available at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/cose

Computers & Security

E^2VoIP^2: Energy efficient voice over IP privacy

Elias Abou Charanek, Hoseb Dermanilian, Imad Elhajj, Ayman Kayssi, Ali Chehab*

Department of Electrical and Computer Engineering, American University of Beirut, Riad El-Solh, Beirut 1107 2020, Lebanon

ARTICLE INFO

Article history:
Received 25 December 2010
Received in revised form
30 May 2011
Accepted 6 July 2011

Keywords:
VoIP
SRTP
E^2VoIP^2
Selective encryption
Handheld devices
Energy efficiency
Security

ABSTRACT

Due to the convergence of telecommunication technologies and pervasive computing, voice is increasingly being transmitted over IP networks, in what is commonly known as Voice over IP (VoIP). Despite many advantages offered by this technology, VoIP applications inherit many challenging characteristics from the underlying IP network related to quality of service and security concerns. Traditional ways to secure data over IP networks have negative effects on real-time applications and on power consumption, which is scarce in power-constrained handheld devices. In this work, a new codec-independent Energy Efficient Voice over IP Privacy (E^2VoIP^2) algorithm is devised to limit the overhead of the encryption process, without compromising the end-to-end confidentiality of the conversation. The design takes advantage of VoIP stream characteristics to encrypt selected packets using a secure algorithm, while relaxing the encryption procedure in-between these packets. We evaluated experimentally the difficulty of conducting known plaintext attacks on VoIP by demonstrating that a sound recorded simultaneously by different sources results in apparently random encoded files. Regarding E^2VoIP^2, experimental and simulation results show a substantial improvement in terms of the number of CPU cycles which results in a reduction of latency and a reduction in consumed power with respect to that of the SRTP. In addition, the proposed method is flexible in terms of the balance between security and power consumption.

1. Introduction

Voice has always been an incontestable way for individuals to communicate. In addition, when we are spoken to, the voice we hear confirms the authenticity of the person speaking. Ever since the invention of the Public Switched Telephone Network (PSTN) to transmit the human voice in a relatively recognizable form (Andrew, 2003), technology has been evolving and giving rise to more opportunities and challenges.

Voice over Internet Protocol (VoIP) is gaining ground, not only because of the cheap call rates it provides but also due to the convergence of telecommunication technologies, pervasive computing and the internet (e.g. WiMax, Long Term Evolution,...) (Steven, 2008). New applications that integrate telephony, computing and ubiquity became viable with VoIP: rich media service, phone portability, and user control interface.

A very detailed survey is presented by Karapantazis and Pavlidou regarding emerging Voice over IP services and protocols. VoIP is compared to PSTN in terms of advantages, disadvantages and the offered services. VoIP services are typically based on monthly fixed costs or are completely free (Karapantazis and Pavlidou, 2009). However, VoIP applications inherit many challenging characteristics from the underlying IP network related to quality of service (QoS) limitations and security threats. The degradation of QoS in VoIP applications is

* *Corresponding author*. Department of Electrical and Computer Engineering, American University of Beirut, P.O.B.: 11-0236, Riad El-Solh, Beirut 1107 2020, Lebanon. Tel.: +961 1 350000x3444.

E-mail addresses: efa01@aub.edu.lb (E.A. Charanek), hmd10@aub.edu.lb (H. Dermanilian), ie05@aub.edu.lb (I. Elhajj), ayman@aub.edu.lb (A. Kayssi), chehab@aub.edu.lb (A. Chehab).

0167-4048/$ – see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.cose.2011.07.002

mainly due to the latency and jitter that voice packets are subjected to. Since voice applications are real-time and interactive applications, a delay that exceeds 150 ms from end-to-end is not tolerated by humans (Patrick, 2008; Glen et al., 2006). As for security threats, they can be categorized into threats against availability, confidentiality, social context, integrity or threats caused by vulnerable components (Patrick, 2008). Threats can also be classified into passive and active, such as denial of service, man in the middle, replay and cut-and-paste attacks, theft of service, eavesdropping, impersonation, poisoning attacks, credential and identity theft, redirection/hijacking and session disruption (Alan and David, 2006).

Traditional ways to secure digital data come at a price when applied to voice traffic. The literature discusses the effects of traditional security protocols that can be used in a VoIP system such as: IPSec, Internet Key Exchange, TLS and Datagram Transport Layer Security (DTLS), Secure Shell, PGP and DNSSec, SRTP, and ZRTP (Alan and David, 2006). Although VoIP uses codecs to digitize voice, the digital output contains much lower information density than text data as discussed in Section 3 (Chung-Ping and Kuo, 2005; Talevski et al., 2007). Encrypting a large number of packets will increase the processing overhead and delay thus affecting the quality of voice. In a mobile wireless context, using handheld devices, this effect is more pronounced. The added processing overhead not only affects the delay but also has direct implications on the power exhaustion in already power-constrained handheld devices.

Secure Real-Time Transport Protocol (SRTP) is one of the most popular security mechanisms used to secure media streams in VoIP applications. SRTP has a very low overhead and it is the secure version of the traditional RTP/RTCP protocol which is mainly used for the real-time transmission of multimedia over IP. SRTP provides confidentiality, integrity, authentication and replay protection for RTP and RTCP traffic. However, evaluating SRTP performance in both protecting the RTCP traffic and cryptographic functions other than the encryption is out of the scope of this paper. In SRTP, AES in counter or f8 mode and HMAC-SHA-1 are the predefined algorithms for encryption and authentication, respectively (Baugher et al., 2004).

This paper proposes an encoder-independent reduced processing encryption method when compared to other VoIP securing methods, such as SRTP. This reduction in processing has a positive impact on the voice encryption latency and on the power consumption in handheld devices, while guaranteeing an end-to-end confidentiality.

The rest of the paper is organized as follow: Section 2 discusses the literature that covers VoIP security and multimedia networking-friendly cryptographic methods. Section 3 presents the proposed design, Energy Efficient Voice over IP Privacy (E^2VoIP^2), and the analyses of its strengths and weaknesses. Section 4 describes the implementation of E^2VoIP^2 and analyzes the experimental results. Finally, Section 5 concludes with a summary of the paper.

2. Related work

In this section we cover two aspects of previous work: (1) VoIP security and (2) multimedia networking-friendly security approaches.

2.1. *VoIP security*

VoIP is prone to different kinds of attacks; some are similar to the privacy threats encountered in the PSTN, others emanate from the IP network, and a few are related to networking protocols that VoIP makes use of (Patrick and Miguel, 2006; Butcher et al., 2007)

The confidentiality of the VoIP conversation can be achieved through several protocols. Barbieri (2002) note that the increase in the packet size due to IPSec headers wastes bandwidth. A compression scheme for IPSec headers that is based on cRTP to improve the effective bandwidth is proposed. This compression scheme renders the compressed header only 2% larger than the plain VoIP.

Ravi et al. (2002) reveal a gap between the computational requirements for securing wireless data transactions and the trends in processing capabilities of embedded processors used in wireless handheld devices. Passito et al. (2005) evaluate speech quality of VoIP traffic (of an OpenH323 Project) over an IEEE 802.11b network with VPN/IPSec. It was found that throughput was reduced by 40%, the delay increased due to the crypto engine and the overhead led to higher packet loss rate. Zeng et al. (2004) enumerate relevant constraints in networks such as delay, power, protocol friendliness, network adaptation, and end-to-end security implications. Because multimedia data exhibits several unique characteristics, like high data rate, power-hunger, time constraints and loss tolerance, Zeng et al. (2004) list various network-friendly security approaches, such as selective encryption, lightweight cryptographic techniques, joint encryption and compression and format-compliant selective encryption/scrambling.

A very detailed security analysis of VoIP was performed by Dantu et al. (2009). They addressed the various protocols operating at different layers in VoIP structure along with vulnerabilities and security considerations associated with each of these protocols. They proposed a high-level VoIP security architecture that provides various security features at different layers. A similar work at the signaling level was done in which three possible denial of service attack scenarios were performed against SIP signaling protocol and analyzed (Ehlert et al., 2010). It was shown that this type of attacks can have full or partial countermeasures.

Palmieri and Fiore (2009) addressed the most well known vulnerabilities in VoIP structure and proposed an end-to-end security solution for voice communication to provide both confidentiality and integrity for both signaling and media streams. The approach introduced a hybrid security solution with symmetric encryption of voice streams and asymmetric cryptographic operations to provide the keys and integrity. The proposed mechanism is tested and evaluated in terms of packet loss and MOS. The processing time for separate SRTP actions was also measured but only on the bare-pc softphone. Results showed that authentication is more time consuming than encryption. However, they have assumed in their conclusion that SRTP has a very low overhead on voice quality. It is also worth noting that negligible variations occurred when different key sizes for both authentication and encryption were used (Alexander et al., 2009).