Secure Communicating Systems

Design, Analysis, and Implementation

Michael R A Huth

Imperial College of Science, Technology and Medicine

CAMBRIDGE
UNIVERSITY PRESS

PUBLISHED BY THE PRESS SYNDICATE OF THE UNIVERSITY OF CAMBRIDGE
The Pitt Building, Trumpington Street, Cambridge, United Kingdom

CAMBRIDGE UNIVERSITY PRESS
The Edinburgh Building, Cambridge CB2 2RU, UK
40 West 20th Street, New York, NY 10011-4211, USA
10 Stamford Road, Oakleigh, VIC 3166, Australia
Ruiz de Alarcón 13, 28014 Madrid, Spain
Dock House, The Waterfront, Cape Town 8001, South Africa

http://www.cambridge.org

First published 2001

Printed in the United States of America

Typeface Times 10.5/13 pt. *System* AMS-TEX [FH]

A catalog record for this book is available from the British Library.

Library of Congress Cataloging in Publication Data
Huth, Michael, 1962–
Secure communicating systems : design, analysis, and implementation / Michael R A Huth.
p. cm.
Includes bibliographical references and index.
ISBN 0-521-80731-X
1. Telecommunication – Security measures.

TK5102.85.H88 2001
005.8 – dc21 2001025484

ISBN 0 521 80731 X hardback

Contents

Preface *page* vii
Acknowledgments xi

1 Secure Communication in Modern Information Societies 1
1.1 Electronic Commerce: The Mantra of Y2K+ 1
1.2 Cryptographic Systems 3
1.3 Legislating Electronic Authentication 6
1.4 The Mathematical Judge 9
1.5 Encryption Policies 10
1.6 Trust and Communities 11
1.7 Bibliographic Notes 13

2 Public-Key Cryptography 15
2.1 Specification of RSA 17
2.2 A Realization of PKCs: RSA 23
2.3 Generating Large Primes 27
2.4 Correctness of RSA 59
2.5 Security of RSA 64
2.6 Integer Factorization 73
2.7 Other Key-Exchange Realizations Based on Discrete Logarithms 76
2.8 Bibliographic Notes 79

3 Symmetric-Key Cryptography 81
3.1 Stream Ciphers 81
3.2 Block Ciphers 95
3.3 Bibliographic Notes 130

4 Security Protocol Design and Analysis 131
4.1 Digital Signatures 131
4.2 Secure Log-In Protocols 142
4.3 Authentication Revisited 149
4.4 Secret-Sharing Protocols 153
4.5 Model Checking Security Protocol Designs 156
4.6 Bibliographic Notes 178

5 Optimal Public-Key Encryption with RSA 179
5.1 A Simple Semantically Secure Encryption 180
5.2 A Plain-Text–Aware Encryption 182

5.3 The Random Oracle Methodology 186
5.4 Exact Security for the Simple Encryption 189
5.5 Exact Security for the Plain-Text–Aware Encryption 199
5.6 Bibliographic Notes 203

6 Analysis of Secure Information Flow 204
6.1 Motivation 204
6.2 A Type System for Analysis of Secure Information Flow 207
6.3 A Semantic Approach to Analysis of Secure Information Flow 227
6.4 Program Certification 255
6.5 Covert Channels 256
6.6 Bibliographic Notes 257

Appendix Primitive Roots 259
A.1 Existence of Primitive Roots 259
A.2 Computing Primitive Roots 269

Bibliography 271

Index 275

CHAPTER 1

Secure Communication in Modern Information Societies

1.1 ELECTRONIC COMMERCE: THE MANTRA OF Y2K+

We are presently witnessing mergers and takeovers of unprecedented speed and extent between companies once thought to have national identities, or at least clearly identifiable lines of products or services. On the day this paragraph was written, the British Vodaphone AirTouch announced an Internet alliance with the French conglomerate Vivendi. The deal was conditional on Vodaphone's hostile takeover of Germany's Mannesmann and, in the end, did establish a branded multi-access portal in Europe. About a week later, the takeover of Mannesmann was official – the biggest ever, and friendly. MCI's attempted takeover of Sprint is another example of a strategically advantageous combination of different information technologies. January 2000 saw CNN, NTV, and the *Deutsche Handelsblatt* (a direct competitor to the *Financial Times*) launch a multimedia product for stock market news that is accessible via television, printed newspapers, and the World Wide Web. And so it goes. Although many differing views are held regarding the causes and consequences of these phenomena, we would probably all agree that they reflect a certain shift of emphasis from production-based economics to one grounded in the processing, marketing, and *access* of information. Whether the products themselves are merely "information" or systems for managing and processing vast amounts of data, information systems are seen as a crucial strategic means for organizing, improving, and maintaining more traditional production cycles.

Such a shift could not have been achieved without the creation of reliable, dense, and global electronic information networks that offer the full spectrum of accessibility modes that conventional information carriers allow. This spectrum ranges from being open to the general public (e.g., a public library) to being open only to members of a very well-defined community (e.g., the NASA engineers who develop the next generation of shuttle thrusters). The Internet and the World Wide Web have become a key medium for the storage, transmission, transformation, and analysis of information of any kind: textual, visual, or auditory. Recently, we even witnessed the release of a device that "interprets" olfactory information transmitted over the Internet! Apparently, we increasingly participate in – and depend on – electronically networked communities. This raises societal and managerial questions pertaining to the rights and responsibilities of network participants. However, it is not clear a priori whether standard practices from offline communities adequately transfer to so-called virtual communities and electronic communication networks. For example, children's bookstores and pornographic shops are typically found at disjoint locations in real cities, whereas such an exclusion principle is hardly implementable on the Internet; this renders online protection and guidance of minors an unresolved issue.

Regulatory efforts, which are mostly confined to sovereign states and trade unions, have little hope of success in a truly global environment unless their legal and moral force is recognized, and enforced, worldwide.

Today's digital networks are adopting an abundance of newly developed information technology tools that facilitate the gathering and creation of meaningful information needed for successful business ventures; yet these tools also provide a platform for *conducting* business. The fashionable term "electronic commerce" denotes any kind of commercial activity that occurs over the World Wide Web, the Internet, intranets, facsimile, telephone, and so forth. Electronic commerce is believed to have the greatest growth rates in any economic sector. E-commerce start-ups are enthusiastically received, and almost indiscriminately so, by investors. As a result, individuals who can install or maintain information systems for e-commerce are much in demand. However, the promises of electronic commerce must be weighed against their possible dangers and inherent challenges.

1. The *locality* and *authenticity* of electronically communicating agents is dubious at best; electronic business interactions make it harder to guarantee that potential business partners are honest about who and where they are.
2. Sensitive information or other private data may be transmitted through unreliable or otherwise *unsecure communication channels.* Not only does this pose a threat in that competitors may be able to access and use confidential strategic or technical information, it also raises grave concerns about the *privacy of individuals* who use those very channels for noncommercial (yet still nonpublic) communications.
3. Even if electronic transactions came equipped with a mechanism of authenticating agents, one needs to ensure that agents cannot subsequently deny any of their properly authenticated actions. We speak of *nonrepudiation* if an authentication scheme has this desirable property.
4. The right to anonymous actions has held an important role in securing free speech and unhindered political discourse. Although mechanisms that implement anonymous interaction may also be subject to serious abuse, they are an important component of democratic processes. Most patents on digital cash realize such electronic cash in an anonymous way. However, the financial services sector (including tax agencies) are quite interested in removing this anonymity feature of such cash, at which point the issue becomes not merely technical but also one of politics, policies, and laws.
5. "The devil is in the implementation" – this means that a secure specification of a cryptographic system (or security-handling computer program) is still a long way from its actual secure implementation.
6. Mobile code, active networks, and extensible operation system kernels require: novel methodologies for specifying safety rules for executing programs that are foreign to the local system; provably correct algorithms for verifying that programs meet such safety specifications; and mechanisms that attach certificates to mobile code so that these certificates can quickly be evaluated locally.

These are only a few (and by no means the most critical) problems that electronic commerce faces. Even if all had acceptable solutions, a host of other pressing questions would remain unanswered. For example, how should businesses protect the integrity, existence, and control of their information systems? – given that they may be distributed globally and have plenty of interfaces to publicly accessible resources. There is also the daunting

task of designing working frameworks for the taxation of Internet sales, given the conflicting interests of stakeholders: local counties, states in a federation, sovereign states, e-commerce companies, and consumers. Guaranteeing privacy of communication and authenticity of agents may be of little use if unauthorized and presumably hostile network agents are able to penetrate the heart of a company's information system. Federal agents recently managed to enter, without proper authorization, sites that are vital to the security of U.S. national infrastructures. We all have read stories of the so-called hackers who gained access to computers of the U.S. Department of Defense and thereby downloaded huge amounts of sensitive data during the initial phase of Operation Desert Storm. Computer security cases in the military sector are not out of place in this section, for defense agencies rely on electronic purchasing and ordering procedures that are increasingly required to interface with the nonmilitary commercial world. At present, it is unclear what the psychological and sociological effects and implications will be of making electronic commerce a main mode of entrepreneurial activity, but the events of May 2000 have already demonstrated the threat that e-mail viruses and worms pose to an economy that depends more and more on the Internet and the World Wide Web. It is not the objective of this text to address these pressing issues; rather, it focuses solely on the six points previously listed. Specifically, we give an introduction to *secure communicating systems* by studying the design, analysis, and implementation of systems that are built to provide solutions to the practical problems of (a) certifying the safety rules of programs, (b) realizing the authentication of secure and perhaps anonymous communication along an open channel, and (c) the nonrepudiation of committed (trans)actions.

1.2 CRYPTOGRAPHIC SYSTEMS

Although cryptology has a rather long history and is a thriving field of sophisticated research, in this text we give only a selective overview by choosing representative designs of cryptographic systems and some forms of their analysis that are accessible to senior undergraduate and beginning graduate students. To be up-front about it, there is an inherent and deplorable tradeoff between the degree to which cryptographic systems realize their stated security goals and the computational overhead they impose on information networks.[1] More often than not, such security goals are left implicit or are formulated with insufficient precision, as the discussion of authentication in Section 4.3 illustrates. Perfectly secure mechanisms for ensuring private communication along a channel are possible; the one-time pad (see page 86), while being perfectly secure, requires an encryption key that is as long as the actual message to be communicated. This burden hardly justifies its use unless perfect security is a minimum requirement, as for the "hotline" between the White House and the Kremlin. More efficient systems don't have such perfect security, so one needs to assess just how secure they are. In concrete terms, such security is often measured in how much money, or time, one would have to spend in order to "break"[2] a cryptographic system; unfortunately, such estimates may only be meaningful

[1] There is an even more disconcerting tradeoff between the security of a communicating system and the convenience of its user-level functionality.

[2] Breaking a system can mean a variety of things: obtaining access to a single message (or fragment thereof) with or without control over which message that should be; corrupting the entire security of the system for an extended period of time, with or without its legal users noticing the break-in; being able to assume someone else's identity; etc.

for a specific method of breaking a system. A useful measure should thus provide cost predictions for *all possible attacks*, independent of whether they are known to the analyst. Evidently, this can only be realized in a very limited manner. This also entails a reasonably clear understanding of how secure the respective communication and authentication components *must be*. Such a quantitative requirement analysis is usually quite difficult; for example, the monetary value of a company's customer database is typically hard to assess and may be a function of who would gain access to it. And how would *you* quantify the loss of privacy if your medical records were to be posted on the World Wide Web?

We mention these issues in passing but more often assess the *computational effort* needed to break certain cryptographic systems. A fundamental difficulty with such analyses is that they must consider some (mathematical) model of the cryptographic system under consideration, or even a specific implementation thereof. Any positive security results drawn from such an analysis are therefore only valid *within the given model or implementation*. Alas, this does not rule out an attack *outside the given model*; the well-publicized attack of RSA encryption implemented on a smartcard is one such alarming example (see pages 68 and 204). In an extreme view, one may even consider such results as helping potential attackers by pointing out to them what sorts of things *won't* succeed; it is wise to assume that attackers read the relevant technical literature.

You may be surprised to hear that the bulk of cryptographic systems make use of rather astonishing facts about natural numbers and some of their computational problems. Thus we need to study a certain amount of number theory and get to know a few important number-theoretic algorithms that form fundamental components of real cryptographic systems. We hasten to point out that we aim to develop such material at a graceful pace and at an accessible level.[3] In this chapter, we mention the role of number theory in cryptography because all the cryptographic systems that use certain "hard" number-theoretical problems – for realizing secure communication, authentication, or nonrepudiation – rest their security on the premise that such hard problems don't have easy solutions. The point is that this premise's validity is still an open (and most difficult) research problem and moreover that even its validity would usually not *ensure* security.

Because this text will not develop the rather advanced concepts required for a precise definition of what "hard" and "easy" problems are, we mean to illustrate this via example. Integer factorization is believed to be a hard problem, and the security of the RSA cryptosystem relies on this belief (see Section 2.5). More specifically, it is believed to be computationally infeasible to find a factor of an integer with 1024 binary digits if that number is the product of two randomly generated primes of about equal size. (Improvements in processor speed and cheaper computer parts, such as memory, may require a future increase in the number of bits needed.) Yet to this day, nobody has put forward any proof of this belief. It is conceivable that somebody will eventually devise an efficient procedure for factoring such large numbers. Similar concerns (and lack of proof) prevail for other "hard" problems used in building cryptographic systems, whether they are grounded in number theory or some other computational structures.

[3] Appendix A may be skipped entirely without compromising the appreciation of our cryptographic designs, but it does fill the explanatory gap of proving the correctness of the Miller–Rabin algorithm for primality testing, one of the "workhorses" in our cryptographic toolbox.

Even if such (unlikely) proofs were to be found, they could only be carried out *relative to a computational model*, such as a conventional personal computer. This means that their resulting safeguards would only apply to that very same computational model. However, various computing paradigms may be vastly different in nature from each other. Some, admittedly small, instances of certain "hard" problems have been solved using chemical reactions based on the processing of DNA. We already have seen computers with up to four states, where computation is driven by the laws of quantum mechanics. If – and that is a big "if" – the development of such machines is scalable in the number of states, then this will provide an efficient engine for factoring large integers. It is debatable whether any of these approaches might pose a real threat to existing cryptographic systems, but only time can tell. In June 2000, a Swiss research team used entanglement of photons[4] to transport an encrypted message from one town to another through ordinary fiber-optic lines. A U.S. team is currently investigating how one can make it harder for eavesdroppers to alter the properties of photons. A German–Austrian team has used such techniques to encrypt an image. This news is exciting, but it also suggests that new technology may only provide new instantiations for familiar players, such as eavesdroppers. It is also unclear whether such technology can be used on large networks that intend to reach ordinary households. It seems rather disturbing (perhaps pleasing, to some) that the realization of electronic commerce and the protection of vital national infrastructures – which rely on secured information systems – may depend on facts about number theory, microbiology, and quantum physics.

Cryptographic components, even if assumed to be perfectly secure as isolated components, raise novel security questions if placed within the context of *interacting networks*. For example, can a *security protocol* be successfully attacked even though none of its cryptographic primitives can be broken in isolation? Indeed, quite a few published protocols were found to have undergone such attacks. Such insights gave rise to research activity similar to that in the design and analysis of concurrency protocols. We therefore present a customized framework for "debugging" security protocols in Section 4.5. Again, such tools are certainly needed by implementors and designers of security protocols; if *they* don't do their homework then attackers will do it for them – and let them know by attacking weaknesses discovered with the aid of those tools.

This point illustrates another peculiarity in the study of cryptographic systems. Historically, such designs (say, a particular encryption algorithm) were kept secret, and knowing the design was often coextensive to knowing how to break it. All such early systems were broken eventually. A conceptual breakthrough was the idea of *key-dependent cryptosystems*. Ideally, such systems are secure even if one knows all the intricate details of their design – as long as one does not know the concrete key with which the system was instantiated. This idea made it possible to publish designs so that the entire scientific community could study and attack them. Although this development can only improve the strength of emerging designs, it takes time for such studies to be of any substantial value. It is fair to

[4] Quantum computing rests on three principles: (i) *superposition* of quantum bits allows for an exponential speed-up factor for certain computations (including the factorization of integers); (ii) *quantum entanglement* enables a reliable and instantaneous communication of quantum bits over arbitrarily long distances; and (iii) *quantum interference* poses the challenge of engineering a system of quantum bits that does not interfere with its environment (decoherence).

say that the Data Encryption Algorithm (featured in Section 3.2.1) and the RSA encryption system (presented in Section 2.2) underwent more than twenty years of public analysis and scrutiny without revealing any fundamental design weaknesses. More recent cryptosystems and cryptographic algorithms, such as the new Advanced Encryption Standard *Rijndael*, may well be far superior to the previous ones, but again only time can tell because we have no single sound and coherent mathematical theory or methodology for reasoning about the strength of such systems. This places consumers and standards committees alike in an awkward position. When and why should one abandon a given cryptographic system in favor of another? If a cryptographic standard is fully implemented and integrated into other network standards, what can be done if the cryptographic design turns out to have serious flaws? Note that this is not just an engineering problem of replacing one system with a different (and, it is hoped, more secure) one, since sensitive data will have been stored in an unsecure manner. This raises several thorny issues, not the least of which is liability.

At the time of this writing, it is anticipated that the Data Encryption Standard (DES) will be replaced by the Advanced Encryption Standard (AES), the cipher Rijndael, which is featured in Section 3.2.2. On 2 October 2000, the U.S. Department of Commerce announced Rijndael as the winner of a worldwide design contest. Pending a period of public comment and final approval, this cipher will become a standard of the U.S. National Institute of Standards and Technology. That the submissions came from all over the world already suggests that national standards and their overseeing national agencies may need to rethink their roles and begin to interface with similar bodies of other nations. It may well be that global economic conglomerates will put pressure on governments to streamline regulation and licensing activities toward standard business practices and to offer approaches that are fairly uniform on a global scale. Indeed, recent policy changes at the White House regarding the export control of U.S. encryption products indicate that governments have already begun to think along those lines. These changes worry national agencies that deal with issues of defense and the protection of vital national infrastructures. We return to the dilemma of encryption policies in Section 1.5.

1.3 LEGISLATING ELECTRONIC AUTHENTICATION

More and more, the Internet and other electronic media provide a platform for ordering products, negotiating contracts, and paying for rendered or anticipated services. Thus consumers, government agencies, and commercial sectors wonder whether there is a need for new legislation that elaborates in which cases, and to what extent, electronic signatures are legally valid. Unfortunately, technical terminology is often misunderstood by legislative bodies, and technicians who consult in a legislative effort find it equally hard to appreciate the legal language. Needless to say, it is crucial that these communities work together in realizing a maximum of clarity in the legislative process. For example, there seems to be some confusion between the concepts of an *electronic signature* and a *digital signature.* The former can be thought of as any technical replacement of the usual handwritten signature functionality in an electronic system: digital pens, PIN numbers, and scanned hand-written signatures are a few examples. In some sense, digital signatures are a special case of electronic signatures in that they use public-key cryptosystems (the topic of Chapter 2) as a mechanism for ensuring the integrity and origin of digital messages; Section 4.1 discusses digital signatures in detail. In another sense, digital signatures are more

appropriately thought of as *digital envelopes*, for the signer may not know, or endorse, the signed message. Upon closer inspection, digital signatures have a much broader range of applications than (electronic) signatures in the narrow sense. Digital signatures can be used to authenticate servers in a computing network, web pages, software, or any data that is stored digitally.

Legislators may take a *technical* approach – declaring, for example, a specific digital signature system as a (possibly required) standard for implementing certain electronic authentication functions. This view generally provides no insights into the legal consequences of using, or misusing, such systems. One of the first laws on digital signatures, the German Digital Signature Law, used a legal instrument to set a technical standard: specifically, for the required security of the public-key infrastructures. The law does not explicitly state any legal consequences that would result from using digital signature systems that are compliant with the standard prescribed by the law.

A *legal* approach, on the other hand, attempts to equate handwritten and electronic signatures and may not impose any restrictions as to which technology may realize electronic signature systems. The Utah Digital Signature Act of 1995 regulates digital signatures based on public-key cryptosystems and legally equates such digital signatures with handwritten ones, provided that the corresponding cryptosystem meets all the requirements described in the Act.[5] The State of Utah has a common law system that often allows a more liberal interpretation of the use of signatures; expressing one's intentions explicitly, for example, may be considered "signing". Unfortunately, the Utah Digital Signature Act does not adequately reflect the different functions of signatures. This kind of law could threaten the development and growth of electronic commerce in that it also identifies functions of handwritten signatures with novel digital functions, such as certifying a web server.

In practice, most (draft) law and directives present a mixture of these approaches, thereby creating both legal uncertainty and possible impediments to the evolution of electronic commerce. The United Nations Commission on International Trade Law (UNCITRAL) crafted Draft Uniform Rules on Electronic Signatures; these rules would be nonbinding and technologically nonspecific, but they would provide guidance to legislative authorities during their own process of designing legislation for electronic authentication. These rules distinguish between "electronic signatures" and "*enhanced* electronic signatures"; the latter must meet a higher standard of security with regard to the signing and signature verification process. It is assumed that data signed with enhanced electronic signatures are legally signed. The EU Directive of the European Parliament and of the Council on a Common Framework for Electronic Signatures gives similar open-ended definitions for an "electronic signature" and for what is now called an "*advanced* electronic signature"; however, the Directive focuses on digital signatures and does not provide legal recognition of electronic signatures pertaining to the validity of contracts requiring signatures. The CA Working Group of the Electronic Commerce Promotion Council of Japan issued guidelines for the operation and management of certification authorities (CAs), an infrastructure used to establish a notion of trust in the authenticity of public keys. This is an example of a *self-regulated* effort, where one hopes that industry will establish common practice in accord with such guidelines.

[5] At the time of this writing, nobody has come forward to register a public-key cryptosystem under this Act.

In the past, one could observe a preference for technology-specific legislation that most often dealt with digital signature systems. The Italian Digital Document Regulations of 10 November 1997 state that, under certain conditions, digital signatures can be legally equated with handwritten signatures. At the same time, these regulations are restricted to public-key cryptosystems with public-key infrastructures used for digital signature systems. The prevalence of a mixed approach is largely due to the fact that digital signature systems are the basis of important tools for electronic commerce: Pretty Good Privacy (PGP), Secure Electronic Transactions (SET), and Secure Socket Layer (SSL) all make crucial use of such technology.

Policymakers often think that the success of electronic commerce depends on having a well-specified technical signature system with well-understood legal consequences. This wishful thinking stands in direct opposition to new technological developments and the need for novel signature roles that electronic commerce is likely to bring about. A variety of alternative approaches to electronic signatures exist already. Virtual Credit Card (VCC), used by the Brazilian bank Unibanco, electronically authorizes credit-card purchases without using the public-key infrastructures (PKIs) upon which digital signature systems rely. Another example is iPIN, an Internet-based payment system for small amounts that can be managed by Internet service providers.

On 30 June 2000, President Clinton signed the Electronic Signatures in Global and National Commerce Act, a bill that recognizes and clarifies the legal status of electronic signatures. This bill requires consumers to agree to electronically signed contracts; they also must consent to receiving records over the Internet. Companies, on the other hand, must verify that customers have a viable e-mail address and the necessary equipment to receive electronic information.

There are a number of biometric approaches to electronic authentication. The idea is to authenticate individuals by means – it is hoped – of dependably unique biological data. For example, fingerprint readers on small chips can be integrated into keyboards, and one may scan a person's iris or palm at an automatic teller machine. It is unclear whether biometrics can replace, or even supplement, cheaper authentication mechanisms that don't rely on biological data. Because useful biometric data ought to remain fixed during a person's lifetime, such information may have to be considered as *personal property* in the legal sense. At any rate, the handling of such data requires reliable legal frameworks that protect the privacy and identity of individuals.

The examples just given show that regulatory efforts need to reflect the possibility of swift and dramatic technological changes. The downside of technology-neutral legislation is that courts may have to develop case law when such legislation cannot achieve a precise definition of legal concepts. Another source of tension is that one country's national law often conflicts with other national (or international) law. The UNCITRAL Model Law on Electronic Commerce was drafted within the larger context of achieving a more uniform and cohesive international trade law; it is technologically nonspecific, thus allowing and anticipating fast and dramatic technological changes. International legislation must also make room for flexible interpretations of legal requirements of form; for example, common law and civic law systems typically offer different interpretations of "legally binding signatures".

Since electronic commerce is, by its very nature, an international phenomenon, we need drafts and guidelines for digital law at an international level. The pressing need for legal

clarity, however, requires national legislation, as this can be enacted much sooner. Additionally, nations may have an inherent cultural and historical outlook on legal concepts. Laws about handwritten signatures, for instance, may emphasize the signer's intention to be legally bound by his or her signature (often the case in common law, as in the United States), or it may stress the security of the actual signing process (often occurring in civic law, as in Germany). When nations draft new digital law, they may also have to "clean up" and streamline some of their existing law. At the time of this writing, a handwritten signature on a document transmitted via facsimile (fax) is legally binding in the Netherlands but not so in Germany. Nations and unions may also have a different view of privacy and civil rights and of their implementation in systems that support electronic commerce.

In the meantime, it appears that legislation should largely be nonspecific about technological details of electronic authentication. It should pay considerable attention to the various functions and features of handwritten and electronic signatures, making clear if and how such functional roles allow for a match between electronic and nonelectronic signatures. This legislative process needs to be internationally oriented but must also reflect the specific intent and nature of national law. Clearly, these objectives have inherent conflicts. It is hoped that a more mature electronic commerce will also see a slower technological change of authentication mechanisms in order for technology-specific legislation to be effective. Whether one believes that legislation (hard law) is necessary or that self-regulation (soft law) – or some combination of both – is needed to aid and oversee the development of electronic commerce, it is evident that these problems require an unprecedented degree of cooperation among technicians, government and nongovernment organizations, industry executives, and legislative bodies. This provides one of the many reasons why computer science professionals and students ought to be informed about the basic concepts, designs, modes of analysis, and implementations of cryptographic systems.

1.4 THE MATHEMATICAL JUDGE

Regardless of whether a security protocol or its cryptographic primitives are secure or not, they will typically be sold and used as a commercial product. So far, software vendors have generally not been liable for flawed software, provided that they could show that they followed established "software engineering practice". However, it is not clear whether such a line of argument will continue to be successful if software erroneously confirms or denies the authenticity of a contract signature, or if it exposes confidential information resulting in physical, monetary, or psychological harm to the sender or receiver. For example, what about cases in which agents sign data electronically and later claim that the signature has been forged? Even if the signature system had a built-in nonrepudiation mechanism, the agent could still claim that its implementation was somehow flawed. Using a digital signature scheme, the agent could also claim that somebody obtained her private signature key – say, by corrupting the public-key infrastructure or some certification authority. Even if the protocol adds more and more protective layers against such possibilities, the agent could always contest the functioning of the *lowest* or at least *some* level. This is in striking contrast to the traditional practice of using pens and handwritten signatures. We can hardly blame the company that manufactured a pen used by someone else to forge our signature! Likewise, we cannot sensibly assert that somebody

acquired the knowledge and skill of reproducing our original signature perfectly. Consequently, the question of establishing the circumstances under which electronically signed documents will be recognized in court as legally binding is more delicate than one may initially suppose.

In the technical part of this text, we see that basically all practical cryptographic systems come with an inherent degree of unsecurity, even if we were to assume a flawless implementation process. Admittedly, the likelihood of a security violation occurring in a perfect implementation may be extremely small, but can we establish a definite threshold saying that a digital signature scheme is legally binding if the probability for the claimed signer *not* to have signed a document using this scheme is smaller than some $\varepsilon > 0$? Who will come up with such a value? Who will assess a given implementation of a cryptographic system to estimate that threshold? Who will certify that the concrete implementations of such abstract digital signature cryptosystems meet all the relevant security specifications? If, say, RSA were used for such a certified signature generation scheme, then how would a jury react to defense lawyers exposing jurors to popular-science and technical articles that describe the occasional success story of "breaking" a large RSA key? Would the jury not feel uneasy about resting their judgment on conflicting presentations on the security of key lengths? And would a substantial number of future court cases require a *mathematical judge*?

Although it may be somewhat of a stretch, electronic signatures could conceivably become key evidence in first-degree murder cases. One may recall that prosecutors have a hard time convincing juries when their only hard piece of evidence is a sample of nonmitochondrial DNA, found at a crime scene, with a "close" match to the DNA of a defendant. Jurors find it difficult to relate sophisticated scientific facts to the concept of "beyond reasonable doubt".

To play devil's advocate, suppose one has legislation that endorses a specific technology and a specific implementation for a digital signature scheme and also states explicitly the legal consequences of electronic signatures produced with the system it describes. Suppose further that, after some time, this implementation turns out to have serious flaws. Who would deal with the long case list of past system users who now contest having signed their mortgages and car loans? It seems that one might have to rely on higher implementation standards than those for software used on commercial aircraft – but meeting such standards is expensive and time-consuming. A more sensible approach may be to make the implementation and verification effort a function of the importance of the data that the tool is intended to sign. Clearly, a system that handles only small-scale transactions requires less effort than one that deals with major stock trading. Even so, the former could see class-action lawsuits by consumer groups and the like. Perhaps car loans and other big-ticket items will still rely, at least partially, on traditional signing methods and evidence provided by the particular (nonelectronic) business context. At the risk of repeating ourselves, only time can tell how people and other agents will sign what – and how successful courts will be in using electronic signatures as hard evidence.

1.5 ENCRYPTION POLICIES

The economic promises of global electronic commerce and its need for uniform interfaces suggest that support for reliable and secure cryptographic components should be available

worldwide; nonetheless, some governments impose restrictions on the use, import, or export of such products. This largely occurs in the context of cryptographic systems used to render text unintelligible to everybody except the sender and receiver of the message. Obviously, such capabilities pose threats to national interests; they can make it hard or impossible for law-enforcement agencies to conduct investigations or to gain convictions; and they can affect national security if used to cover up terrorist activity. They also can facilitate extortion schemes: former or current employees of some company or agency may encrypt important data and then demand money from their employer for making the data legible again. But let us not forget that the same tools that aid terrorists are also instrumental in protecting the privacy and confidentiality of people's speech and their lawful participation in democracies – not to mention the protection this technology offers to pro-democracy activists in certain parts of the world. This is clearly a political point of friction that will not go away, but the interests of democratic movements and existing democracies must not be taken lightly.

The reference to the Crypto Law Survey (given in the bibliographical notes to this chapter, Section 1.7), provides an excellent resource for finding out what nations apply what sorts of encryption control at present. The current U.S. government went through an interesting learning process that caused it to change its encryption export policies. Interestingly enough, digital *signature* systems were never controlled in this manner in the United States. Encryption systems for functions other than signing, formerly classified as *ammunition*, can now be exported (after a technical review) to commercial firms and other nongovernment end users unless they reside in states named on the U.S. State Department's evolving list of supporters of terrorism. If the key-length of the cryptosystem is longer than 64 bits – which is true of the new AES Rijndael – then the vendor may be required to submit a post-export report that is facilitated by reflecting standard industrial practice. Foreign nationals no longer need a license if they want to work for U.S. firms on the development and maintenance of cryptosystems. Fortunately, the idea of mandatory recovery keys (which would have allowed the authorized decryption of text even if the keyholder refuses to hand over the key) seems to have been abandoned, much to the dismay of U.S. agencies concerned with national security. For details, see the press release of the U.S. Department of Commerce dated 12 January 2000.[6] Encryption policies have their own dilemmas. They must be strong enough to adequately protect law enforcement and national security but at the same time liberal enough to maintain or improve a nation's political structures and processes – as well as its competitiveness in the lucrative global market of electronic security products and resulting e-commerce. This may well be the principal reason why the U.S. government solicits public comments on these regulations for 120 days before final revised policy rules are implemented.

1.6 TRUST AND COMMUNITIES

Today, we witness a fierce global economy with large multinational conglomerates that encourage governments to provide incentives for setting up shop within their territory. For example, the German car manufacturer BMW let European states "bid" for hosting

[6] http://www.bxa.doc.gov/Encryption/regs.htm

their new production facility. AOL Europe asked the German government to enact policies that would lower the base access rate to the Internet within Germany, identifying current rates as a major obstacle to the growth of German e-commerce. Major companies nervously try to find strategic partners that complement and strengthen their competitiveness worldwide. The World Trade Organization (WTO) may see China as a future member, and worldwide free trade and mobility seem within reach. At the same time, however, international, national, and regional interest groups actively campaign against the possibly harmful sociological, environmental, and economic implications of increasingly global production and management structures. The riots at the WTO meeting in Seattle (United States) and the voices of protest at the last World Economic Forum in Davos (Switzerland) are indicative of such concerns. Through meetings such as the Davos forum, top executives are beginning to appreciate that the concerns of communities are a serious component of their managerial decision processes. The customer boycott of Shell in Europe, triggered by Shell's plan to dump a polluted oil rig in the North Sea, suggests that consumer values can affect company policies.[7] The Internet and other digital communication technologies give traditional and emerging communities a powerful tool for reaching their constituency and other affected groups they mean to impact; these technologies also enable the creation of novel interest groups and communities at a speed and to an extent that were previously impossible.

All these communities, even the ones based on business relationships, critically depend on working notions of *trust*. This may seem ironic, considering that the current economic climate conjures up images of Manchester Capitalism. However, even the most aggressive and hostile parties depend on some form of trust if they want to communicate at all. Vodaphone AirTouch placed considerable trust in the publicly available reports issued by Mannesmann regarding its financial performance and marketing goals. If you were to apply for admission into the graduate school at Tulane University and then received mail – on 100% cotton paper emblazoned with the crest of Tulane University of Louisiana – informing you of your acceptance or rejection, you would trust that this mail is coming from that university, *all things being equal*.

Such trust has practical advantages; it would simply be impossible to be "perfectly paranoid" and still maintain a productive and meaningful life. We tend to question trust when all things are *not* equal! – as when your bank inspects your signature more closely on a check for $10,000 than on one for $10. In the rapidly evolving realm of electronic commerce, we have seen attempts to provide business websites with stamps of approval given by some generally trusted certification or accreditation company. TRUSTe[8] is one such (nonprofit) service provider; its certification vouches for certain privacy policies that consumers can expect to be met. However, companies are often hesitant to attain such a certification; among other things, clearly stated privacy policies open the door to lawsuits if the company violates those policies. In July 2000, there were alleged cases of failed e-commerce businesses that – in order to appease creditors – sold private consumer data in violation of company policy.

The need for trust evidently poses a dilemma for implementing systems that hold any value at all, be they production facilities, information systems, or strategic centers such as

[7] www.ens.lycos.com/ens/nov98/1998L-11-27-03.html
[8] www.truste.org/

the NATO headquarters. The widespread use of mobile code (e.g., by accessing active web pages) also implies trusting that the evaluation of foreign code on a local system does not compromise the security or safety rules of that local system. Even if such code is authenticated prior to its execution, we still have to trust its execution behavior. *Proof-carrying code* – though for now a mere research topic – has the potential to provide a platform for the specification of local safety rules, the verification that programs meet these rules, a means of communicating this fact by attaching a certificate to code, and an efficient way of checking such certificates. One may then confine the need of trust to those aspects that are not expressed or implied by the formally specified safety policy.

The design and use of cryptographic systems does not dispense with such security-threatening needs. Digital signature systems were invented to eliminate the need to trust a third party with the job of delivering a secret key from one agent to another. Ironically, and not surprisingly, this solution created a new need for trust. Such systems have no mechanism for certifying that the public key, which an agent advertises as belonging to him, actually is associated with that agent. The protocol attack described on page 22 illustrates the need for third parties that vouch for such correct matchings of agents and their keys. Commercial products realize this through certification authorities, a "web of trust", or other public-key infrastructures. In that sense, cryptographic systems render the same dilemma of possibly extreme needs for protection and security and a concurrent need for trust. We believe that this dilemma cannot be entirely resolved qualitatively, but only to certain degrees. As D. Denning put it so aptly in her statement before the Subcommittee on Courts and Intellectual Property (Committee on the Judiciary, U.S. House of Representatives) regarding the Security and Freedom Through Encryption Act: "In short, encryption is no silver bullet." The reader of this text will be well advised to keep this in mind.

1.7 BIBLIOGRAPHIC NOTES

A good descriptive account of the shift from production-based to access-based economies has been given by Rifkin (2000). Denning (1999) discusses information systems in general, provides a systematic exposition of their threats, and competently presents possible strategies (and their tradeoffs) for countering a possible corruption of their security. Her website "The Cryptography Project",[9] contains well-organized and topical material on national and international encryption policies. Schneier (2000) gives an entertaining and revealing analysis of information security in the networked world. Also recommended is B.-J. Koops' Crypto Law Survey,[10] an up-to-date discussion of legislation pertaining to cryptographic systems that protect information against unauthorized access. The details on U.S. encryption policy given in Section 1.5 of this chapter reflect the Fact Sheet issued on 16 September 1999 by the Office of the Press Secretary of The White House and the press release of the U.S. Department of Commerce from 12 January 2000.[11] B. P. Aalberts and S. van der Hof have conducted an analysis of legislative approaches to electronic authentication, providing evidence that the emphasis on digital signature schemes

[9] www.cosc.georgetown.edu/~denning/crypto/index.html
[10] http://cwis.kub.nl/~frw/people/koops/lawsurvy.htm
[11] http://www.bxa.doc.gov/Encryption/regs.htm

may impede the growth and progress of electronic commerce and increase legal uncertainty;[12] Section 1.3 largely draws from that work. The books by Negroponte (1995) and Roszak (1994) represent two rather extreme – and opposing – positions regarding the role of information technology in modern societies. Denning and Lin (1994) present a compact but rich overview of the moral and legal challenges that come with the participation and management of (electronically) networked communities. For a discussion of the security features of the Java programming language, see McGraw and Felten (1997). Last, but not least, M. Curtin's website[13] contains a nice survey on "Snake oil warning sign: Encryption software to avoid".

[12] http://cwis.kub.nl/~frw/people/hof/ds-fr.htm
[13] http://www.interhack.net/people/cmcurtin/snake-oil-faq.html