ELSEVIER

Computer Communications 26 (2003) 552–565

computer communications

www.elsevier.com/locate/comcom

# Performance analysis of secure session initiation protocol based VoIP networks

Mohan Krishna Ranganathan*, Liam Kilmartin

*Communication and Signal Processing Research Unit, Department of Electronic Engineering, National University of Ireland, University Road, Galway, Ireland*

Received 19 February 2002; revised 18 June 2002; accepted 25 June 2002

**Abstract**

The commercial deployment of voice over internet protocol (VoIP) networks (and associated packet switching technologies) has gathered pace in the recent years. However, a major concern with such networks is the issue of the security of networks based on such open standards. Little research has been carried out into examining the options for securing VoIP networks and, more specifically, the impact which implementing such security architectures and protocols will have on the performance of such secure networks. This paper describes the research, which has been carried out into the development of a realistic model for carrying out simulations of the performance of secure session initiation protocol based VoIP networks. The results of the performance analysis obtained using this model are presented with a discussion of the implications of these results for designers considering implementation of real secure VoIP networks.


*Keywords:* Voice over internet protocol; Security; IP Security; Security protocol performance analysis

## 1. Introduction

> Starting as a hobbyist movement five years ago, "Voice over Internet Protocol" is quietly remaking the telephone system worldwide. It is one of the venerable network's biggest overhauls in decades—but not its last by a long way.
>
> The Economist, March 2001.

The recent years has seen the growth of internet protocol (IP) based networks (e.g. Internet) at a thriving pace. The rapid proliferation and ubiquitous nature of the Internet, for example, has now given rise to strong interest in using IP based networks for carrying non-conventional information like the voice, multimedia, etc. The use of the Internet as a transport network for speech signals is currently in its infancy. The sharing of existing network infrastructure between data applications and voice calls, and the sharing of access and transport services helps in reducing implementation, management and support costs. This also provides an opportunity for new services and applications, which were not feasible with traditional circuit-switched telephony networks, to be developed. Even with all these benefits, wide spread commercial deployment of voice over IP (VoIP) is still restricted [1] due to the challenges posed by the nature of the Internet. However, it is widely accepted that next generation networks will use the Internet Protocol, or some variant thereof, as the networking protocol of choice for supporting multimedia traffic, and voice traffic in particular.

There remains a great deal of research, which still needs to be carried out into the particular problems which need to be solved for VoIP networks to be a technical and commercial success. The non-deterministic nature of the Internet, and the impact, which this specifically has on voice traffic, is one major area of concern. Inherent problems with security due to the 'open' nature of public IP networks are also of equal importance. This paper focuses on the challenges and impact of employing security services into VoIP networks. The security requirement considerations of VoIP networks are highlighted along with the available security service options for the different VoIP architectures. A simulation model of an IPSec secured session initiation protocol (SIP) based VoIP network is presented along with a discussion of the simulated network performance as

* Corresponding author. Tel.: +353-91-750326; fax: +353-91-750511.
*E-mail address:* mohan.krishna@nuigalway.ie (M.K. Ranganathan).

0140-3664/03/$ - see front matter 
PII: S0140-3664(02)00146-9

obtained from this model. A number of implications for real secure network designers and operators arising from this research are highlighted.

## 2. VoIP architectures

Unlike the circuit-switched PBX scenario, the IP world is dominated by open systems, and hence the need for *standards* to ensure interoperability between devices manufactured by various vendors. Currently, many organisations are in the process of developing standards for session signalling over packet based networks, and the choice for VoIP vendors and manufacturers would be to select the architecture that would support emerging trends.

H.323 [2], adopted in 1996 by the international telecommunications union (ITU), was the first call control standard developed for VoIP. H.323 is an umbrella recommendation suite, which defines audio, video and data communications across local area networks (LANs) that do not guarantee quality of service (QoS).

Even though H.323 was the widely deployed signalling protocol for VoIP, a newer signalling protocol, SIP has gained significant momentum [3] recently as an alternative to H.323, mainly due to its simplicity and efficiency.

The SIP [4,5] is a generic application layer session management protocol, developed by the internet engineering task force (IETF) multi-party multimedia session control (MMUSIC) working group and was standardised in 1999.

SIP provides for advanced signalling and control functionality for a wide variety of multimedia services including VoIP. The syntax and semantics of SIP are heavily borrowed from the popular HTTP and SIP works on the same request–response model as in HTTP. The functionality of SIP is similar to telephony signalling protocols, such as Q.931 [6], but only in an Internet context. SIP also differs from the traditional telephony signalling protocols, in that it does not reserve resources or establish circuits (virtual or real) in the network.

### 2.1. The VoIP protocol stack big picture

As seen in Fig. 1, there are two main aspects of VoIP (1) the call signalling and call controlling information and (2) the media (speech) information. The protocol stack defines the method of carrying both the signalling and media information.

The well-established VoIP protocol stack can support a variety of underlying network types (typically LAN standards) below the network layer. A VoIP terminal, connected to such networks, has traditionally been a PC equipped with audio peripherals (i.e. speakers and microphones) but many networking manufacturers are now supplying standalone VoIP terminals. The internetworking protocol (IP) networking layer operates above whichever networking technology is in operation. The user datagram protocol (UDP) operates in the transport layer in order to provide a suitable end-to-end protocol for this type of multimedia application. UDP does not, however, adequately support some of the needs of real-time audio being transported over an IP network. Hence, a companion transport layer protocol operates above UDP to provide specific support required by such real-time multimedia applications as VoIP. This additional transport layer protocol suite actually consists of two protocols, namely the real-time transport protocol (RTP) [7] and the real-time transport control protocol (RTCP).

The application layer effectively implements whichever audio or speech codec is in use by the VoIP terminal (e.g. G.711, G.721, G.728, etc.), and uses the RTP layer to transport the media stream. The application layer also implements the call signalling and control protocol to establish, control, and terminate VoIP calls, as well as to invoke any of the multitude of supplementary services that can be supported by a VoIP network (e.g. SIP, H.323).

## 3. VoIP security requirements

As every VoIP network is essentially an IP network, VoIP network and terminals face the same security threats

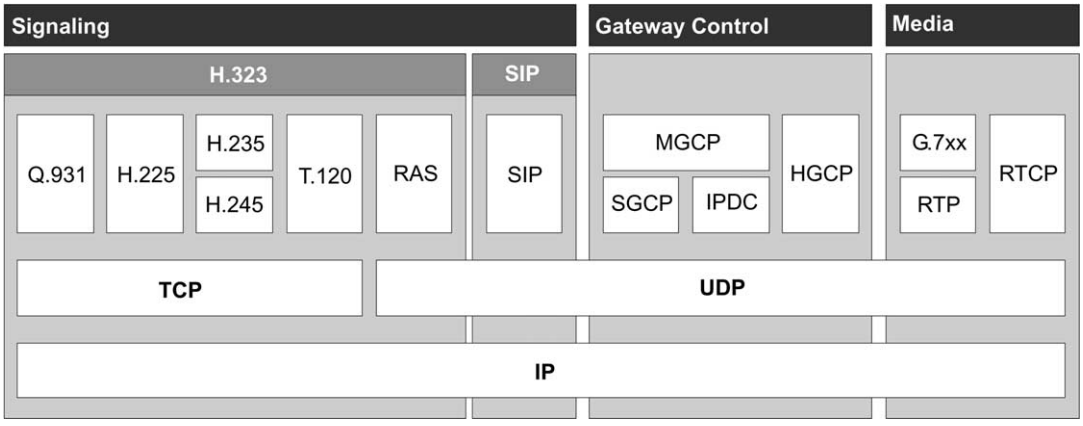


Fig. 1. VoIP protocol stack.

inherent with any IP network. For example, the media (RTP) packets of an ongoing VoIP call on a LAN could be easily picked up and recorded by a simple packet sniffer.

In naive terms, VoIP calls need to be at least as secure as circuit-switched calls with respect to anonymity and privacy. With added security threats because of the open nature of the underlying IP network, the functional VoIP security requirements could be stated as:

(1) Protection of privacy of the call conversation.
(2) Authentication of call end entities.
(3) Protection from misuse of network resources, or in other words, access control by the service provider.
(4) Ensuring correct billing by the service provider, and protecting billing information from unauthorised access.
(5) Protection of caller behaviour or statistical information from unauthorised access.
(6) Protection of network servers and terminals from well-known threats, such as 'denial of service' and 'man in the middle attacks'.

In essence both the media stream and the signalling stream of a VoIP call must be protected from unauthorised access.

Most of the security requirements highlighted above are not specific to VoIP alone, but are general requirements of any IP network wishing to protect the interests of its end entities. However, unless the underlying network is secure, either the VoIP call is not secure, or it has to have its own security features implemented (e.g. application layer security).

### 3.1. Available security options for VoIP

A significant amount of research and development has been carried out into the issue of providing security protocols for specific applications on IP networks (e.g. secure socket layer (SSL), which is primarily used in eCommerce applications and pretty good privacy (PGP), which is primarily used for email). These already developed security protocols could be employed for VoIP as well. However, the characteristics and resource requirements of securing voice traffic (and indeed multimedia traffic in general) are distinctly different from traditional IP traffic. In general, the available options for VoIP security are,

(1) Integrating the core security mechanisms of authentication and encryption, into the VoIP protocols itself.
(2) Use existing application layer security protocols (e.g. PGP), or similar for providing security services to VoIP signalling and media streams.
(3) Generic transport layer security protocols, like SSL/TLS could be employed.
(4) More flexible and scalable alternative is to carry VoIP over 'secured' networks, i.e. use security services built into the network layer (IPSec).

#### 3.1.1. SIP security

Incorporating security services into SIP signalling [8] is under consideration by the IETF SIP working group and is still under development. As SIP borrows heavily from HTTP for its messaging syntax, it can also employ the challenge–response authentication model used by HTTP. Ref. [8] suggests the use of basic or digest authentication mechanisms. In addition, authentication using the CHAP (challenge–response) mechanism is proposed. Encryption of SIP message bodies and some of the headers is also suggested, but some of the headers have to be in clear text. The other choice would be use to PGP for SIP message encryption.

For encrypting the media stream, session keys could be exchanged as part of the session description protocol (SDP), but then this would require the SIP signalling messages to be encrypted. In addition, Ref. [8] suggests the use of developed security protocol, either IPSec or RTP Security, for this purpose.

### 3.2. Restrictions due to VoIP dynamics

Real-time applications, such as VoIP are highly sensitive to network packet delays and delay jitter. The acceptable one-way delay for approximate toll quality speech is no more than 150 ms [9]. Codec delay, serialisation delay, queuing delay, propagation delay, etc. all contribute to the overall network packet delays, which could vary dynamically depending on the network dynamics.

Encryption, authentication and key exchange algorithms are by nature computationally intensive and when employed for VoIP only add up to the overall packet delay. More specifically, public key encryption (used for key exchange) requires very high computation power. A choice has to be made such that the security protocol overhead does not affect the packet processing delay beyond acceptable limits with the associated negative impact on the transmitted voice quality. One option is to employ end-to-end encryption, so that the computation power requirement is widely distributed. However, problem arises with system management, key exchange management and investment and maintenance costs. The other option is to employ security services at edge routers, gateways, etc. In such a case, when the network throughput is very high, the edge routers/gateways could get clogged with large number of queued packets and that would only add higher delays to the overall network delay.

## 4. Overview of SIP

SIP is a simple textual based session management protocol developed by the IETF. The main purpose of SIP

Table 1
SIP request messages

| SIP request type | Purpose |
|---|---|
| INVITE | Used to request a callee to join a particular session, or establish a two-party conversation |
| ACK | To confirm that a caller has received a final response for the INVITE message sent by it |
| OPTIONS | Server being queried about capabilities |
| BYE | Client indicates to the server to release the call |
| CANCEL | Cancels a pending request |
| REGISTER | Client registers address with a SIP server |

is to initiate, modify and terminate sessions between two (or more) Internet end entities. SIP itself is independent of the type or characteristics of the session and handles the session description as an opaque body. The actual session description is handled by a companion protocol called the SDP [10].

### 4.1. SIP architecture

SIP identifies four types of logical entities as the network session participants: the user agents, registrars, proxy servers and redirect servers.

*The user agent* consists of two entities, the user agent client (UAC) and the user agent server (UAS). The UAC is the caller application that initiates and sends SIP requests and is the final destination of a call. The UAS receives and responds to SIP requests on behalf of the clients, accepts, redirects and refuses calls.
*Registrars* keep track of users within their assigned network domain. A registrar is typically co-located with a proxy or redirect server and may provide location services.
*Proxy servers* are application layer routers that forward SIP requests and responses. They interpret the SIP requests and determine the next SIP server/terminal to which the SIP request has to be forwarded to.
*Redirect servers* map the destination address of a received SIP request into one or more new addresses, and return these new addresses to the client. In contrast to proxy servers, redirect servers do not route/forward the SIP request to the next hop.

SIP also uses another entity called the location server, which provides information about the caller's possible locations to proxy and redirect servers. The location server may be co-located with a SIP server. In a typical SIP session, the user agent initiates the SIP messages, which traverse one or more proxy servers before it reaches the destination user agent. The proxy server may interact with the location server to get the routing information. Typically, a user agent may decide to send all requests to a fixed, local outbound proxy server.

### 4.2. SIP signalling

SIP borrows most of its interface (and its advantages) from the well-known HTTP, which is used for web pages. Hence, SIP uses simple textual commands to setup and tear down call sessions. The SIP end entities/users are identified by e-mail like identifiers or telephone numbers, which are assigned while registering with the network or telephony service provider. The SIP URI (or address) takes a form similar to a mailto or telnet URL, i.e. user@host. The user part could be a user name or a phone number. The host part is either a domain name or a numeric network address. All SIP messages will carry the destination URI, which will be looked up by the proxy servers on route to make the routing decision. SIP messages are mainly request methods and responses, in plain text format. SIP identifies six request messages, INVITE, ACK, OPTIONS, BYE, CANCEL and REGISTER. Table 1 describes the purpose of each of these request messages.

The SIP response messages, as shown in Table 2, are identified by a numeric status code and are used to indicate the response for a received request method. The messages are classified into different status code ranges based on the purpose of the response message.

### 4.3. SIP call flow

Fig. 2 shows a typical SIP call setup flow involving a proxy server. It also highlights the usage of the common SIP request and response messages. SIP signalling is independent of the underlying transport protocol, i.e. SIP can be transmitted over UDP or TCP. Generally, UDP is preferred since it avoids the TCP connection setup and tear down overhead. As SIP messages can be transmitted over unreliable transport protocols like UDP, SIP has to take care of reliability on its own. SIP specifies the method of

Table 2
SIP response numeric codes

| | |
|---|---|
| 1xx | Informational |
| 2xx | Success |
| 3xx | Redirection |
| 4xx | Client error |
| 5xx | Server error |
| 6xx | Global failure |

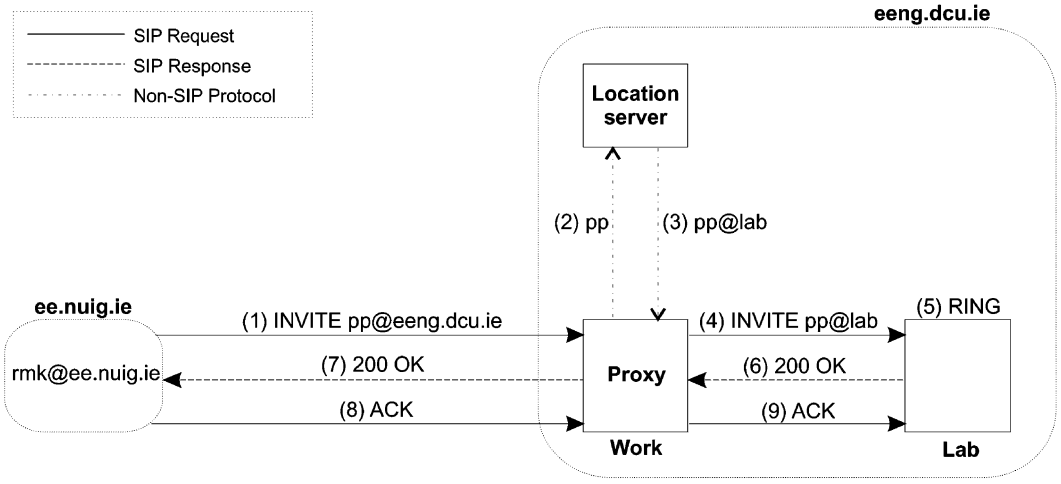


Fig. 2. SIP call signalling flow.

retransmitting messages with an exponentially increasing time gap between successive retransmissions. The retransmission ceases when a definitive response is received or if the number of retransmissions crosses a preset limit.

## 5. Overview of IPSec

IPSec [11] stands for 'IP Security', which was standardised by the IETF to address the security problems inherent in IP. IPSec defines a security framework architecture, using a collection of protocols to provide security at the network layer for any application. IPSec has been made mandatory for IPv6 and is optional for IPv4. As IPSec operates at the IP layer, every packet of information passing through the IP layer can be secured. IPSec was designed to be independent of underlying network topologies and provides transparent services to the application layer. In addition, IPSec is independent of cryptographic algorithms. This allows for seamless integration of new and stronger encryption algorithms into the IPSec architecture.

IPSec defines three technologies authentication header (AH), encapsulation security protocol (ESP) and internet key exchange (IKE) to provide the required security services, such as data origin authentication, confidentiality, integrity, protection against replay and limited traffic flow confidentiality. AH and ESP are the two traffic security protocols and IKE is used for session key management. IPSec defines the concept of security association (SA), which is a contract between two communicating entities. SAs are defined by various security parameters like the IPSec protocols to be used, the transforms, keys and validity of keys to name a few. IPSec implementations rely on two databases called the security association database (SAD) and the security policy database (SPD), for determining the security services to be provided to every packet. The SPD selector fields define the granularity of packet selection for applying security services. For example, all packets to a particular subnet could be set to be encrypted using the stronger triple DES, or packets to/from a particular host destined for a particular port could be set to be completely discarded.

### 5.1. Authentication header overview

AH [12] defines a mechanism for providing cryptographic authentication to IP (v4 and v6) datagrams. The authentication data are computed by using any of the standard message digest algorithms, such as HMAC-MD5 [13] and HMAC-SHA, and the computed authentication data are appended to the IP datagram. Note that AH does not provide confidentiality service to the IP datagram. AH can be operated in any of the two modes, transport mode or tunnel mode. In transport mode, the authentication service is provided to the upper layer protocol (transport layer) only and the AH is inserted immediately prior to the transport layer protocol header. In tunnel mode, the authentication service is provided to the whole of the IP packet. The whole of the authenticated IP packet is inserted as the payload of the AH and a new IP header is affixed before the AH. Tunnel mode implementation is generally used at gateway implementation, where packets from all hosts behind a gateway are offered the same security service.

### 5.2. Encapsulating security protocol overview

ESP [14] offers a mix of security services, which includes authentication and confidentiality. ESP can be applied alone, or in combination with IP AH or in a nested fashion (i.e. using a tunnel mode, Fig. 3). ESP can be configured to provide confidentiality alone, or authentication alone, or both. Encryption algorithms, such as DES in CBC mode [15], CAST and IDEA can be used with ESP for providing confidentiality. For authentication, message digest algorithms, such as HMAC-MD5 and HMAC-SHA can be employed.

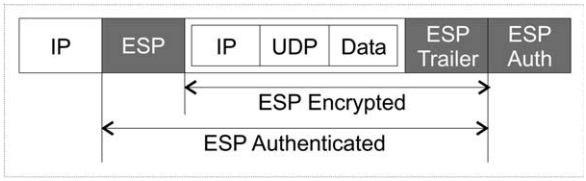


Fig. 3. ESP tunnel mode packet structure.

### 5.3. Security association

For correct encapsulation and decapsulation of IPSec packets, it is necessary to have a way to associate security services and a key, with the traffic to be protected, and the remote peer with whom IPSec traffic is being exchanged. Such a construct is called a SA [11]. IPSec SAs reside in the SAD. The SPD in turn, specifies the policies that determine the disposition of all IP inbound and outbound traffic.

The SPD has to be referred for every IP packet, and the policy determines whether the packet should be allowed to bypass IPSec processing, or should be discarded, or should undergo IPSec processing. If the policy dictates that the packet has to be IPSec processed, then the SPD entry points to one or more SAs (SAD entries), which have to be applied for the packet.

### 5.4. Internet key exchange overview

IKE [16] is used to establish the security parameters and the authenticated keys, in other words, SAs, between IPSec entities. IKE operates in a framework defined by the internet security association and key management protocol (ISAKMP) [17], which defines the packet formats and message construction requirements. IKE uses the ISAKMP grammar to establish a shared, authenticated key between the communicating entities. Diffie-Hellman public key exchange mechanism is used for the purpose of negotiating the shared secret key.

## 6. Secure VoIP network simulation model

Currently, there is no single VoIP signalling protocol, which has been exclusively adopted by the networking community. However, it is widely accepted that the SIP has a number of distinct advantages over H.323 and MEGACO, most notably its simplicity. Additionally, the IPSec framework of security protocols and architectures offer a myriad of flexible options when it comes to secure a VoIP network. For this reason, SIP and IPSec were chosen as the signalling and security protocols/architecture of choice on which the

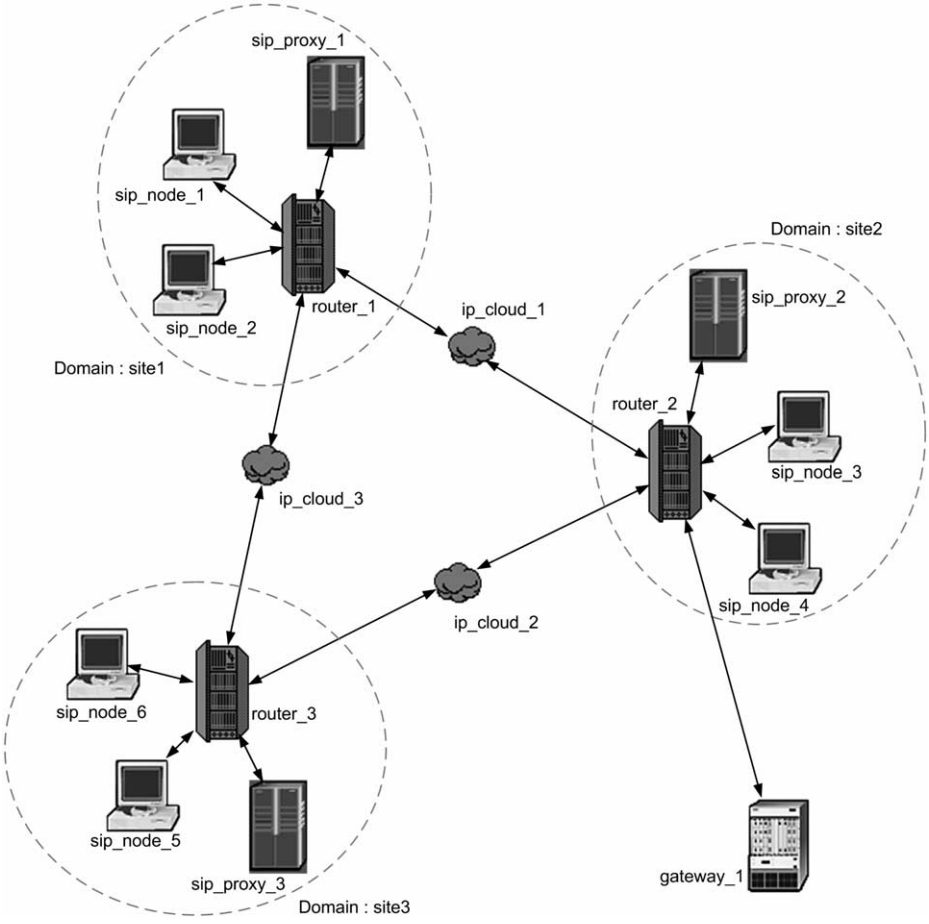


Fig. 4. SIP–VoIP network model topology.

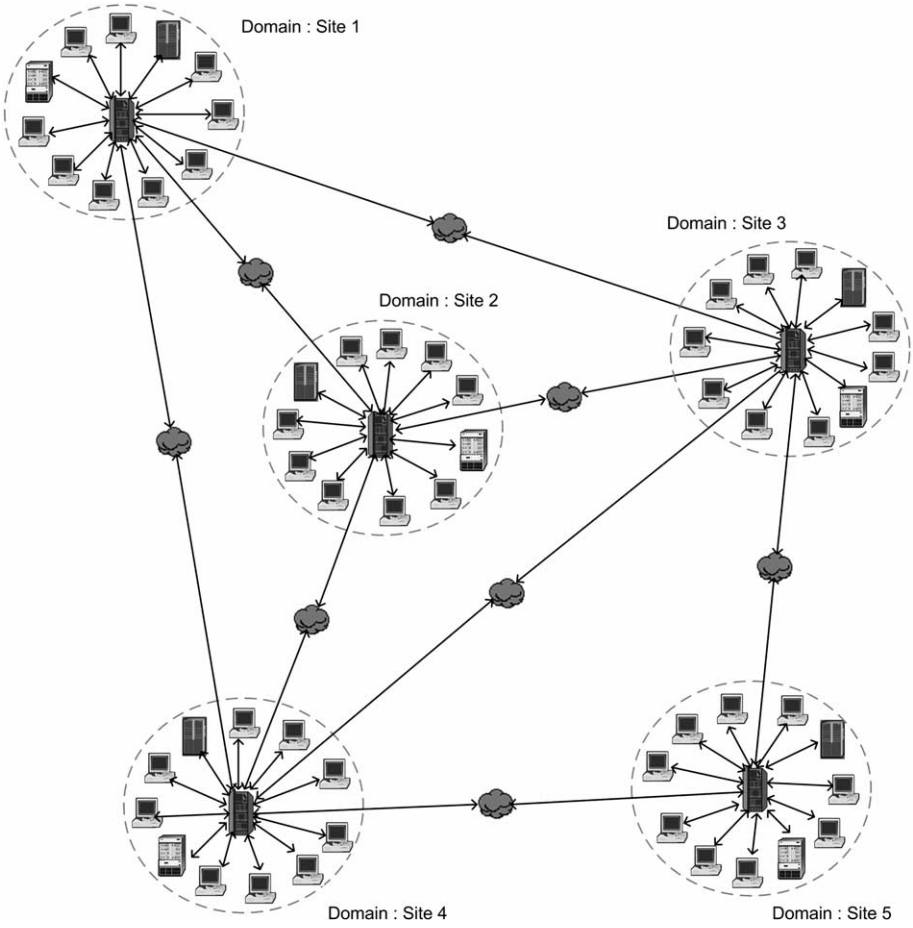


Fig. 5. Higher complexity SIP–VoIP network model topology.

simulation model for the secure VoIP network was based. The simulation model was developed using the OPNET Modeler® network simulation tool.

### 6.1. Network topology

Fig. 4 shows the reference structure of the SIP–VoIP network model built, emphasising the basic network building blocks and their organisation. The shown network topology consists of three different administrative SIP domains, 'site1' to 'site3', every pair of which is interconnected by an IP router. The SIP proxy server with a co-located location server acts as the functional core within each SIP domain. The location server at each proxy server is configured with the location details, i.e. IP addresses of all the SIP nodes within the local domain and also the IP addresses of all the other SIP proxy servers in the entire network neighbourhood.

The SIP nodes are assigned SIP user identifiers of the form id@domain, and all SIP nodes within a domain are configured to transmit their SIP messages to the local SIP proxy server. The SIP proxy server is entirely responsible for routing the SIP messages appropriately. If a SIP message is addressed to a SIP node within the domain, the proxy server routes it to the corresponding host machine. If the incoming SIP message is addressed to a SIP node in a different domain, the message is forwarded to the proxy server of the destined domain.

The network topology shown in Fig. 4 is solely for the clarity of illustration of the network elements and their organisation. The model can be easily scaled to obtain more complex network architectures, as shown in Fig. 5. The results presented in this paper were obtained with simulating networks having five SIP domains with 10 nodes and one VoIP–PSTN gateway in each domain.

### 6.2. Network components overview

The main components that build up the network model are the SIP nodes, SIP proxy servers, VoIP–PSTN gateways, IP routers and IP cloud models. The VoIP protocol stack within the SIP related nodes is implemented over a UDP/IP infrastructure. Figs. 6 and 7 represent the protocol stack implemented at SIP nodes and SIP proxy servers, respectively. Proprietary OPNET process models were developed for all the layers in the stack except for the physical layer, for which standard OPNET process models were used.

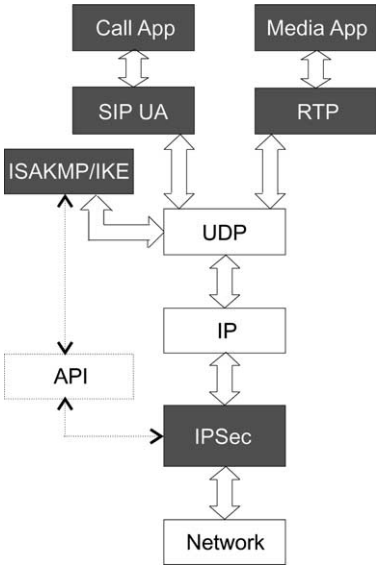


Fig. 6. SIP node protocol stack.

### 6.2.1. Call application layer model

The call application is the initiator of the voice call and also the call terminator at the destination host. The call application layer interacts with the SIP layer for call setup and disconnection and also directs the media application layer for generation and transmission of voice media.

The configurable simulation parameters at the call application layer are the total number of calls to simulate and their occurrence distribution. The specified number of calls is spread over the entire simulation duration, typically with an exponentially varying intercall occurrence interval. The source and destination for each call are picked up randomly from the entire list of SIP nodes in the network.

### 6.2.2. Media application layer model

The media application layer is responsible for source and sink of the voice call media stream. At the source end, once the SIP connection is setup, the media application generates G.711 media packets at a regular interval of 20 ms during talkspurts and no packets are generated during silence periods. The distribution of lengths of the talkspurt and silence periods are based on the model suggested by Sriram and Whitt [18], i.e. talkspurt lengths are exponentially distributed with a mean of 352 ms and silence lengths are exponentially distributed with a mean of 650 ms. The duration of the entire call itself is distributed exponentially with a configurable mean.

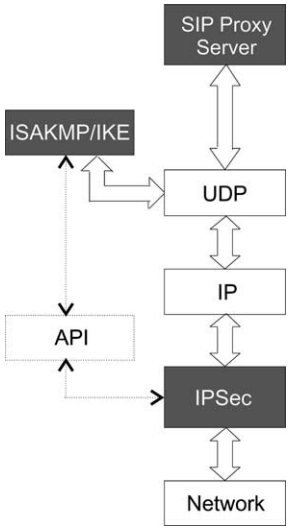


Fig. 7. SIP proxy server protocol stack.

### 6.2.3. SIP layer process models

The SIP layer models include the SIP user agent (UA) at the SIP nodes and the proxy server process model at the SIP proxy server. The implementations are based on a subset of recommendations from Ref. [5], required to setup and tear down call connections within a network architecture as described in Section 6.1. The configurable parameters of the SIP UA process are the SIP user ID, the name of the SIP domain it belongs to and the IP address of the local SIP proxy server. The SIP proxy server process model requires the domain name and the location server details to be configured.

### 6.2.4. RTP process models

The RTP process model implementation is based on the recommendations from Ref. [7] and helpful guidelines from Ref. [19]. The RTP process model provides services, such as timing reconstruction, loss detection and content identification of media stream. RTP utilises the sequence number and timestamp fields in the RTP header to provide these services.

The main functionality of RTP at the receiving end is to reconstruct a continuous stream of audio from the received audio packets, which may have undergone varying end-to-end network delays and packet loss. Buffering of received audio packets at the destination, and hence delaying their playout is a prudent mechanism used to compensate for varying network delays.

### 6.2.5. IPSec layer

The IPSec layer, implemented between the IP and physical network layers in all network components, is responsible for providing confidentiality and authentication security services. It allows for various configurations of security services, such as providing confidentiality alone, or authentication alone, or both, by using ESP in transport or tunnel mode. The security policy at each end host can be configured such that security services are applied to only SIP signalling stream, to media stream, to both, or to none. The cryptographic algorithms implemented are DES in CBC mode for encryption and HMAC-MD5 for authentication. The processing latencies introduced into the model by the MD5 and DES algorithms were based on values reported by Touch [20] and Schneier [21], respectively. The IPSec

layer supports both manual keying of SAs and dynamic setup of SAs using the IKE process.

### 6.2.6. IKE layer

The IKE process provides automated key exchange, SA creation and SA management services to clients, e.g. IPSec. IKE is based on the ISAKMP framework, and uses the ISAKMP phases and message construction syntax. The key exchange is carried out using the Diffie-Hellman public key exchange mechanism [22] over a MODP (prime modulus) group with 768-bit modulus. An efficient library for multiple precision integer arithmetic required for Diffie-Hellman exchange was implemented using the algorithms suggested in Ref. [23]. Ref. [23] also specifies the computation requirements for the various algorithms, which were used to include realistic Diffie-Hellman processing latencies into the IKE layer process model. An API between the IKE and IPSec layer serves for transferring of SA requests and resultant SAs.

As IKE messages are carried over UDP, which does not guarantee packet delivery over lossy networks, a retransmission strategy with exponential backoff mechanism was used for ensuring reliable IKE message transmission. IKE messages are retransmitted starting with a suitable time interval and the interval is doubled for each subsequent retransmission. Retransmissions cease when either there is an appropriate response from the IKE peer or a maximum limit of seven retransmissions is reached. The efficiency of retransmission mechanism is vastly dependent on the choice of the initial value for the retransmission interval [24]. For example, the graph of Fig. 8 shows the percentage additional time required for the IKE setup under lossy network conditions, due to message retransmissions, for various choices of the initial retransmission interval.

As seen from the graph, if the initial retransmission interval is set to approximately half the underlying mean network delay, around 7% increase in IKE setup times could be expected due to message retransmissions, considering that the network exhibits a packet loss of 5%. Delay due to IKE retransmissions could inadvertently affect the transmission delays of SIP/RTP packets, which are awaiting IKE setup, but could be kept to a minimum with a suitable choice of initial retransmission interval.

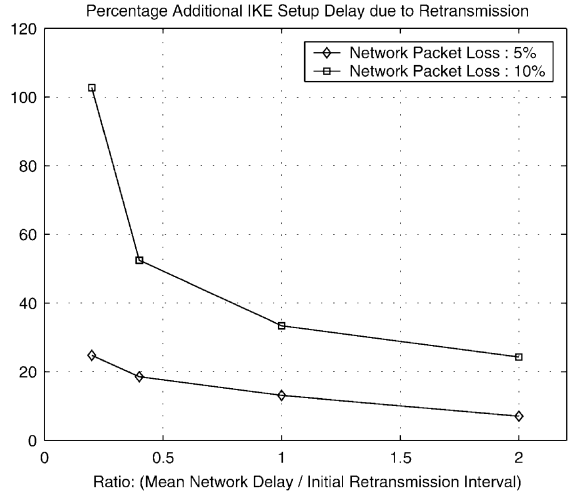


Fig. 8. Percentage additional IKE setup delay due to message retransmissions.

### 6.2.7. Gateway model

The gateway model is a representation of the bridge between the PSTN and the VoIP network. The main purpose of gateway is to translate call control signalling and media information between the PSTN and VoIP protocol architectures. The traffic-processing limit of the gateway is expressed in terms of the number of voice channels (which is configurable with the model), in other words, the maximum number of simultaneous calls that could go through the gateway. The gateway is a likely location where queue length build up will occur, as a single encryption engine at the IPSec layer will be shared by all the multiple call channels.

The simulation gateway node model hides the PSTN interfacing from the VoIP network model. The gateway simulation model is primarily used to study the impact on the call setup and media stream delays due to the shared encryption engine at full load conditions.

### 6.2.8. IP cloud model

The IP cloud model implementation is used to induce packet delay and packet loss into the transmission, and hence to simulate a near real-time degradation of packet transmission. Individual packet delays, were computed using normal probability distribution model as

$$\text{Packet delay} = K + \text{normal}(\text{mean} = 0, \text{var})$$

where normal(mean $= 0$, var) is a random value calculated using normal distribution with mean equal to zero and a specified variance. The mean delay contribution, $K$, can be configured to remain constant, or set to vary over simulation time. For the time-varying case, $K$ could be configured to vary continuously or to change at discrete time points during the simulation lifetime.

Packet loss introduced by the IP cloud model is based on the often used two state Markov chain model, also known as the Gilbert model for bursty packet loss, as suggested by Ref. [25].

In the Gilbert model, as shown in Fig. 9, $p$ is the probability that the next packet is lost, provided the previous one is not lost, and $q$ is the reverse. $(1 - q)$ is the conditional loss probability. Normally $(p + q) < 1$. The average packet loss rate for such a model is given as $lr = p/(p + q)$. The probability of getting a burst of length $n$, is equal to $q(1 - q)^{(n-1)}$. The IP cloud model allows the average loss rate, $lr$ and the burstiness factor, $q$ to be configured.

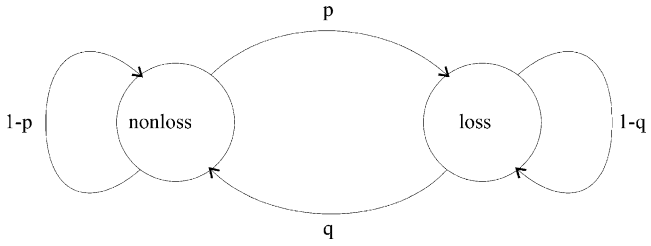


Fig. 9. Gilbert model for bursty packet loss.

## 7. Simulation strategy and results

### 7.1. VoIP network performance measures

Generally, the performance of VoIP networks is quantified in terms of the delay experienced by the media packets. An average one-way delay of around 150 ms for media packets is considerable acceptable [9]. An additional aspect of a VoIP call which, would impact upon the user's perception of service provided by a VoIP network is the average call set up time. As regards to acceptable values for call set up delays, there is generally only a single fully standardised recommendation for a maximum acceptable call setup delay, namely ITU-T recommendation E.721 [26]. However, strictly speaking this recommendation is applicable to circuit-switched telephony networks and hence is not directly applicable to VoIP networks, where it would be expected that call setup delays will be significantly shorter under normal loading conditions. A number of standard bodies, such as the IETF, are currently studying the issue of providing recommendations [27] for maximum acceptable call setup delays for various types of VoIP environments.

In order to evaluate the simulated secure VoIP network model, the following performance parameters were monitored for the model:

(1) SIP call setup time.
(2) Average delay per call experienced by media packets.

The SIP call setup time is measured as the time difference between the instant at which the SIP INVITE message is sent and the time instant at which the SIP 200 OK (confirmation) message is received. The average media delay per call is computed as,

$$\text{Average media delay per call} = \frac{\sum_{\forall \text{ calls}} \{\text{Mean media packet delay for the call}\}}{\text{Total number of calls}}$$

### 7.2. Secure VoIP network performance analysis criterion

The criterion for performance analysis were:

(1) Effect of employing encryption and authentication algorithms, on the VoIP network performance
(2) Additional effect because of the use of dynamic key exchange algorithms, IKE
(3) VoIP–PSTN gateway analysis.

Any encryption or authentication algorithm, which operates on media/signalling packets, results in a definite delay overhead. Assuming the pre-existence of all required cryptographic keys, the first criterion isolates the effect of the delay due to encryption/decryption/authentication alone. The simulation results for this criterion were obtained by manually configuring the IPSec layer at network edge routers with SAs holding 'valid forever' static keys.

The key exchange mechanism, when employed, adds to the delay due to the encryption and authentication layers. Public key exchange mechanism in particular is very expensive in terms of time and processing requirements. The second criterion analyses the effect of employing the IKE mechanism in association with IPSec. The use of IKE results in dynamically created SAs, which usually have a validity duration or lifetime associated with them. Once the SAs expire, they have to be re-keyed resulting in a fresh Diffie-Hellman public key exchange again. As Diffie-Hellman exchanges are demanding on processing resources, the validity duration of the SAs emerges as a key factor for the VoIP network performance. The simulation results for this criterion were obtained for different refresh rates for the dynamic SAs.

The shared encryption engine between multiple call channels at the VoIP–PSTN gateway results in high probability of occurrence of packet clogging. The third criterion for performance analysis, studies the impact of employing IPSec, on the average packet delay introduced by the gateway.

### 7.3. Secure VoIP network simulation results

The simulation results were obtained with the IP cloud models configured to induce a mean network delay of

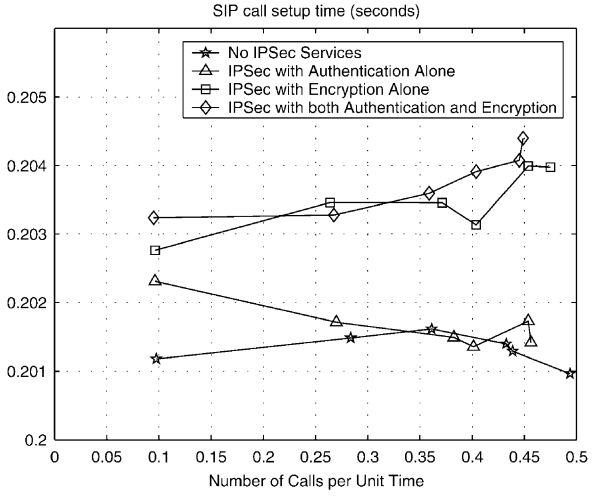


Fig. 10. SIP call setup time comparison for different IPSec configurations.

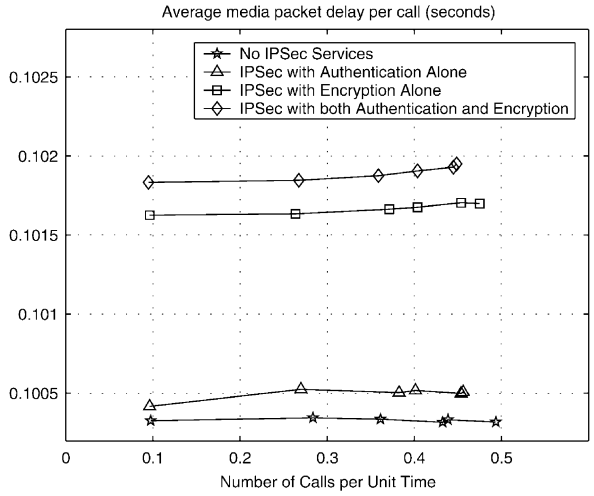


Fig. 11. Media packet delay comparison for different IPSec configurations.

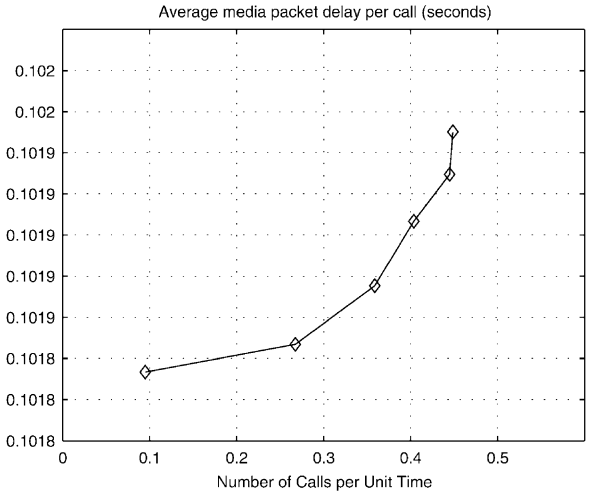


Fig. 13. Media packet delay analysis for full security IPSec configuration.

100 ms, and a packet loss of 1%. Figs. 10–13 illustrate the impact of employing encryption and authentication algorithms, on the VoIP network performance. The simulation results were obtained with network edge routers configured to apply IPSec ESP protocol in tunnel mode operation, with manually keyed SAs. Fig. 10 gives a comparison of the effect on SIP call setup times for the different basic combinations of IPSec configurations, i.e. (a) with no security services, (b) with both encryption and authentication employed, (c) with encryption alone, and (d) with authentication alone. Fig. 11 shows a comparison of the effect on the media stream delays for the same set of IPSec configuration options. Figs. 12 and 13 highlight the isolated graphs for SIP call setup time and media stream delay, respectively, for the case where both encryption and authentication employed.

Figs. 14 and 15 study the impact of using IKE dynamic key exchange mechanism along with IPSec. The graphs are parameterised according to the validity duration of the dynamic SAs, i.e. the lifetime of SAs. The simulations were run with the SA lifetime set as different multiples of the mean VoIP call duration. Fig. 14 highlights the impact on SIP call setup times and Fig. 15 highlights the impact on the media stream delays.

Fig. 16 exhibits the impact of using IPSec at the VoIP–PSTN gateway terminal. The graph represents the variation in average per packet delay introduced by the gateway with increasing number of calls (or call density) going through the gateway.

## 8. Performance analysis

The impact analysis of employing IPSec encryption and authentication services for VoIP signalling and media streams (Figs. 10–13) show that, between encryption and authentication, encryption is the more expensive operation. When both encryption and authentication services are employed, an increase of around 1.4% in the SIP call setup times, and an increase of around 1.6% in the media

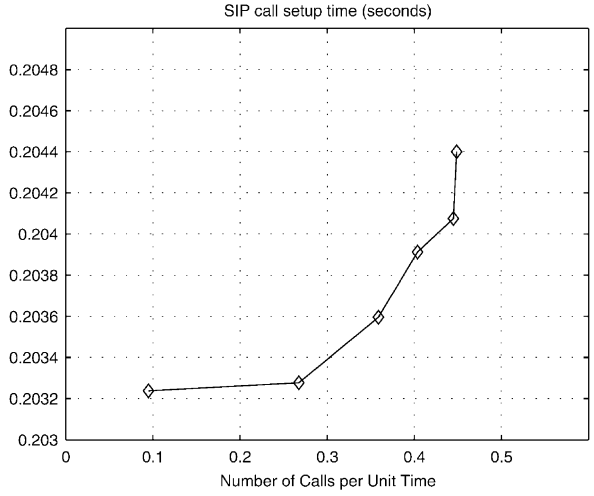


Fig. 12. SIP call setup time analysis for full security IPSec configuration.

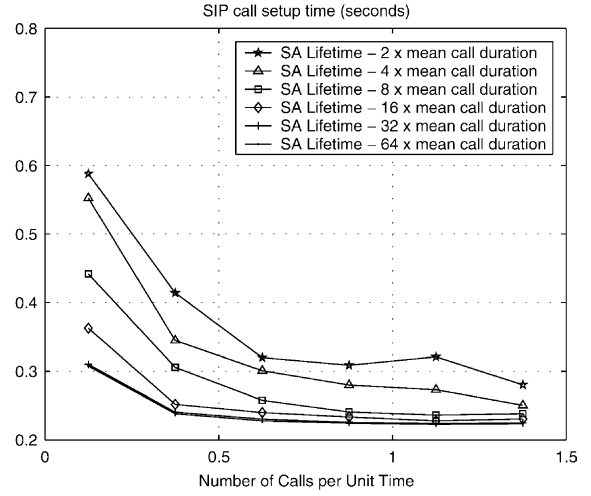


Fig. 14. SIP call setup time variation with IKE SA validity duration.

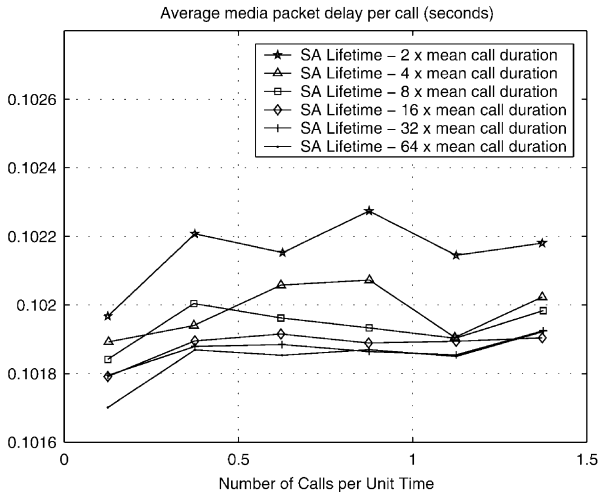


Fig. 15. Media packet delay variation with IKE SA validity duration.

stream delays is seen. From Figs. 12 and 13, the SIP call setup times and media stream delays seem to increase exponentially with increase in the network call density. This increase is prominently due to the packets getting queued up at edge routers waiting to be processed by the IPSec security engine.

### 8.1. SIP call setup time analysis

Impact analysis of employing dynamic key exchange mechanism, such as IKE, along with IPSec shows more alarming effect on VoIP network performance. Firstly, the scale of graphs in Fig. 14 shows a huge increase in SIP call setup times, when compared to the scale of graphs in Figs. 10 and 12. The graphs in Fig. 14 illustrate a decreasing trend of SIP call setup times with increasing call density. The reason for this is the way in which dynamic SAs are created. The creation of SAs is triggered off, when a call is to be made and there is no suitable SA available. Hence, the first call, or rather the first packet has to wait until the Diffie-Hellman exchange is complete and the SAs are created. Once the SA is created, the subsequent calls, which suit the same SA, can go through without being held up, and the call setup delay is governed only by the network delay. As the call density increases, the number of calls secured using every individual SA, increases, i.e. the utility of the SA increases. The performance cost of setting up a dynamic SA is very high and dominates the effect on performance. Also, it is observed from the graph that a very low SA life duration results in SAs being setup very frequently, resulting in huge delays.

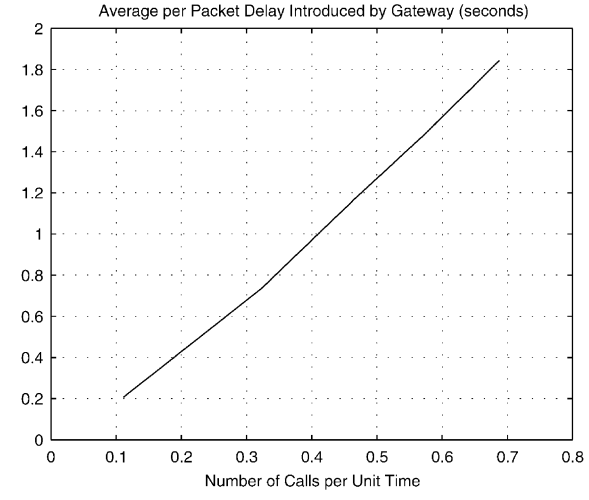


Fig. 16. Gateway performance analysis.

### 8.2. Average media packet delay analysis

Fig. 15 presents the impact which IKE has on the average delay experienced by media packets. As media packets are generated at spaced intervals, every 20 ms in this case, and considering their heavy volume, the impact of SA lifetimes on the average media packet delay is slightly different than the impact on SIP call setup times. As in the case of SIP, when there is no SA available, the initial media packets are held up at the IPSec layer until the SA is created. Once the SA is created, all the queued up packets are processed and transmitted on to the network, and hence the subsequently generated packets are only affected by the network delay. While the initial media packets are delayed heavily during SA setup, the majority subsequent packets, which are generated at spaced intervals, do not experience much additional delay due to the previous SA setup. Hence, the average media packet delay per call, shows an increase, albeit comparatively by a smaller amount, with decreasing SA lifetimes.

Also it is seen that the average media packet delay increases with increasing call density. This is attributed to the fact that all the media packets undergo additional queuing delays at the IPSec layer due to encryption and authentication. The cost of encryption and authentication dominates the effect on performance in this case.

It should be emphasised that the results presented in Fig. 15, are ‘averages over all calls’ of the mean media packet delay per call. However, for some calls, specifically for the ones, which await SA setup, the media packets could have undergone excessive delays. This excessive impact is represented by the maximum mean media packet delay per call and the corresponding 95th percentile, as in Table 3.

The worst impact is seen for the case where the SA validity is configured to be twice the mean call duration. For this case, at the maximum, the mean media packet delay for a call has increased by approximately 27 ms, over the configured network mean network delay of 100 ms. Though the resulting worst impacted delay of 127 ms is well below the standard QoS limit of 150 ms, a higher underlying mean network delay could result in average media delays for some calls falling below the required QoS limit.

Table 3
Worst impact media packet delay statistics

| SA validity duration | Maximum delay (ms) | Corresponding 95th percentile (ms) |
|---|---|---|
| 2 × mean call duration | 127.275 | 107.023 |
| 4 × mean call duration | 122.571 | 102.405 |
| 8 × mean call duration | 106.788 | 102.439 |
| 16 × mean call duration | 104.865 | 102.291 |
| 32 × mean call duration | 105.905 | 102.266 |
| 64 × mean call duration | 104.682 | 102.307 |

### 8.3. Gateway performance analysis

The gateway analysis (Fig. 16) shows a linearly increasing average delay introduced per packet by the gateway with increasing call density through the gateway. These results are as expected given that the gateway could be analysed as a simple M/D/1 queuing model. The shared encryption engine between the multiple number of call channels, is a concern in terms of gateway performance.

## 9. Conclusions

This paper has described an OPNET Modeler® simulation model of an IPSec secured SIP based VoIP network. The performance analysis of secure SIP–VoIP network, aided by this tool has been presented along with results obtained for various IPSec configurations. The most predominant and alarming effect on VoIP network performance was seen in the case where dynamic public key exchange mechanisms, such as IKE was used for establishing shared and authenticated secret keys. The other performance bottleneck observed was the shared encryption engine at edge routers and VoIP–PSTN gateways, which could very well be averted by using multiple encryption engines.

As a security option for VoIP, the IPSec framework offers a myriad of choices and options for providing security services. Performance analysis of the different configurations of IPSec for VoIP networks, using simulation models is a prudent method for basing decisions on, when implementing real IPSec secured VoIP networks.

## Acknowledgments

This research has been carried out with the support of funding from Enterprise Ireland and Nortel Networks, under Enterprise Ireland's Applied Research Programme. The use of OPNET Modeler® in this research was facilitated through OPNET's University programme.

The authors would like to thank Burkhard Springer of the Department of Electronic Engineering in NUI, Galway for his input on the issue of IKE re-transmission strategies and packet burst loss modelling.

## References

[1] M. Atiquazzaman, M. Hassan, A. Nayandoro, Internet telephony: services, technical challenges and products, IEEE Communications (2000) April.
[2] G.A. Thom, H.323: the multimedia communications standard for local area networks, IEEE Communications (1996) December.
[3] H. Liu, P. Mouchtaris, Voice over IP signaling: H.323 and beyond, IEEE Communications (2000) October.
[4] H. Schulzrinne, J. Rosenberg, The session initiation protocol: internet-centric signaling, IEEE Communications (2000) October.
[5] M. Handley, E. Schooler, H. Schulzrinne, J. Rosenberg, STP: session initiation protocol, IETF RFC 2543 (1999).
[6] Q.931: Digital Subscriber Signaling System No. 1 (DSS 1)—ISDN User—Network Interface Layer 3 Specification for Basic Call Control, ITU-T Recommendation, Q Series, 1998.
[7] S. Casner, R. Frederick, V. Jacobson, H. Schulzrinne, RTP: a protocol for real-time applications, IETF RFC 1889 (1996).
[8] J. Rosenberg, SIP Security, URL http://www.dynamicsoft.com/resources/pdf/SIP2000-Security.pdf.
[9] G.114: One-way Transmission Time, ITU-T Recommendation, G Series, 2000.
[10] M. Handley, V. Jacobson, SDP: session description protocol, IETF RFC 2327 (1998).
[11] R. Atkinson, S. Kent, Security architecture for internet protocol, IETF RFC 2401 (1998).
[12] R. Atkinson, S. Kent, IP authentication header (AH), IETF RFC 2402 (1998).
[13] R. Glenn, C. Madson, The use of HMAV-MD5-96 within ESP and AH, IETF RFC 2403 (1998).
[14] R. Atkinson, S. Kent, IP encapsulating security protocol (ESP), IETF RFC 2406 (1998).
[15] N. Doraswamy, C. Madson, DES-CBC cipher algorithm with explicit IV, IETF RFC 2405 (1998).
[16] D. Carrel, D. Harkins, The internet key exchange (IKE), IETF RFC 2409 (1998).
[17] D. Maughan, M. Schertler, M. Schneider, J. Turner, Internet security association and key management protocol (ISAKMP), IETF RFC 2408 (1998).
[18] K. Sriram, W. Whitt, Characterizing superposition arrival processes in packet multiplexers for voice and data, IEEE Journal on Selected Areas in Communication 4 (6) (1986) 833–846.
[19] H. Schulzrinne, Some Frequently Asked Questions about RTP, URL http://www.cs.columbia.edu/~hgs/rtp/faq.html.
[20] J. Touch, Performance analysis of MD5, Proceedings of ACM SIGCOMM '95 (Cambridge, MA, 1995), 1995, pp. 77–86.
[21] B. Schneier, Applied Cryptography, Wiley, New York, 1996, pp. 278–279.
[22] E. Rescorla, Diffie-Hellman key agreement method, IETF RFC 2631 (1999).
[23] A.J. Menezes, P.C. van Oorschot, S.A. Vanstone, Handbook of Applied Cryptography, CRC Press, Boca Raton, 1996, pp. 591–634.
[24] B. Springer, Performance Issues Relating to IKE Re-transmission Strategies in Secure VoIP Networks, Private Correspondence, Department of Electronic Engineering, National University of Ireland, Galway, Ireland, May 2002.
[25] G. Carle, R. Koodli, H. Sanneck, A framework model for packet loss metrics based on loss runlengths, Proceedings of SPIE/ACM SIGMM Multimedia Computing and Networking Conference 2000 (San Jose, CA, 2000), 2000, pp. 177–187.
[26] E.721: Network Grade of Service Parameters and Target Values for Circuit-Switched Services in the Evolving ISDN, ITU-T Recommendation, E Series, 1999.
[27] A. Broscius, C. Huitema, H. Lin, T. Seth, VoIP Signaling Performance Requirements and Expectations, Internet Draft, Internet Engineering Task Force, 1999. Work in Progress.

**Mohan Krishna Ranganathan** received the BE degree in Electronics and Communication Engineering from University Visvesvaraya College of Engineering, Bangalore University, India, in 1998. He is currently pursuing the PhD degree with the Department of Electronic Engineering, National University of Ireland, Galway, Ireland. From 1998 to 2000, he was with Infosys Technologies Limited, Bangalore, where he was involved with Internet Programming. His research interests include VoIP, security and performance modelling.

**Liam Kilmartin** has been a lecturer in the Department of Electronic Engineering at the National University of Ireland, Galway since 1992. His research interests include modelling of advanced communication networks, speech and image processing and neural networks. He also acts as a senior software architect for Tango Telecom Limited in Limerick, Ireland designing next generation applications for converged and 3G mobile networks.