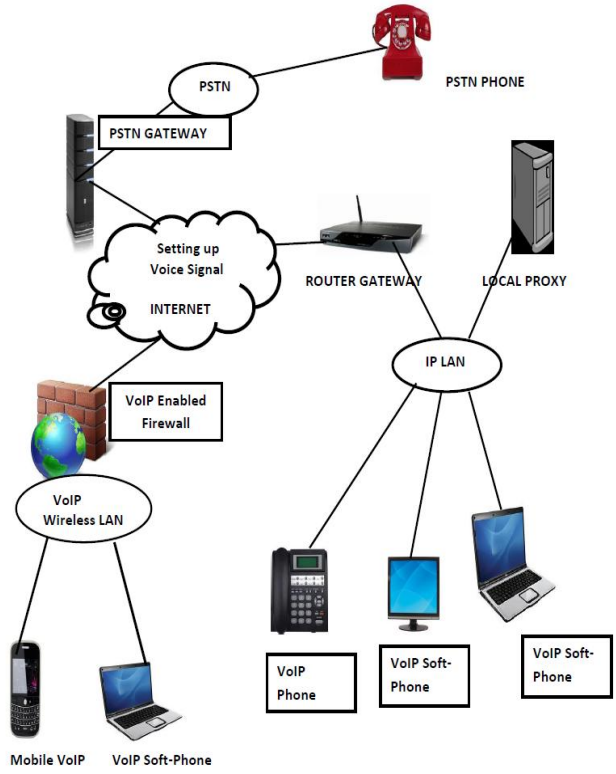
ABSTRACT
This paper presents a brief survey about the existing key exchange protocols namely MIKEY, ZRTP and SDES. The core features of these protocols and their suitability to SIP-VoIP Networks are analyzed in this paper. The focused research area in VoIP is related to the Security and Quality of service of the Voice data. Among these areas VoIP security and confidentiality of voice data turns to be a challenging one .As VoIP delivers the voice packet over the public internet, using the transparent IP protocol suite the confidentiality of the voice data is at risk. Moreover exchanging cryptographic keys to encrypt the media stream in the Session Initiation Protocol has proven to be quite difficult task. There is a need for stronger key management protocols which will secure the voice data from all types of attack and which also provides a feasible key exchange mechanism. Each of these three key management protocols is surveyed and in addition its resistant against Man-In-The-Middle Attack has also been analyzed.

General Terms
Key Exchange Protocol, Man-In-The-Middle Attack.

Keywords
SIP, SRTP, VoIP, ZRTP, SDES, MIKEY

1. INTRODUCTION
Voice over IP (VoIP) technology involves the transmission of digitized voice data, which is obtained by converting the analog speech signals [7]. Then the voice data is compressed by applying voice codec and the compressed voice data is sent over the Internet. Compared to the traditional Public Switched Telephone Network (i.e. PSTN) the VoIP has an added advantage due to the low cost devices and unlike the traditional PSTN network the VoIP does not require dedicated lines to carry the voice data traffic. In today's world many business organizations prefer the VoIP networks because of its low cost.

The VoIP infrastructure is composed of terminals or endpoints (i.e. phones) proxy servers, gateways and IP-based network as shown in Fig 1. Interactions with the telephone in Public Switched Telephone Network (PSTN) are done using the gateways. The signaling protocol like Session Initiation Protocol (SIP) which is an application layer protocol is applied to initiate and manage VoIP connections.



`Fig 1: Typical Architecture of VoIP Network

Compared to the conventional PSTN telephone network, the VoIP calls are easier to be threatened by eavesdroppers as the voice data packets of the VoIP networks are delivered and exposed to the unsecured public internet. It is usually quite difficult to tap a phone call in a circuit switched network at any place except at the final leg of the analog circuit. Moreover achieving end to end security in VoIP session is a demanding task. During this endure a combination of different protocols are involved in VoIP session establishment. All of these protocols must interoperate properly to provide a good voice quality

The VoIP protocol stack can be categorized into four layers as show in Fig 2, secure media transport, key exchange, session description and signaling [13]. Each of the above layers are implemented by a separate protocol. [6]SIP (Session Initiation protocol) is the commonly used signaling protocol in VoIP networks. In VoIP network for creating, modifying and

terminating sessions with one or more participants the Session Initiation protocol (SIP) is used which is an application layer protocol. In order to provide a cryptographically secure way of establishing secret session between the various participating entities in an insecure environment, various key exchange protocols like MIKEY, SDES, and ZRTP are used. Typically the key exchanged between the two or more participants are used in a symmetric encryption technique, the confidentiality of the voice data is dependent on the secrecy of the shared key between the legal communicating parties. The ultimate goal of the key secrecy is that nobody other than the communicating participants in that particular session are able to distinguish the key from a random bit string.

The confidentiality and integrity of the actual voice data stream which transmitted via the public internet is provided by the secure media transport layer. In the case of VoIP, this stream usually carries voice datagrams. Confidentiality means that the privacy of the encrypted voice data is maintained and nobody other than the two participants who have the specific key is able to decrypt the information of the voice data. Integrity of the voice data implies that any change in the content of the actual voice datagrams must be detected by the recipient. Secure Real Time Protocol (SRTP) which an extension of the RTP protocol is an example of the secure media transport protocol and which provides the functionalities of voice data confidentiality and integrity.

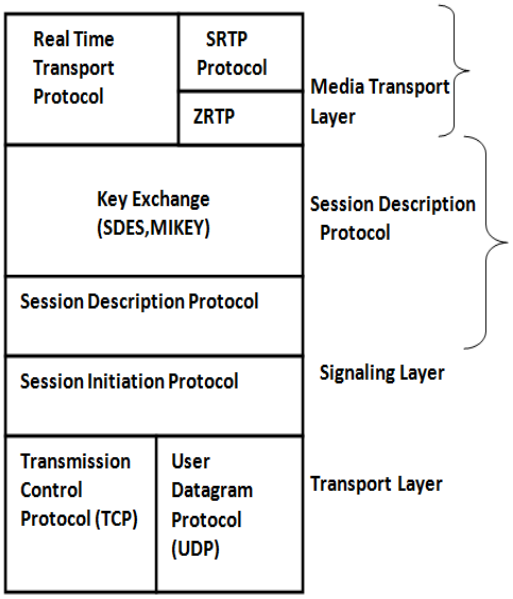


Fig 2: Voice-over-IP protocol stack

2. BACKGROUND

2.1 Session Initiation Protocol (SIP)

SIP [6] is an Internet protocol present at application layer, it is used to set up maintain and terminate multimedia sessions. SIP was designed to be a rendezvous protocol; i.e., given a user to locate, the protocol would use any means at its disposal to find the user. SIP is more flexible and less complicated compared to H.323 protocol.

2.1.1 Establishing a SIP session

A SIP ecosystem consists of proxy servers, user agents, redirect servers and registrars [10]. There are two types of SIP user agents: User Agent Client (UAC) and User Agent Server (UAS). These user agents are software programs that execute on s computer, an Internet phone or any devices. In order to initiate a call the User Agent Client sends an URL addressed INVITE to the intended recipient. A User Agent Server accepts the request and acts upon it. Typically the User Agent Servers register themselves with a registrar, this registration information is required by the SIP proxy servers to route the request to an appropriate UAS.

Proxy servers are SIP intermediaries that provide critical services like forking, authentication and routing. Upon the receipt of the incoming call request, the SIP proxy will determine how best it can route the request to a downstream UAS. The INVITE request is used to establish a SIP session; it can generate one or more responses. The status code of the responses indicates success or failure. The responses with the status code 1xx are termed provisional responses and are used to update the progress of the call. Responses with status code 2xx indicate success and the status code with higher number indicate failures. 2xx-6xx responses are termed as final responses and serve to complete the INVITE request. The INVITE request is forwarded by a proxy possibly through a chain of proxies until it gets to its destination. Finally the destination sends one or more provisional responses followed by exactly one final response.

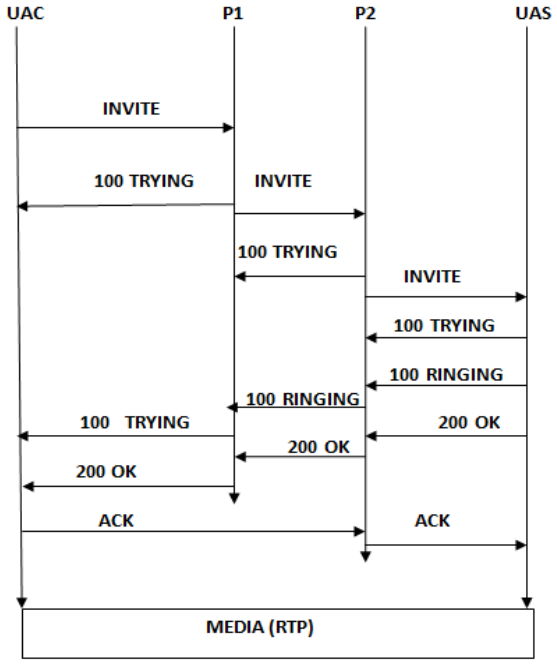


Fig 3: Session Establishment in SIP

Figure 3 provides a time line of call establishment between the User Agent Server and User Agent Client. The request from UAC is forwarded through a chain of proxies. With reference to Figure 3, the User Agent Client sends INVITE to P1 and P1 routes the call further downstream. With reference to User Agent Client's reference, P1 is called an outbound proxy. The

proxy server P1 determines that request should be forwarded to P2, further when the request arrives at proxy server P2 it queries it location server. From the User Agent Server point of view, P2 is an inbound proxy. The User Agent Server issues a provisional response followed by the final response.

2.2 RTP and SRTP

The media is transported end to end using Real Time Transport Protocol (RTP) [10].RTP exchanges packets in clear text, so there were many security issues as the confidentiality of the voice data was at serious risk, Therefore they developed a Secure RTP to provide message authentication, replay protection and confidentiality to the clear text RTP traffic .Conceptually, SRTP can be seen as a "bump in stack" implementation as shown in Figure 4, that exists between the RTP layer and the transport layer. SRTP intercepts RTP packets and then forwards that packet on the sender side and intercepts SRTP packets and passes an equivalent RTP packet up the stack on the receiver side[10].

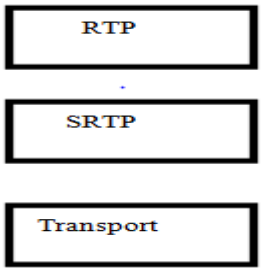


Fig 4: SRTP as "bump in stack"

In order to achieve the goals of confidentiality, replay protection and message authentication. SRTP defines extensions to the RP packet format to encrypt the RTP payload. The SRTP stream requires that the sender and receiver to maintain crypto-graphic state information. The cryptographic context provides all the necessary parameters such as the chosen cipher, its mode of operation and the block size, the master key and the session key etc. There are two types of keys used by SRTP one is the session key and other is the master key. The session key is used directly in the payload encryption or message authentication and the random bit string provided by the key management protocol from which the session keys are derived in a cryptographically secure manner .The salt key, master key and other parameters in the cryptographic context are provided in context by the various key management mechanisms such as SDES, ZRTP.

The cryptographic context is selected by a 32 bit numeric field called Synchronization Source (SSRC) which is carried in the fixed RTP header, which identifies the source of a RTP stream. Some of the key management protocols RTP in particular provide this to SRTP, while in the other cases SSRC is obtained dynamically when SRTP packet arrives at a receiver .Figure 4 depicts the primary components of SRTP framework and their relationship. Note that while the key management protocols such as ZRTP are able to agree on the master key, salt and other parameters independently at the peers, some amount of information to tie the media stream to the signaling channel to prevent from inserting false media packet can be provided by the signaling layer. To successfully accomplish this ZRTP computes a 256-bit hash across its initial handshake message and carries it as a=zrtp-hash SDP attribute.

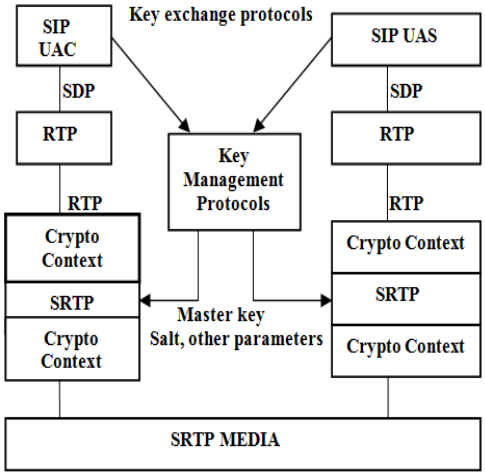


Fig 5: Components of SRTP framework

SRTP has been used in SIP increasing in recent times, however its wide spread adoption has been slow due to the inability of the various keying protocols to negotiate security contexts while preserving the semantics of certain SIP features .We should note that each SRTP stream requires the sender and the receiver to maintain a separate cryptographic context. Central to this context is a master key, essentially a secret random bit string provided to SRTP by an external key management protocol .Session keys, used in securing the communications, are in fact derived from the master key .Session keys which are typically derived from the master key are used for securing the confidentiality of the voice data .Settling on one specific key management protocol is still an challenging issue in VoIP.

3. EXISTING KEY MANAGEMENT TECHNIQUES

The key management protocols are the fundamental security mechanism for protecting the confidentiality of the voice data. Therefore in this part we describe and analyze the key exchange protocols which are specific to VoIP in detail. In the below section, we are mainly going to focus on the three frequently used key generation protocols i.e. ZRTP MIKEY and SDES. Moreover the key features of these protocols are analyzed and their immunity towards Man-In-The-Middle attack is also studied.

3.1 ZRTP

The ZRTP is a protocol which enables the two communicating parties to set up a shared secret, by which the confidentiality of the voice data is secured. In order to agree upon a common session key the Diffie-Hellman key exchange mechanism is used. As we know that Diffie-Hellman key exchange mechanism is inherently vulnerable to MITM attacks, ZRTP claims to be immune from such attacks by introducing some extra mechanism to detect such attacks, one of such mechanism is called as Short Authentication String (SAS) [1]. Short Authentication String is a group of characters that is derived from public values of Dif fie-Hellman key exchange. If the SAS

value on both the sender and the receiver side are equal, then there is a high chance of a MITM attack. Moreover ZRTP offers a feature for key continuity by caching the key material from previous session for future use.

Another notable feature of the ZRTP protocol is that it does not require a Public Key Infrastructure. This feature is a clear advantage of ZRTP as the user levels of the Public Key Infrastructure are very hard to maintain and here the users are free from the certification authority. If the certification authority is in use, then there is a need for each user to hold his or her unique digital certificate, moreover the operational cost of maintaining such an authority is high and it requires high cost of issuing, revoking and renewing user level certificates.

When ZRTP is used for SRTP session establishment, it can be categorized in to four phases. After the RTP session has been initialized using the Session Initiation Protocol, the ZRTP participants exchange information with each other in the discovery phase about the capability of their client. In the following phase which is called the commitment phase, the initiator commits himself to a set of parameters like the public value 'pv' of the Diffie-Hellman key exchange scheme. The commitment in this phase is done by computing a hash value of the particular message that is to be sent in the following key negotiation phase. In the key negotiation phase, the relevant parameters for the actual key exchange are shared with various participating entities. The Final phase is the confirmation phase where the process is completed; it ensures that the key derivation has been done successfully.

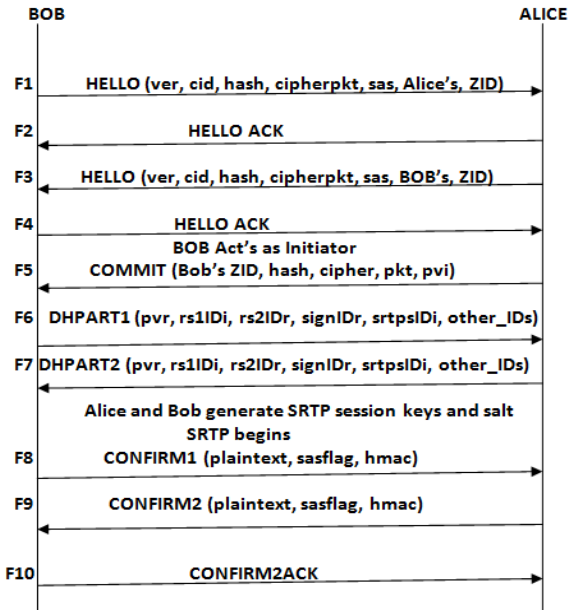


Fig 6: Establishment of SRTP session key using ZRTP

3.1.1 Key Features of ZRTP

3.1.1.1 Authentication

This is the feature of ZRTP by which it claims to be immune from MITM attacks. ZRTP uses a Short Authentication String (SAS) to avoid MITM attack, SAS is cryptographic hash value of the two Diffie-Hellman public values. This SAS value is generated on both the participants' side. To carry out authentication, both the participants read aloud the computed SAS value over the established voice connection. If the values on both the ends of the voice connection are equal, it indicates that the both the participating entities use the same key for encryption. If indeed there was a Man-in-The-Middle attack on the voice connection there would be different encryption keys in use on both ends, i.e. one between the initiator and the attacker and another between the responder and the attacker.

3.1.1.2 Key Continuity

A "baby duck security" model is used by ZRTP to authenticate the Diffie-Hellman key exchange process. Both clients participating in the voice communication process cache the shared secret key that is used in session and it uses that captured share key in the next session to derive a new Diffie-Helman value. This means that in order to carry out a successful MITM the attacker must be present in all the sessions starting from the first session, this process is considered to be difficult for the attacker.

3.1.2 MITM ATTACK ON ZRTP

Despite the above mentioned countermeasures to prevent a successful MITM, researchers have found that still it is possible to launch a successful MITM Attack on ZRTP [18]. In order to launch a successful MITM attack the attacker has to first get into the data path somehow .There are a number of different possibilities to achieve this .One approach is to carry out an ARP poisoning attack between the victim host Alice and the SIP Proxy server. For carrying out a successful MITM attack the attacker has to have access to the local area network, otherwise ARP cache poisoning mechanism would be difficult to carry out. Other scenarios favorable for a ARP cache poisoning is the use of a rouge WLAN access point or some similar techniques. For an attacker point of view, it is not mandatory for an attacker to insert a bogus component into a signaling path; it is conceivable to exploit vulnerability and to take over an existing SIP proxy server.

Now lets us discuss a number of possibilities as shown in Table 1, for the attacker how to relay the media data between the two communicating parties.

1. Direct Relay
2. Direct Relay +Initializing SAS
3. Repeating

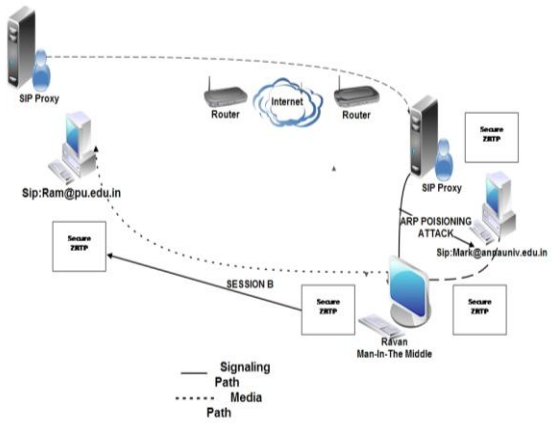


Fig 7: A Typical Man-in-the-middle Attack secanrio

Table 1: Various attack methodologies for carrying out Successful MITM attack

Method Name	Description	Advantage	Disadvantage
Direct Relay	The simplest method is to take the RTP payload from one session and feed into other session	1. There is a little delay 2. The interaction quality is normal. 3. Automation of attack is possible.	If SAS comparison is carried out there is a possibility of finding out the attack
Direct Relay +Initializing SAS	This method is similar to the above method ,but during the conversation the attacker replaces the SAS value which is exchanged with the one appropriate for a particular session	1. Liitle Delay 2. Normal Interaction quality	1. Independent of the applied method manual interaction is required. 2. Insertion of faked SAS comparison has to be time accurately.
Repeating	In this case the attacker listens to both participants almost concurrently repeats every single word for BOB .Only when one side utters the SAS to provide appropriate SAS value for each call	SAS comparison is not broken	1. Larger latency is observed due to delay when repeating spoken words. 2. Bad interaction quality.

3.2 SDES

SDES is the simplest form of key management protocols to understand. SDES is the key transport extension of the SDP protocol. SDES defines a new SDP attribute called "crypto" which is used to signal and negotiate cryptographic parameters for SRTP media streams[10]. This attribute transports the encryption and the authentication algorithms, master key and salt of the sender (i.e. the receiver should use the said master key and salt to derive session keys for decryption .The crypto attribute for SRTP is defined as: a= crypto :(tag) (crypto-suite) (key-params) [(session-params)]. The most important component is key-params, which specifies one or more cryptographic keys as (key-info).

In this simplest form, the UAC inserts this parameter in the SDP of the INVITE request and send it to UAS; the UAS inserts this parameter in the 200 OK responses and transmits it to the UAC. Consequently SDES provides unique keys for each media stream in each direction. The example below shows "crypto" attribute in INVITE from Samuel to Mark :

INVITE sip : <mark@pu.edu.in> SIP/2.0

To: Mark<sip:samuel@pu.edu.in>

From:Samuel <sip:mark@annauniv.edu.in>;
Tag=oij8z
Via: SIP /2.0/UDP a.example.org;
Branch =z9hG4bKnash
CSeq: 48884 invite
Call-ID : 7823299@pu.ed.in
Content-type: application/sdp
V=0
O=alice 2890844526 2890844526 IN IP4
a.example.org
C= IN IP4 192.0.2.101
T=0 0
M=audio 49170 RTP/SAVP 0
A= crypto: 1 AES_CM_128_HMAC_SHA1_80
Inline: NZbd2n8jsy8nhdu8nh899jhj99+jdjj9jjs | 2^20

The crypto attribute above identifies the encryption and authentication algorithm and specifies the master key, salt and the lifetime of the master key (2^20). The master key and salt are concatenated and base 64 encoded .The sender of the "crypto" attribute uses the master key to derive the session key for encryption and the receiver uses it to derive the session key for decryption

A malicious eavesdropper can gain access to the master key, if the SIP request or response carrying the "crypto" attribute is transmitted in the clear. Thus there is a need for the cryptographic keys and other parameters should be secured pn a hop-by-hop link using TLS. By this mechanism the unauthorized eavesdroppers from sniffing of the cryptographic keys are prevented, though it does not afford complete privacy or confidentiality to the media. Session, because the intermediaries at the end of the hop-by-hop TLS link will have access to clear text cryptographic keys.

3.2.1 Man-In-The-Middle Attack in SDES

SDES suffers a serious problem from Man-in-the-Middle attack. Consider a scenario where an adversary is able to inject itself as a next hop in the intermediary chain; it will have complete access to the cryptographic parameters. From this point of view, SDES may be considered the least secure of the key exchange protocols. It should be noted that the use of the TLS-secured channel across the intermediary chain does not guarantee secure and private delivery of session keying material .The adversary may be able to obtain a legitimate certificate from a certificate authority and then insert itself in the intermediary chain by using techniques such as DNS cache poisoning.

3.3 MIKEY

SDE MIKEY is an another key management protocol which is used in VoIP .The Multimedia Internet Keying [RFC 3830] was designed to meet the requirements of initiation of secure multimedia sessions [13]. In MIKEY the parameters for the security protocol should be exchanged in one round trip. It was designed to make the means of key exchange simpler and straight forward. The MIKEY protocols provides end-to end security for the keying material and moreover it is independent from any specific security functionality of the underlying transport .Another notable feature of MIKEY is that it consumes low bandwidth consumption and low computational workload.

MIKEY supports three types of key agreements

1. Pre-shared key (PSK) – In this method the key derivation for both the encryption and the integrity of the message purpose the pre-shared secret is used and. The randomly generated TGK of the MIKEY message is securely transported. This key agreement mode is considered to the most cost effective scheme but the drawback of this scheme is that the shared key distribution leads to scalability issues.
2. Public-Key Encryption (PKE) - This method is mostly similar to the above method , the only difference is that here the initiator makes use of a random key for encryption and integrity.
3. Diffie-Hellman (DH) - This method provides perfect forward secrecy. This method can be used in peer-to-peer keys, but this method proves to be resource consuming. Further for the purpose of a message signing the existence of PKI is a must.

3.3.1 Man-In-The-Middle Attack in MIKEY

The ultimate aim of the cryptographic secure key exchange protocol is to establish a session key which is indistinguishable from a random bit string by anyone other than the participants [13]. It is notable that when MIKEY is executed in Diffie-Hellman mode it is vulnerable to the Man-In-The-Middle attack. The shared key that is derived using the joint Diffie-Hellman value is used directly as the key. But researches have proved that the key value derived from the joint Diffie-Hellman vales does not provably produce an output which is indistinguishable from a random bit string. Finally, we observer that MIKEY used in the pre-shared key mode obviously doesn't satisfy perfect forward secrecy because the compromise of the pre-shared secrets leads to compromise of all previous sessions. Another drawback of the MIKEY is that it provides very limited protection against Denial of Service attacks because performing a large number of digest verifications can exhaust memory resources.

Table 2: Feature Set Evaluation of ZRTP, SDES, MIKEY

Feature	**ZRTP**	**SDES**	**MIKEY**
MAN –IN-THE MIDDLE ATTACK	Attack is possible	Attack is possible	Attack is possible
Forking	Key Leakage occurs	Key Leakage does not occurs	Key Leakage does not occurs
Conferencing	Not supported	Not Supported	Not supported

4. CONCLUSION

In this paper the key-generation and key management protocols used in VoIP have been evaluated .We have come to notice that the three key generation protocols ZRTP, SDES and MIKEY are vulnerable to the Man-In-The Middle attack. Our analysis suggests that the key management protocols that operate in the media layer are indeed suitable media keying protocols despite their operational differences. It is very important to note that the, key management protocols are the fundamental security mechanism for protecting the confidentiality of the voice data. Therefore we feel that there is a need to find an alternate secure key generation protocol for Mobile VoIP where the bandwidth and the resource consumption should be very minimal. Currently ZRTP is used in mobile VoIP which has been found vulnerable to Man-In-The Middle attacks and more over the bandwidth consumption of ZRTP is more. Therefore there is a growing need for a new key management protocol for Mobile oriented VoIP devices given that it is immune to all kinds of attacks. In our future work we are planning to implement an ECC based key management protocol for Mobile SIP VoIP devices, as ECC has inherently faster computing power and demands only smaller storage space in the same security level than any other PKI (Public Key Cryptography System).

5. REFERENCES

[1] P. Zimmermann, A. Johnston, and J. Callas, "ZRTP: Media Path Key Agreement for Secure RTP", draft-zimmermann-avt-zrtp-09 (Internet- Draft), Internet Engineering Task Force, September 2008.

[2] J. Rosenberg, H. Schulzrinne, G. Camarillo, A. Johnston, J. Peterson, R. Sparks, M. Handley, and E. Schooler, "SIP: Session initiation protocol," RFC 3261 (Proposed Standard), June 2002. [Online]. Available: http://www.ietf.org/rfc/rfc3261.txt

[3] M. Baugher, D. McGrew, M. Naslund, E. Carrara, and K. Norrman, "The Secure Real-time Transport Protocol (SRTP)," RFC 3711 (Proposed Standard), Mar. 2004. [Online]. Available: http://www.ietf.org/rfc/rfc3711.txt

[4] M. Petraschek, T. Hoeher, O. Jung, H. Hlavacs, and W. Gansterer, "A man-in-the-middle attack on ZRTP," in Wor--ks hop on Socio-Economic Issues of NGI, Santander, Spain, 2007.

[5] P. Zimmermann, A. Johnston, and J. Callas, "ZRTP: Media Path Key Agreement for Secure RTP," draft-zimmermann-avt-zrtp-09 (Internet-Draft), Internet Engineering Task Force, September 2008.

[6] Butcher, D.; Xiangyang Li; Jinhua Guo, "Security Challenge and Defense in VoIP Infrastructures", Publication Year 2007

[7] Chia-Hui Wang a ,n , Yu-Shun Liu b Publication "A dependable privacy protection for end-to-end VoIP via Elliptic-Curve Diffie-Hellman and dynamic key changes", Year :2010.

[8] Walsh TJ, Kuhn DR, "Challenges in securing voice over IP". Security & Privacy, IEEE Magazine 2005.

[9] Kokkonen, E.; Matuszewski, " Peer-to-Peer Security for Mobile Real-Time Communications with ZRTP" Publication Year: 2008.

[10] D. Wing, S. Fries, H. Tschofenig, and F. Audet, "Requirements and analysis of media security management protocols," RFC 5479 (Infor-mational), 2009.

[11] Gurbani, V.K.; Kalashnikov, "A Survey and Analysis of Media Keying Techniques in the Session Initiation Protocol (SIP)", Publication Year: 2011.

[12] R. Bresciani, "The ZRTP protocol: Analysis on the diffie-hellman mode," Computer Science Department Technical Report TCD-CS-2009-13, Trinity College Dublin, 2009. [Online]. Available: http://zfoneproject.com/docs/TCD-CS-2009-13.pd

[13] P. Gupta and V. Shmatikov, "Security analysis of voice-over-IP proto-cols," in Proc. Computer. Security Foundations Symp.. IEEE, July 2007,pp. 49–63.

[14] M. Bellare and P. Rogaway, "Entity authentication and key distribution," in Advances in Cryptology – CRYPTO 93, ser. LNCS, vol. 773. New York, NY, USA: Springer-Verlag, 1994, pp. 232–249

[15] A. Datta, A. Derek, J. Mitchell, and D. Pavlovic. "A derivation system and compositional logic for security protocols". J. Computer Security, 13(3):423–482, 2005.

[16] F. Andreasen, M. Baugher, and D. Wing. "Session Description Protocol (SDP) Security Descriptions for Media Streams". IETF RFC 4568, July 2006.

[17] P. Zimmermann., "ZRTP: Extensions to RTP for Diffie-Hellman Key Agreement for SRTP". http://www1.tools.ietf.org/html/draft- zimmermann- avt-zrtp- 01 , March 2006.

[18] Jung, O.; Petraschek, M.; Hoeher, T.; Gojmerac, I. "Using SIP identity to prevent man-in-the-middle attacks on ZRTP" Publication Year: 2008.