Security for VoIP transmissions over 802.11 and Bluetooth networks

Andrêa Nascimento, Loide Verdes, Alexandre Passito, Edjair Mota, Arlen Nascimento, Edjard Mota, Leandro Galvão

Federal University of Amazonas, Computer Science Department, Networks and Embedded Systems Group

{agon, lmvj, passito, edjair, aon1, edjard, galvao}@dcc.ufam.edu.br

Abstract

This paper presents a study on IP Security Protocol used to guarantee secure transmissions of Voice over IP on wireless networks. We have investigated and quantified the impact of IPSec security mechanisms on multimedia traffic and selected robust configurations to 802.11 and Bluetooth networks.

KEYWORDS: VoIP, wireless, security, quality-of-service.

1 Introduction

With increasing popularity of Voice over IP (VoIP) technology, we are even more using tools that allow multimedia traffic transmission through the Internet. These services offer two main advantages in relation to the commuted public telephony network: reduction of long-distance calls costs and possible development of smart services, like voice-mail, calls redirect, conferences, etc (Carvalho, 2004). This new telephony service was developed based on the TCP/IP structure, so it inherits many problems from this paradigm, like delay and loss of packets due to the best-effort behavior of the Internet, as well as its security problems.

This paper encloses security problems inherent to the TCP/IP networks, and how these problems can affect VoIP communications. More specifically we have evaluated the use of a very robust security standard protocol, the IPSec. The evaluation of this protocol was done using a measurement tool developed in our laboratory (MedQoS, 2006), that has quantified in an objective way the impact of the security mechanisms of IPSec using Quality of Service (QoS) as metric.

Based on previous works that have evaluated VoIPSec in wired networks, in this paper we enclose two technologies that have been widely used in mobile communication: 802.11b and Bluetooth. These technologies have been spread in the networks and embedded systems at UFAM (Federal University of Amazonas). We have evaluated the impact on the quality and the IPSec capacity of reducing the number of simultaneous calls on these systems.

The main goals of this paper are the measurement of the impact using the E-Model and the MOS (Mean Opinion Score) metric (Carvalho, 2004), the evaluation of the IPSec in many different contexts and a reflection on the way this mechanism can be used to improve the secure communication in the Amazon.

This paper is organized as follows: section 2 presents related works in VoIP security. Section 3 presents motivations to implement secure VoIP communication in the Amazon. Section 4 briefly presents features of 802.11 and Bluetooth networks and the motivation to use IPSec. Experiments and analysis of results are presented in Sections 5 and 6, respectively. We present conclusion and ideas for future works in the Section 7.

2 Related Works

A first work which has presented research results on the IPSec impact on VoIP traffic over Ethernet networks was Barbieri et al. (2002), more specifically evaluating the encrypting mechanisms. All the results have pointed that the traffic delay can be extremely affected due to the fact that scheduling with priority to voice traffic in the IPSec encrypting machine is not possible.

In Xiao and Zarrella (2004) and Rajavelsamy et al. (2005) the use of VoIP over WLANs was evaluated and the results have pointed that some variations of the IPSec parameters can affect the whole system performance, and the IPSec is an effective solution in comparison to other security mechanisms like WEP or WPA.

In Klaue and Hess (2005) audio and video protected by IPSec in a WLAN were evaluated. Results have presented that IPSec do not introduce substantial damage to the VoIP traffic, but we argue that such scenario is not realistic enough. The reason it is because it presents
restrictions on the number of nodes and on channel conditions. Our experiments have not presented such limitation, and as we can see in this paper, using a robust QoS tool the impact of using IPSec on VoIP traffic is most likely to be better evaluated.

In Passito et al. (2006) we have presented preliminary results about the VoIPSec performance over 802.11b networks. In this paper we have extended our work using Bluetooth networks and have made a comparative analysis.

3 Voice Over IP and Applications

In the VoIP technology, the analógical voice is transmitted through the Internet after passing for many processes, that go from coding and compression by the codecs, to the construction of frames with IP, UDP and RTP headers, beyond the payload. When these frames arrive at the receiver, the inverse process is made and the analógin signal is recovered.

During the way between the microphone, at the source, and the loudspeaker, at the receiver, many metrics of performance can be affected during the process of sending packets over the network. Among the metrics directly related to the end-to-end speech quality, the ones we chose for performance parameters are end-to-end delay, jitter and loss of packets rate in the network.

Beyond the real time multimedia traffic, VoIP systems also must to have mechanisms to establish and finish calls. These mechanisms are performed by signaling protocols. Many signaling protocols for multimedia sessions were developed, but only two of them have been spread: H.323, based on many ITU-T recommendations, for its complete architecture, and SIP (Session Initiation Protocol), proposed by IETF, for its simplicity and many functionalities for multimedia services.

VoIP has been successfully employed on wired networks and backbones. Many institutions like universities and public agencies have extended the services of the conventional telephony to their internal networks. There are many benefits, as the employment of smart services and reduction of costs on maintenance and interurban calls. The Federal University of Amazonas is an example. It offers VoIP service to its students and teachers, improving research productivity due to the easy communication with other researchers around Brazil.

The main motivation for this paper is the experience acquired with VoIP and wireless networks researchers, and the possibility of incorporating this experience in many scenarios in the Amazon. More specifically, we tackle the problem of security in some wireless network standards and how this can affect the VoIP system performance.

4 VoIP Security

A voice service implementation must offer robust security mechanisms and almost always these mechanisms have to be optimized for VoIP. That is due to intrinsic features of this technology that can be used for intruders on new attacks: causing a delay bigger than 150 ms, that is the biggest acceptable delay (Barbieri et al., 2002) for a good quality communication, the jitter variation, or a burst loss of packets. These attacks can be very harmful to the voice traffic.

There are many other vulnerabilities:
– eavesdropping: the intruder can collect voice packets and reproduce them using tools like VOMIT (Vomit, 2006);
– information theft;
– denial of service (DoS);
– jamming.

The main challenge is to offering security to the VoIP technology, with solutions to all problems mentioned above, on challenging environments with, for example, wireless networks and users mobility. In our work we have used 802.11 and Bluetooth networks, a well-known and an emergent technologies, respectively, that present many security problems on the physical and link layers. Besides these problems, the diffusion of packets feature allow intruders to collect data inside the antennas range.

In the next subsections we briefly describe the wireless communication technologies we have used in this work, and we identify some security failures already shown in other works, like authentication and communication confidentiality failures, and additional security mechanisms to these networks.

4.1 Bluetooth Networks and Security Problems

Bluetooth works on the ISM band (2,4 GHz) and is inclined to interference. In order to minimize this problem, Bluetooth uses a technique named FHSS (Frequency Hoping Spread Spectrum). This technique divides the channel in a random sequence of 79 frequencies with a rate of 1600 hopes per second. Each channel is divided in time slots of 625µs. Bluetooth works on the rate of 1Mbps using GFSK (Gaussian shaped Frequency Shift Key) method for modulation.

Topology configuration of Bluetooth defines two operation modes for devices: master or slave. A master device executes a search in its range area looking for slave devices that must answer the request. A master is a central unity responsible for data traffic. A basic Bluetooth network is called piconet, and allows one master device
interconnect with up to seven active slave devices and 255 inactive slave devices.

Piconets support synchronous connection for real-time communication and asynchronous connection for normal data flow. When we execute real-time applications, like voice applications, SCO (Synchronous Connection Oriented) channels are created. If they are not real-time applications, ACL (Asynchronous Connectionless) channels are used.

A SCO link offers low latency but does not guarantee data integrity. The low latency is due to two mechanisms: packets are queued for transmission on specific time slots and are never relayed. SCO channels also define some different types of packets (High Quality Voice Packets – HV), and in this paper we have used HV3 packets, based on previous works (Kapoor et al., 2001; Wu and Todd, 2004).

Bluetooth security has three modes: not secure (Mode 1), service level security (Mode 2) and link level security (Mode 3). Authentication and encryption in the link layer are performed through four basic entities:

- Bluetooth device address, a 48 bits identifier associated to each device.
- A private authentication key (random number).
- A private encryption key (random number).
- A 128 bits random number frequently changed and randomly generated by each device.

A PIN (Personal Identification Number) code, size between 1 and 16 octets, must be inserted on the initialization of all Bluetooth devices. There is also a possibility of maintain this PIN on hardware.

Bluetooth adopts EO stream cipher. For each session an encryption key is generated and packets are protected. This method is superior in comparison to WEP used on 802.11 networks, but many important failures were reported in Hager and Midkiff (2003). These failures are related to device validation, exposition of keys, random numbers generating. Vulnerabilities also include address and state of devices validation failures (Jakobsson and Wetzel, 2001).

4.2 802.11 Networks and Security Problems

The IEEE 802.11 standard offers wireless connectivity to those devices that need quick and easy installation, like mobile computers, PDAs, or any mobile device in a wireless LAN (WLAN). The standard defines MAC as the physical media access, that can be infrared or radio frequency. The available bandwidth is partitioned in 14 partially overlapped channels, each one with 22MHz width. Every device in a range uses the same channel. For modulation, following techniques can be used:

- DSSS: adopted for rates of 1 and 2 Mbps;
- CCK: defined in the 802.11b standard, it is adopted for rates of 5.5 and 11 Mbps.
- OFDM: defined in the 802.11g standard, it is also used for rates of 6, 9, 12, 18, 24, 36, 48 and 54 Mbps.

A 802.11 WLAN is based on a cellular architecture where each cell is called basic service set (BSS). A BSS is a set of 802.11 mobile or fixed stations. The network can work in ad hoc mode or in infrastructure mode. In the last case, the network depends on an access point (AP) to reach other mobile stations in a different BSS, or in a wired network.

In order to provide security to the data traffic on the level of physical and link layers, WEP (Wired Equivalent Privacy) was proposed to WLANs. Using WEP could avoid some kinds of security attacks as eavesdropping, denial of service, and non-authorized location of users. Unfortunately, many failures and different attacks were presented in Fluhrer et al. (2001), and more recently in Shin et al. (2006), for example, authentication and access control failures, which make the protocol quite insecure.

4.3 VoIPSec on Wireless Networks

Considering that call end-to-end confidentiality is a VoIP security requirement, 802.11 and Bluetooth security mechanisms could protect only the communication between the end mobile user and the access point. Due to this requirement, and due to the necessity of repairing technology security failures, it is also necessary to use security mechanisms from upper layers.

Using virtual tunnels, or VPNs (Virtual Private Networks), is our alternative to avoid this problem in this work and in previous works. VPNs offer security with the integration of authentication, encrypting, access control and session management services. Users that do not know directly each other can establish a communication due to different configurations of the VPNs (Vatn, 2005).

IPSec is a protocol used to implement virtual private networks. It is formed by the protocols: AH (Authentication Header), ESP (Encapsulating Security Payload) and IKE (Internet Key Exchange). IPSec supports many encrypting and authentication algorithms and they are largely tested on the most recent Linux kernel version.

AH provides integrity, source authentication and protection against reproduction attacks. ESP offers besides the same of AH, confidentiality (encrypting) of the data traffic.

In order to use IPSec with VoIP applications, it is necessary a native support for the security protocol by the operating system. Secure tunnels also can be initialized in a independent way from the application, or else can be initialized by the application. We are interested on the behavior of this scheme (VoIPSec) over the used wireless networks.
5 Experiments

5.1 General Configurations

In order to measure the impact of IPSec on the multimedia traffic in both wireless environments, we have made experiments on scenarios presented in Figures 1 and 2, at the UFAM campus.

In scenario I, two Bluetooth access points are directly linked by an Ethernet link. Each access point acts like a master on a piconet, changing informations using a LAN profile. The position of the access points in this scenario represents a real VoIP application on a public domain where many access points are connected in order to provide a bigger range.

The scenario II has consisted of two WLANs and the positions of the access points were similar to the ones in scenario I.

During each experiment we have had three different entities involved:

- 6 laptops, AMD 1.6 GHz and 256MB RAM. Each computer was configured with a Bluetooth USB device or with a 802.11b PCMCIA.
- Bluetooth access points (D-Link) with default configuration.
- 802.11 access points with transmission rate of 11 Mbps.

The operating system used in laptops was Debian GNU/Linux, 2.6.11 kernel version, with Bluetooth support through the library Bluez. The IPSec has a stable support on this kernel version, and it was configured with shared keys. Security configurations used for the experiments are presented:

- AES-CBC (192) and HMAC-MD5 (128)
- 3DES-CBC (192) and HMAC-MD5 (128)

For encrypting we have used 3DES (Triple Data Encryption Standard) on CBC mode (Cipher Block Chaining), and AES (Advanced Encryption Standard). For authentication we used HMAC-MD5.

Calls between mobile users were established through extended versions of known softphones. We have made modifications on the Ohphone code, so that it provides information about VoIP traffic, like end-to-end delay, jitter and loss packets rate. Besides that, these modifications have led to the development of a computational tool based on the E-Model that is able to calculate the MOS score for each call. Our tool is largely explored in Carvalho et al.(2005) and Lustosa et al.(2004).

Calls were realized with an automatic calls generator (Callgen323). It takes an audio file with conversation features and sends this file to other computer in a VoIP call. Using a file with conversation features like periods of speech and periods of silence has improved the trustworthiness of our experiments.
Full-duplex calls were made using codec G.711 μ-Law (64 kbps) with frames of 30 ms. Figure 3 shows the configuration used on the laptops and presents a layer distribution for the entities involved.

5.2 Dynamics of the Experiments

Experiments were made in both scenarios, with the same set of security configurations, in order to obtain a comparative analysis of Bluetooth and 802.11 technologies. At a first moment, calls were made without a security mechanism, and after that calls were made using IPSec with AES and 3DES. The HMAC-MD5 authentication protocol was the same to all the experiments.

Our experiments were performed in two parts: the first one was to provide a comparative analysis of the impact of IPSec protocol on the technologies with only one call in the network. The second part was to evaluate the reduction of the channel capacity in the case of many simultaneous calls.

In the first part of the experiments, calls were established between two different cells in the network, between two WPANs on the scenario I, or between two WLANs on the scenario II. Callgen323 has generated data voice traffic from WPAN1/WLAN1 to WPAN2/WLAN2 where the voice data traffic was received by Ohphone, after passing the Ethernet link. 100 calls of 5 minutes were analyzed for each different security configuration, a reasonable number of calls to calculate the MOS mean score. According to Carvalho (2004), common calls have only a few minutes of duration.

On the second part of the experiments we have established simultaneous calls between mobile stations on WPAN1/WLAN1 and WPAN2/WLAN2, for each security configuration. Callgen323 is able to generate simultaneous calls.

6 Result Analysis

Analyzing the impact of IPSec was based on the total delay of packets, jitter, loss packets, etc. Evaluating VoIP calls quality is not simple, considering that the final user evaluates the quality of the whole call; the user does not know about the details that influence the general quality. That leads to a necessity of evaluating many metrics, like the ones cited above, related to the communication in order to have a complete measurement of the call quality. In order to avoid this problem the tool MedQoS (MedQoS, 2006) was implemented based on the E-Model to evaluate, in an objective way, the quality of VoIP communications.

E-Model calculates the speech quality through the R factor, that is converted to the MOS score. MOS is a specific score to evaluate the speech quality and it ranges from 1 (very poor) to 5 (excellent) (Carvalho et al., 2005). R Factor is obtained as follows (Carvalho, 2004):

\[ R = R_o - I_s - I_d - I_{e,eff} + A \]

where:
- \( R_o \) represents effects of the signal to noise ratio;
- \( I_s \) represents impairments simultaneous to voice signal;
- \( I_d \) represents impairments related to the end-to-end delay;
- \( I_{e,eff} \) represents the lost related to the equipment used;
- \( A \) corresponds to the advantage factor (attempts to account for caller expectations).

Figure 4 presents a comparative analysis of Bluetooth and 802.11b technologies, with the results of MOS score for the calls on our experiments. As we can see, the speech quality is above 4,25 when there is no security mechanism on both scenarios. Using IPSec, the MOS score decreases to below 4,0 using AES encrypting algorithm, and 3,7 using 3DES.

We can also notice that the MOS value for 3DES algorithm on scenario II (802.11b) was 3,39, a value below the minimum necessary for a good communication, that in according to Carvalho et al. (2005) must be above 3,5.
We have realized many tests with simultaneous calls on both scenarios between mobile stations in different cells, in order to evaluate the impact of IPSec on calls passing through the wireless and wired channels.

Simultaneous calls were established between one mobile station in WPAN1 and another mobile station in WPAN2 on scenario I with Callgen323. MOS score values for these calls are presented in Figure 5. We can analyze comparatively the speech quality in relation to the security configurations (without security, with AES algorithm for security, with 3DES algorithm for security).

We could notice that the MOS score decreases as the number of simultaneous calls increases. In Figure 5, for 2 simultaneous calls the MOS score was above 4.0, decreasing to below 2.0 for 12 simultaneous calls.

It is important to cite that Callgen323 tries to establish the requested number of calls, but due to different overheads caused by different configurations of IPSec, the mean number of established calls is different.

In Figure 6 we present a graphic that shows the mean MOS score for simultaneous calls on scenario II. Experiments presented in this graphic were made between WLAN1 and WLAN2. The initial MOS score for 2 simultaneous calls without security was above 4.0. This number has decreased to below 1.5 for 24 simultaneous calls. We have gotten MOS between 3.5 and 4.0 for 2 simultaneous calls using AES and 3DES encrypting algorithms. For 24 simultaneous calls, the MOS score was below 1.5 for both encrypting algorithms.

In Wang et al. (2005) we find an analysis of the G.711 codec, with frames of 30 ms and $BER = 10^{-4}$, where the maximum number of simultaneous calls on 802.11b networks is 9. Our results with the standard IEEE 802.11b have presented an approached number.

Two factors in the E-Model can be used to analyze the results. The first one is the $I_e$ factor. When IPSec on tunnel mode is used, ESP header is added to the voice packet. It increases the ratio between packet size and payload reducing the effective bandwidth. Analyzing the numbers in the tool traces, we can relate the increasing loss of packets with the new size of the packets, with IPSec headers. $I_e$ factor is very sensible to the increasing loss of packets rate.

The second parameter in the E-Model affected by the IPSec overhead is the $I_d$ factor. This factor is related to all delays in a VoIP system. The bigger the total delay of packets, the bigger the $I_d$ factor, and consequently the smaller the MOS value. When IPSec is used, each small voice packet has to be encrypted and decrypted. This process in the encrypting machine takes many time, and its latency introduces a bottleneck in the system, affecting the speech quality in the receiver. It can be more evident when we take expensive algorithms, like 3DES, into account.

Figure 7 presents numbers for an important metric in E-Model, the delay of packets. We can see the difference of the delay introduced by the encrypting machine for each encrypting algorithm. The graphic shows that 3DES algorithm introduced a bigger delay in comparison to delay with AES and without IPSec. Analyzing all graphics of this paper we...
can see the importance of this metric for a good quality communication, its influence in the E-Model value is visible.

The scheme in Figure 8 shows that there is no priority for voice packets on the queue (FIFO) in the encrypting machine. D1 and D2 values represent delay added to voice packets in the encrypting and in the reconstruction of headers. This reconstruction influences in the Id factor value.

7 Conclusion and Future Works

We have demonstrated in this work that IPSec can be used in 802.11 and Bluetooth networks to guarantee a secure communication in these wireless environments. Results have showed that despite the negative impact on the speech quality, if algorithms are selected in an optimized way, IPSec presents a reasonable performance on presented scenarios.

For future works we can cite the evaluation of IPSec impact on VoIP system over embedded systems, due to limited resources of these devices, encryption can not obtain enough resources and introduces worse impairments.

For the problems with the encryption machine, we can propose modifications on the scheduler in order to prioritize VoIP traffic.

References


