Available online at www.sciencedirect.com
ScienceDirect
Computer Communications 29 (2006) 2759–2769
computer communications
www.elsevier.com/locate/comcom

Some common attacks against certified email protocols and the countermeasures ☆

Min-Hua Shao [a,*], Guilin Wang [b], Jianying Zhou [b]

[a] *Department of Management Information Systems, National Pingtung University of Science and Technology, No. 1, Shuehfu Rd., Neipu, Pingtung 91201, Taiwan*

[b] *Infocomm Security Department, Institute for Infocomm Research, 21 Heng Mui Keng Terrace, Singapore 119613, Singapore*

Available online 6 December 2005

Abstract

Certified email is a value-added service for standard email systems, which guarantees the fairness, i.e., the intended recipient gets the mail content if and only if the mail originator receives a non-repudiation receipt showing that the message has been received by the recipient. As far as security is concerned, fairness is the most important requirements, though some other properties are also desirable in practice. Recently, a number of certified email protocols have been proposed. However, most of those schemes have more or less weaknesses and/or security flaws. In the worst case, fairness cannot be achieved since one dishonest party can mount some attacks to cheat the honest party such that the latter cannot get the expected items. In this paper, we analyze two latest certified email protocols to demonstrate some common attacks, and then propose some improvements to avoid those security problems. We further give several informal but useful guidelines to counter those common attacks in the design of certified email protocols.

Keywords: Certified email; Fair exchange; Non-repudiation; Digital signature; Security protocol

1. Introduction

The lack of evidence for message receipt is a missing piece of the infrastructure required for the more professional use of email [27]. Certified email uses the notion of a digitally signed receipt and strengthens the binding between the evidence and the mail being certified. In other words, the main purpose of a certified email scheme is to achieve the fair exchange of a message and an irrefutable receipt in the sense that either the sender obtains the receipt from the receiver and the receiver accesses the content of the email simultaneously, or neither party gets the expected item. This undeniable but publicly verifiable receipt serves as an evidence for *non-repudiation of receipt* (NRR). Namely, if the receiver denies having received the delivered message from the sender, the sender can provide NRR evidence to an arbitrator to show that this claim is untrue. Some email schemes also provide evidences for *non-repudiation of origin* (NRO) [15,20,21]. Similarly, NRO evidence protects the receiver from the sender's dishonest denial that he/she did not deliver a particular message to this receiver, though this is the fact. We remark that certified email schemes supporting NRO service are functionally equivalent to non-repudiation protocols [34,22,19].

Although fairness is probably the most important requirement for certified email, there are other requirements to be considered as well. Specifically, certified email schemes are usually designed to satisfy some or all of the following properties [34,6,15,17,26,29]:

☆ Part of this work appeared in [30]. The first author's work was done during her visit to Institute for Infocomm Research, Singapore, and funded by the National Science Council of Taiwan under the contract of NSC 93-2917-I-009-001.

* Corresponding author. Tel.: +886 8 7703202; fax: +886 8 7740306.
E-mail addresses: mhshao@alumni.nccu.edu.tw (M.-H. Shao), glwang@i2r.a-star.edu.sg (G. Wang), jyzhou@i2r.a-star.edu.sg (J. Zhou).
URLs: http://www.i2r.a-star.edu.sg/icsd (G. Wang), http://www.i2r.a-star.edu.sg/icsd (J. Zhou).

0140-3664/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.comcom.2005.10.027

- *Fairness*. The protocol should be *fair* in the sense that either each party receives the expected item or neither party receives any useful information about the other's item.
- *Non-repudiation*. Neither the sender nor the receiver of a message is able to deny the transmission.
- *Timeliness*. Both the sender and the receiver of a message should have the ability to reach the end of a protocol run in a finite amount of time unilaterally without losing fairness.
- *Authenticity*. The players should be guaranteed of their reciprocal identity.
- *Confidentiality*. Besides the two parties involved, anybody else (including the trusted third party) cannot get access to the content of the delivered email message.
- *Integrity*. Message transmission should be protected against unauthorized operations in order to guaranty the correctness and authenticity of data.
- *Temporal authentication*. The sender can obtain the evidence to prove the time at which the message was sent.
- *Sending receipt*. The sender can obtain the evidence to prove that he/she started the process of sending a certified email.

Certified email has been discussed for years, and there are two major classes of schemes to address this technical problem: schemes that require the existence of a *trusted third party* (TTP), and schemes that do not require the existence of a TTP. Oppliger [27] showed clearly that the second class, i.e., either based on a simultaneous secret exchange or trusted system is inappropriate to provide certified mail services for the Internet. Specifically, there are two major disadvantages in the schemes based on a simultaneous secret exchange. One is that they are highly interactive between participants, and the other is that those schemes are based on an unrealistic assumption: both of participants have equal or very similar computing power. As for certified mail schemes based on trusted systems, their usefulness and security are often overestimated. Therefore, the use of TTPs seems advantageous and various types of TTPs can be considered according to their involvement in the certified email protocol: schemes with *in-line* TTPs [8], schemes with *on-line* TTPs [34,13,2,27], and schemes with *off-line* TTPs [3,35,15,20,21,24,7,9].

An in-line TTP acts as a delivery authority that involves in each message's transmission during the protocol. The main advantage of in-line TTPs for certified mail is to ensure strong fairness since the TTP collects all information necessary before forwarding them to the concerned entities; and further, the in-line TTP has full control over the message flows and likely provides the sender with services of temporal authentication and anonymity. However, it also implies a communication and computation bottleneck due to the heavy involvement of the TTP.

An improvement to reduce the TTP's involvement is the use of an on-line TTP. The on-line TTP is actively involved during each session of the certified email protocol but not at each step of the protocol. Its task may only deal with signaling information, such as cryptographic keys and/or receipts [27]. In the academic literature, there is often an emphasis on reducing the role and the expense of a TTP. Protocols with a light-weight TTP have been proposed. For example, Abadi and Needham [1] proposed an efficient certified email scheme with a light on-line TTP. A key feature of their scheme is not to deploy any public-key infrastructure; and further, Imamoto and Sakurai [20] revised their scheme in order to provide the non-repudiation of origin service.

A big step towards more efficient solutions was the introduction of off-line TTPs. That is, an off-line TTP involves in a protocol only in abnormal cases, e.g., a dishonest party is trying to cheat or the communication channel fails to work. There are two examples of cheating: a recipient does not issue a receipt though he/she has already obtained the message, or an originator deliberately refuses to reveal the full message to the recipient despite he/she indeed got a valid receipt. The TTP can help the victim to achieve a fair result in these abnormal cases. In other words, a fair certified email scheme with an off-line TTP guarantees that cheating is not actually beneficial to the cheater. Therefore, it is expected that such a protocol will be executed normally in most of the time. Based on this observation, the approach using an off-line TTP is also called *optimistic* [4].

In this paper, we review two certified email schemes based on off-line and on-line TTPs, respectively: the GG scheme proposed by Galdi and Giordano [17] and the OS scheme proposed by Oppliger and Stadlin [28]. The GG scheme is an improved optimistic protocol for certified email. Their effort is to introduce a feature of "temporal authentication" into a four-message optimistic protocol. The GG certified email scheme is effective against misbehavior of one of the players in some cases. However, we demonstrate in this paper that it suffers from a few severe security problems so that some of the security properties mentioned above cannot be satisfied. For example, the receiver can get the email content by *replay attacks* even though he/she did not give the sender a receipt of the message.

The OS scheme employs an on-line TTP to keep the asynchronous nature of email and to minimize the need for interaction between the originator and the recipient of a message, and uses dual signatures to cryptographically link the message key to the certified message. However, there are some security problems in the OS scheme as well. For example, the recipient could also get the message key and then access the certified message without issuing a valid receipt. On the other hand, a dishonest originator can obtain a valid receipt such that the corresponding recipient does not get the message at all. Therefore, the OS scheme is actually unfair for both originator and recipient.

After that, we summarily discuss some common security problems happened frequently in existing certified email schemes, including assumption on communication chan-