Certified Electronic Mail

Jianying Zhou and Dieter Gollmann

Department of Computer Science
Royal Holloway, University of London
Egham, Surrey TW20 0EX
United Kingdom
email: {zhou, dieter}@dcs.rhbnc.ac.uk

Abstract. This paper examines certified mail delivery in postal systems and derives the essential requirements that may be met by a service called certified electronic mail. Protocols are presented to demonstrate how various flavours of certified electronic mail services may be implemented.

Keywords: certified electronic mail, communications security, protocol design

1 Introduction

In many areas, 'pen-and-paper' procedures are being replaced by electronic mechanisms which can improve efficiency and provide better service. However, this transfer is not always easy. Much of the tradition, culture, and law that has been developed to provide protection in social contexts cannot readily be adapted to electronic procedures. On the other hand, cryptographic mechanisms can provide a new range of protection services in the electronic world. Indeed, Diffie has argued that communication security is "the transplantation of fundamental social mechanisms from the world of face to face meetings and pen and ink communication into a world of electronic mail, video, conference, electronic funds transfers, electronic data interchange, and, in the not too distant future, digital money and electronic voting" [6].

Electronic mail has today become an important application of computer networks. There are two main non-proprietary electronic mail (email) systems in widespread use, namely the Internet email system, specified in various RFCs, e.g. [1, 4, 10], and X.400, specified in the CCITT X.400 series recommendations [2]. Once email is used as a standard means of communications, it is natural to ask for the electronic equivalent of more specialised services offered by postal agencies. One such service, providing added protection for mail items, is certified mail. In this paper, we examine the requirements certified electronic mail should meet and the features that may distinguish certified electronic mail from other related services. This discussion not only leads to the specification of certified electronic mail protocols, it also highlights general issues in the definition and design of communications security services.

The paper is organized as follows. The next section gives a model of certified mail delivery in postal systems. In Section 3, we discuss a fictitious application of electronic mail and state the essential requirements that should be met by certified electronic mail. The relation between a certified mail service and other security services is also analysed in this section. We propose three versions of a certified electronic mail protocol (CEM) in Section 4, demonstrating a trade-off between the security properties achieved and the cryptographic infrastructure required.

2 Certified Mail

By choosing the name certified electronic mail (CEM) for a communications security service, we invoke the images of a familiar postal service. It is then only fair to ask that the objectives of CEM should not diverge too much from those of its precursor. We therefore start with a look at three kinds of certified mail services provided by the Royal Mail in the United Kingdom.

Recorded Mail This is a traceable delivery service, suitable for sending important documents. It has the following properties:

- provides a certificate of posting as proof that a letter or packet has been posted,
- signature collected on delivery,
- a *Local Call* telephone enquiry centre for confirmation of delivery,
- available with First or Second class post.

Special Delivery This is a guaranteed next day delivery service, suitable for sending urgent documents. It has the following properties:

- guarantees next day delivery by 12.30pm to most UK destinations,
- a *Local Call* telephone enquiry centre for up-to-date status and confirmation of delivery,
- signature collected on delivery.

Registered Mail This is a guaranteed next day delivery service with compensation, suitable for sending urgent documents, money, and valuable items. It has the following properties:

- guarantees next day delivery by 12.30pm to most UK destinations,
- compensation for loss or damage up to a maximum of £500,
- a *Local Call* telephone enquiry centre for up-to-date status and confirmation of delivery,
- signature collected on delivery.

In the above services, if nobody is available to receive the item, a card is left to advise the recipient that a mail item can be collected from the nearest Royal Mail delivery office. If the item is not collected within 3 weeks (1 week for Recorded mail items), it will be returned to the sender. The sender can track the progress of delivery by making a phone call to the enquiry centre or by paying an extra fee for automatic information.

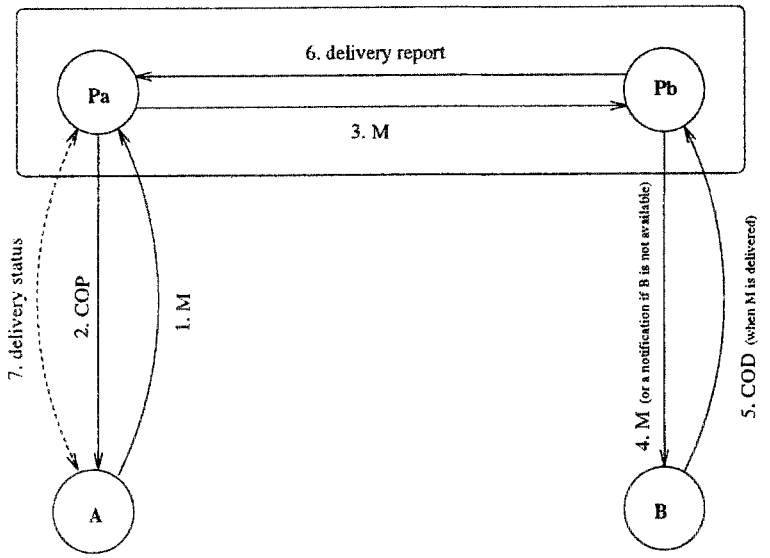

Fig. 1. Schematic Description of Certified Mail

A schematic description of certified mail is given in Figure 1. The sender A submits its mail M at a local post office P_a. P_a gives a certificate of posting (COP) to A, then sends A's mail to the recipient B's local post office P_b. P_b delivers A's mail to B and collects B's signature on a certificate of delivery (COD), or leaves a message asking B to collect its mail. Delivery is a face to face transaction between P_b and B, therefore the exchange of the mail and the COD with B's signature is guaranteed. A delivery report could be sent back from P_b to P_a by request. If necessary, A may make an enquiry to P_a about the delivery status of its mail which could be in delivery, awaiting collection, delivered with COD, non-deliverable, or lost.

The security properties guaranteed by certified mail are limited. Integrity and confidentiality of a mail item are only protected by placing the mail item in an envelope. Certified mail itself will not detect manipulation of a mail item. Usually, the sender and recipient have to combine to establish such an incidence.

Certified mail establishes a contractual relationship between the sender and the postal agent. A certificate of posting proves that the postal agent undertook to deliver a particular mail item. However, the content of the mail item is not verified in any way by the postal agent, so the sender cannot use *COP* as a proof of submission. The certificate of delivery proves that a particular mail item was delivered to the recipient. Again, *COD* cannot be interpreted as evidence sufficient to establish non-repudiation of receipt of the contents of the mail item.

3 Requirements

In postal systems, ordinary mail items get delayed or lost for many reasons. For important mail items, we thus may use certified mail which provides

- report of delivery, and
- compensation for lost mail items.

Mail sent over computer networks may also be lost or mis-routed due to network faults or system crashes. In current electronic mail systems, such as the Internet mail system, if a mail item fails to be delivered, it will be returned. If a mail item is delayed, the sender will receive a warning message. In X.400, the sender can specifically ask to be notified about the successful delivery of a mail item. This option is usually switched off because of the extra communication burden. Compensation for lost mail is not part of these services. As mail systems can, and will, copy mail items, compensation would not refer to the loss of a physical item but to a situation where the mail service lost track of an item and failed to advise the sender in time. We can formulate the essential requirements of CEM as:

> *Report of delivery and compensation for lost mail* — The sender collects evidence that the delivery agent undertook to deliver the mail item and that the mail item was delivered to the intended recipient. The delivery agent can also discharge its duty to the sender by giving a notification before an agreed deadline that the mail could not be delivered.

3.1 A Fictitious Application of CEM

A fictitious application may help to demonstrate the purpose of a certified electronic mail (CEM) service. A software house A sells its products over the Internet. A potential buyer B enquires about one of A's products. A ships an inspection copy. What could happen next?

- The inspection copy arrives intact, B installs it, tests it, is satisfied, and places an order with A.
- The delivery agent employed by A returns the mail as undeliverable. If A suspects a problem with the delivery agent, it resends its inspection copy via a more competent service provider.

- The inspection copy never arrives. B turns to a different supplier and A has lost a business opportunity.
- The inspection copy arrives in a corrupted state. (Maybe, one of A's competitors has replaced it with a particularly lousy piece of code.) Again, B may be unhappy and turn to a different supplier.

Evidently, A should protect itself against

- loss of the mail item containing its software, and
- corruption of that mail item.

Certified mail addresses the first threat. If a mail item has been lost by the delivery agent, i.e. if A has not been informed about the status of delivery before a given deadline, A can claim compensation.

3.2 Relation to Other Security Services

Let us now consider integrity protection. Of course, this service could be tied in with CEM. However, it is definitely not necessary to do so. In our fictitious application, the sender A could attach its signature to the software shipped. B then can detect changes to the software and verify that it came from A. The delivery agent has no role to play in this verification. Even more, a delivery agent that works incorrectly will not corrupt the integrity service. (Messages may be corrupted, but corrupted messages will not be accepted as genuine.)

Equally, confidentiality protection need not be an aspect of certified mail. In the electronic world, it could be provided by a separate service, like PEM (Privacy Enhanced Mail) [7] or MOSS (MIME Object Security Services) [5].

We can also distinguish a certified mail service from a service for non-repudiation of origin and receipt. Non-repudiation of origin and receipt prevent entities from denying that they have sent or received certain messages [12]. This is a relation between the originator and the recipient, referring to the content of a message. Trusted third parties only help in establishing non-repudiation evidence and settling disputes. In contrast, a certified mail service is offered by the postal agent to the mail sender to guarantee the delivery of a mail to its destination. This is a relation between the mail sender and the delivery agent, without any reference to the content of the message. The mail recipient only helps the delivery agent to prove to the mail sender that the mail was delivered.

In electronic commerce, we may want a payment on delivery service. NetBill [3] defines *certified delivery* as an atomic method to guarantee payment when electronic goods are delivered. A trusted third party, the NetBill server, is involved in this protocol. Both merchant and client have a contractual relationship with the NetBill server and the protocol combines a non-repudiation service with a delivery service so that the client can obtain an item only when merchant and client are committed to that transaction.

4 Certified Electronic Mail

When designing a certified electronic mail protocol, a few important differences between postal delivery and electronic mail have to be noted. In postal systems, mail items exist physically and value is attached to the mail item itself. Postmen deliver the mail to the recipient in a face to face meeting. In computer networks, mail items can exist in several copies and senders are less concerned about the loss of such a copy than about timely delivery. Furthermore, when an electronic mail message is delivered, the recipient may be reluctant to acknowledge the message after having read its content. If this is deemed to be a problem, direct delivery has to be replaced by some other process to establish a certified electronic mail service.

The security elements built into X.400 [2] enable the provision of a number of different security services. The elements which may be of relevance for a certified electronic mail service are *proof of submission, non-repudiation of submission, proof of delivery,* and *non-repudiation of delivery.* Proof of delivery, or its stronger version non-repudiation of delivery, allows the originator of a message to obtain from the recipient a proof that the message has been delivered unaltered. This is more than originally required from certified mail. We will consider a weaker service that provides delivery reports referring to message labels rather than to the message contents.

In store and forward systems like X.400, messages are delivered by the message transfer system first to a message store (MS) and may later be retrieved by a user agent (UA). We thus can distinguish between services that report on delivery to a message store and those that require evidence from the user agent. A certified electronic mail service may use both options. As pointed out in [8, 9], proof of delivery to an MS is not possible if the message content is encrypted (unless the MS is given the UA's private key) because the message receipt has to be generated at the point in time when the message is delivered to the MS and is required to include a signature of the cleartext message content, which is only available to an entity possessing the recipient's private key. Our certified mail protocol avoids this problem as it only refers to message labels.

4.1 Notation

In our electronic model of certified mail (see Figure 2), A and B are mailbox programs rather than human users. $P_1, \ldots, P_n$ are a set of network mail servers acting as delivery agents to relay messages between A and B. The notation in the protocol description is as follows.

- X, Y: concatenation of two messages X and Y.
- $[X]$: message X is optional.
- $H(X)$: a one-way hash function of message X.
- $sK(X)$: digital signature of message X with the private key K.

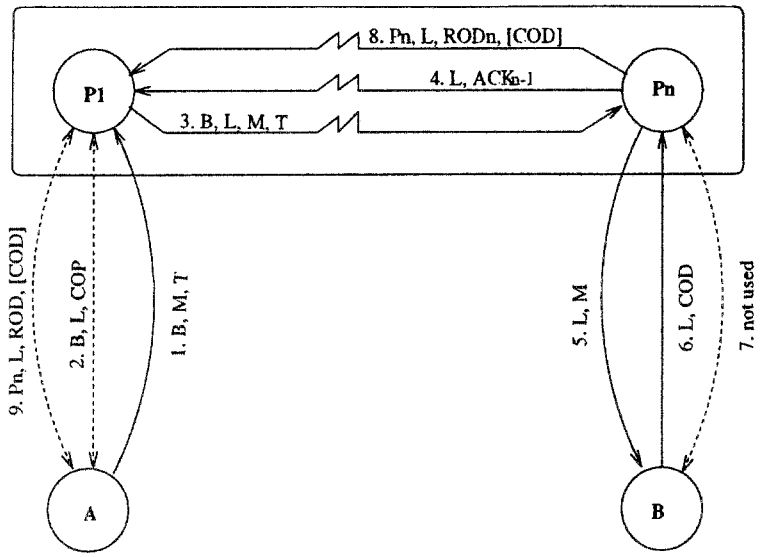

Fig. 2. Protocol CEM

- S_A: the private signature key of principal A.
- $A \rightarrow B : X$: principal A sends message X to principal B.
- $A \leftrightarrow B : X$: principal A pulls message X from principal B.
- M: mail item sent from A to B.
- L: a unique label chosen by P_1 to identify mail item M.
- T: deadline by which a delivery report or a non-deliverable notification should be available, chosen by A and agreed by P_1.
- $COP = sS_{P_1}(B, L, M, T)$: Certificate of Posting for M. In addition to the service offered by ordinary mail where COP and M do not have a direct link, the message is included in the signature. Thus, the mail sender can check whether the delivery agent acknowledged the submitted mail. It also includes a deadline for providing delivery status.
- $ACK_i = sS_{P_{i+1}}(P_i, B, L)$: Acknowledgement of the labeled mail item from P_i, $i = 1, \ldots, n-1$.
- $COD = sS_B(P_n, L, M)$: Certificate of Delivery of the labeled mail item.
- $ROD_j = sS_{P_j}(P_n, L)$: Report of Delivery of the labeled mail item. We use ROD for the Report of Delivery generated by P_1.

4.2 Cryptographic Infrastructure

Our protocol employs digital signatures to generate certificates like COP, COD, and ROD. We assume that all users and mail servers have the ability to digitally sign messages. The mail servers have to be able to verify signatures of neighbouring mail servers and of the users they deliver mail to. Users only need to be able to verify signatures of the mail servers they submit mail to but do

not necessarily have verification keys for the parties they send mail to. A user wanting to query a report of delivery may thus have to invoke another service to check the certificate of delivery signed by the recipient.

4.3 The Protocol

Protocol CEM has four phases. Steps 2 and 9 are *"ftp get"* operations [11]. We assume that network failures are not permanent and that every message will *eventually* arrive at the intended recipient. If A and B are in the domain of the same mail server, i.e. $n = 1$, Steps 3, 4 and 8 will be omitted.

(1) Submission:
   1. $A \rightarrow P_1 : B, M, T$
   2. $A \leftrightarrow P_1 : B, L, COP$

(2) Relay: $i = 1, \ldots, n-1$
   3. $P_i \rightarrow P_{i+1} : B, L, M, T$
   4. $P_{i+1} \rightarrow P_i : L, ACK_i$

(3) Delivery:
   5. $P_n \rightarrow B : L, M$
   6. $B \rightarrow P_n : L, COD$
   7. not used

(4) Confirmation:
   8. $P_n \rightarrow P_1 : P_n, L, ROD_n, [COD]$
   9. $A \leftrightarrow P_1 : P_n, L, ROD, [COD]$

This protocol can be elaborated depending on the application, e.g. by setting a deadline to limit the time mail items and delivery reports are available to the public so that the delivery agents do not need to store all mail items and delivery reports forever. We now describe the steps of protocol CEM in more detail.

1. $A \rightarrow P_1$: A sends the mail item, the name of the recipient and the delivery deadline to P_1. A could resend this message until it has retrieved the certificate of posting from P_1.
2. $A \leftrightarrow P_1$: If P_1 agrees with the delivery deadline, P_1 issues the certificate of posting with a unique label L and puts it in a publicly readable directory. A can fetch COP to check whether P_1 received the mail unchanged. If the check fails, A can resend the mail.
3. $P_i \rightarrow P_{i+1}$: P_i relays the mail item, the unique label, the name of the recipient and the delivery deadline to P_{i+1}. P_i could resend this message until it has received an acknowledgement from P_{i+1}.
4. $P_{i+1} \rightarrow P_i$: P_{i+1} acknowledges the labeled mail item received from P_i. If P_i does not receive this acknowledgement within its timeout period, it will inform A that the mail could not be delivered.
5. $P_n \rightarrow B$: P_n sends the mail item M and the label L to B.
6. $B \rightarrow P_n$: B returns the signed certificate of delivery COD to P_n. If P_n does not receive COD within its timeout period, it will inform A that the mail could not be delivered.

8. $P_n \rightarrow P_1$: A report of delivery, and optionally *COD* are relayed back from P_n to P_1. Then P_1 will store its *ROD*, and optionally *COD*, in a publicly readable directory. If P_1 does not receive the report of delivery before the deadline T, it will inform A that the mail could not be delivered.
9. $A \leftrightarrow P_1$: A fetches *ROD*, and optionally *COD* from P_1. The delivery agent is responsible for maintaining consistency among *COP*, *ROD* and *COD*.

In this protocol, the delivery agent either notifies A that the mail item could not be delivered or indicates delivery by presenting a report of delivery. The certificate of delivery signed by B can be included as supporting evidence. The delivery agent can do so only if the mail had actually been delivered but A would need B's verification key to check the link between *ROD* and *COD*. Such a check would also detect delivery of a corrupted mail item. The recipient has no way of verifying where the message came from or whether it was modified during transport.

As the recipient sees the message before signing the certificate of delivery, the protocol suffers from the *selective receipt* problem, a problem that also exists in X.400 security services like proof of delivery and non-repudiation of delivery. In our fictitious application, the following course of events is possible. The buyer B likes the software but does not want to pay for it. So, B does not acknowledge receipt. After a timeout period, the delivery agent informs A that its mail could not be delivered. B hopes that A will not check B's premises later on.

4.4 CEM-nsr: No Selective Receipt

The following CEM version prevents selective receipt by changing the certificate of delivery to $COD = sS_B(P_n, L)$, and the delivery phase to:

5. $P_n \rightarrow B : L$
6. $B \rightarrow P_n : L, COD$
7. $B \leftrightarrow P_n : L, M$

In this version, P_n informs B that a mail item labeled L is awaiting collection. We assume that the delivery agent places mail item M in a publicly readable directory after receiving *COD* from B. Now, B can retrieve the mail item. If the mail item is put into the directory before, there is no protection against selective receipt. If the mail item is not placed in the directory at all, then B does not get the mail item. When B is unable to retrieve the message, it will take steps beyond the protocol to alert the user community about its communications problems with that particular delivery agent.

The recipient still cannot verify the integrity of the mail fetched from the delivery agent. The mail sender obtains evidence that the recipient acknowledged a notification for a mail item with the correct label but has no evidence that the recipient did indeed receive the correct mail.

4.5 CEM-ip: No Selective Receipt with Integrity Protection

When the sender A wants B to acknowledge that the correct item was received, A could use a service providing non-repudiation of receipt. Such a service requires further security infrastructure to be in place, and may thus be quite expensive. Hence, A could settle for less and only ask for evidence that the correct mail item was ready for collection by B. There is still no guarantee that B will collect this message. Given that CEM is an arrangement between A and the delivery agent, this requirement is reasonable. The delivery agent has no authority over B and may be reluctant to pay compensation if B is at fault. Again, integrity protection inserted by A may alert B when it has retrieved a wrong message.

To provide the mail sender with sufficient information for integrity verification while avoiding the selective receipt problem, we change the certificate of delivery to $COD = sS_B(P_n, L, H(L, M))$.

5. $P_n \rightarrow B : L, H(L, M)$
6. $B \rightarrow P_n : L, COD$
7. $B \leftrightarrow P_n : L, M$

In Step 5, the delivery agent sends $H(L, M)$ with the unique label L. The recipient can use $H(L, M)$ to verify the mail fetched at Step 7. Alternatively, P_n could send M encrypted under a key K in Step 5, release the key in Step 7, and include the key in its report of delivery. However, we prefer the version using a cryptographic hash function because of its lower computational overheads.

When the delivery agent supplies a wrong message M' with hash value $H(L, M')$, B will accept M' as genuine. However, a check of COD will detect that B was notified about a wrong message and the sender can take steps beyond the protocol to rectify the situation. When the delivery agent supplies M' with $H(L, M)$, B will note the discrepancy and retry to fetch M. If this fails persistently, or if the message cannot be retrieved at all, B will take steps beyond the protocol to alert the user community about its communications problems with that particular delivery agent.

When a bogus delivery agent has sent $L, H(L, M')$ to B and B responds with $COD = sS_B(P_n, L, H(L, M'))$, P_n will note the discrepancy and A will not be notified about the success of delivery. However, B will not detect that it retrieves the wrong message M'. When B replies to the notifications $L, H(L, M)$ from P_n and $L, H(L, M')$ from a bogus delivery agent, B cannot determine which notification was genuine. As far as P_n and A are concerned, the message M was delivered but B may wrongly retrieve M'.

It is not difficult for B to protect itself against such a situation. Whenever B is notified about a mail with label L, B will reject any further notifications with the same label until it has retrieved the corresponding mail item. Of course, genuine delivery agents may be concerned about bogus competitors and offer

integrity protection mechanisms of their own. For example, P_n could sign the notification to B. However, this implies that B is able to verify signatures and is in possession of a certified verification key for P_n. As a matter of fact, B will require those keys for all agents that may deliver certified mail to B.

4.6 Compensation

The delivery agent agrees to compensate A if A has a *COP* with label L and the delivery agent fails to supply a *ROD* and *COD* with label L for the correct message, if applicable, signed by the intended receiver within an agreed period of time. Timestamping of *ROD* is outside the scope of this protocol. Instead of *ROD* and *COD*, the delivery agent can also supply a notification that the mail could not be delivered. Of course, the delivery agent can always create such a notification before the deadline. In our fictitious application, A then at least knows about the delivery problem and can use another delivery agent.

The delivery agent ought to be reluctant to *guarantee* to deliver A's mail. It would have to rely on A to specify B's address correctly and on B to be ready to receive mail. Such a guarantee only makes sense if B is controlled by the delivery agent to the extent that the delivery agent can check the correctness of B's address and guarantee that B will be ready. Typically, such a situation arises when B is a message store under the delivery agent's control.

5 Conclusion

There are three aspects to this paper. First and foremost, it uses certified electronic mail in an attempt to show how to properly identify requirements for a secure communication service. It is often quite tempting to add any reasonable protection requirements to the list of intended features. There is no doubt that such an all encompassing service may be useful, but it will quite likely also be rather expensive. Every new encryption, every new signature demands appropriate management of the new keys introduced to the system. The more cryptographic functions we use, the more elaborate our security infrastructure will be. Given that the customers will have to pay for this infrastructure, we should place our services where it is most cost effective. This observation may be obvious, but it is worth repeating.

Secondly, we have defined CEM to protect only against loss and delayed delivery of mail items. We decided to treat integrity protection and non-repudiation of receipt as different services, which may be offered in conjunction with CEM, but are not an integral part of CEM. If we only ask for a report of delivery, then present electronic mail systems already include CEM as an option.

Finally, we have given protocols to demonstrate how various flavours of CEM may be implemented. We observe a trade-off between the guarantees the sender

can obtain and the assumptions it wants to make about the behaviour of the recipient and of the delivery agent. Protocol CEM does not make any assumptions about the delivery agent, but the recipient may decide not to acknowledge a message it has received. Protocols CEM-nsr and CEM-ip avoid the selective receipt problem but rely on the delivery agent to make messages available once the recipient has acknowledged the corresponding notification.

Acknowledgements

We have benefited from discussions with colleagues in the Information Security Group. Comments from Chris Mitchell and Wenbo Mao on the draft of this paper are appreciated. We are also grateful to the anonymous referees for their important suggestions for improvements to this paper. The first author would like to thank the British Government and the K C Wong Education Foundation for their support through an ORS Award and a K C Wong Scholarship.

References

1. N. Borenstein and N. Freed. *Multipurpose Internet mail extensions (MIME).* RFC 1521, September 1993.
2. CCITT. *Recommendation X.400: Message handling system and service overview.* November 1988.
3. B. Cox, J. D. Tygar and M. Sirbu. *NetBill security and transaction protocol.* Proceedings of the First USENIX Workshop on Electronic Commerce, July 1995.
4. D. Crocker. *Standard for the format of ARPA Internet text messages.* RFC 822, August 1982.
5. S. Crocker, N. Freed, J. Galvin and S. Murphy. *MIME object security services.* RFC 1848, October 1995.
6. W. Diffie. *The impact of a secret cryptographic standard on encryption, privacy, law enforcement and technology.* Hearings before the Subcommittee on Telecommunications and Finance of the Committee on Energy and Commerce, House of Representatives, One Hundred Third Congress, First Session, Serial No. 103–53, pages 111–116, April 29 and June 9, 1993.
7. S. T. Kent. *Internet privacy enhanced mail.* Communications of the ACM, 36(8):48–60, August 1993.
8. C. J. Mitchell, M. Walker and D. Rush. *CCITT/ISO standards for secure message handling.* IEEE Journal on Selected Areas in Communications, 7(4):517–524, May 1989.
9. C. J. Mitchell, D. Rush and M. Walker. *A secure message architecture implementing the X.400-1988 security features.* The Computer Journal, 33(4):290–295, 1990.
10. J. B. Postel. *Simple mail transfer protocol.* RFC 821, August 1982.
11. J. B. Postel and J. K. Reynolds. *File transfer protocol.* RFC 959, October 1985.
12. J. Zhou and D. Gollmann. *A fair non-repudiation protocol.* Proceedings of the 1996 IEEE Symposium on Security and Privacy, pages 55–61, Oakland, CA, May 1996.