PART Three

In the final section of the book, I cover three themes: politics, management, and assurance. Given that we now have some idea how to provide protection, the three big questions are: what are you allowed to do? how do you go about organizing it? and how do you know when you're done?

There has been much public debate recently about whether cryptography should be controlled in the interests of law enforcement. The evolution of U.S. and European law and policy on cryptography makes an interesting tale, but I'll cover it at some speed. It's only the tip of an iceberg.

Other places at which security engineering is coming into conflict with politics abound. In what circumstances should legal recognition be given to digital signatures? What sort of mechanisms are feasible to protect people from inappropriate material on the Web (and who's to say what's inappropriate in any case)? What are the implications for commercial system designers of the threat of "information warfare" by hostile powers or substate groups? And how will individual privacy be protected? This last question is being answered quite differently in the United States and Europe. In the former, it's left to corporate "self-regulation," while in the latter, the experience of World War II has led to privacy being entrenched as a constutional principle. "Data protection," as it's called in Europe, threatens a major ruction between the two continents. Successive U.S. administrations have tended to see privacy as something on which "a deal could be done" or that could be fudged or just swept under the carpet—not realizing that Germans can be as inflexible on data protection as many Americans are on gun control.

Our next chapter is about management. This has become a dirty word in the information security world; there are endless vapid articles written in "managementese" that say nothing at great length. But management issues are important. Organizational and economic incentives often determine whether secure systems get built. A large number of systems have failed because the protection was tacked on as an afterthought, or because the real purpose of the system was not its advertised purpose, or because the people who controlled the system design were not the people who suffered when it failed. Economics provides a number of insights; for example, security engineers often work with imperfect information, and network externalities are particularly savage. The management of residual risk, and the retention of organizational know-how, are two of the other problems that frequently cause expensive failures.

Assurance is a huge political can of worms. On the face of it, it's just an engineering issue. How do you go about finding convincing answers to the questions: are we building the right system? and, are we building it right? These questions are familiar from software engineering (which can teach us a lot), but they acquire new meaning when systems are exposed to hostile attack. Also, most of the organizational structures within which assurance claims can be made, or certified, are poisoned one way or another. Claims about system security properties are often thinly veiled assertions of power and control, so it should surprise no one if the results of evaluation by equipment makers, insurers' laboratories, military agencies, and academic attackers are very different. So it's really important for the security engineer to set out at the start of a project not just what the objective is, but the criteria by which it will be judged a success or a failure.

CHAPTER

21

E-Policy

Experience should teach us to be most on our guard to protect liberty when the government's purposes are beneficient.... The greatest dangers to liberty lurk in insidious encroachment by men of zeal, well meaning but without understanding.

—SUPREME COURT JUSTICE LOUIS BRANDEIS

The arguments of lawyers and engineers pass through one another like angry ghosts.

—NICK BOHM, BRIAN GLADMAN, AND IAN BROWN [124]

21.1 Introduction

Information security is about power. It's about determining who will be able to grant (or deny) the use of a resource. In the past the implications went largely unexamined and uncontested. Banks built systems that failed in their favor, rather than the customers'; hospitals harvested patient data for management and research, without telling their patients; and governments bullied phone companies into making their networks easy to tap. But since the early 1990s, many of these assumptions have begun to be challenged. Contributory factors include increased public awareness, the greater importance of IT in people's lives and businesses, and the fact that computing power is now more distributed. Cheap computers meant that small businesses could balance their bank account and check their interest calculations, making it harder for the bank manager to quietly add a percent or two to the agreed rate; the contempt that many health insurers and hospitals had for medical privacy was exposed once family doctors started competing for control of the electronic health record; and wiretapping became an issue once ubiquitous PCs, email, and encryption software made it practical for individuals to defeat some kinds of government surveillance.

The role of government in e-commerce and the Internet generally has become a source of much argument. For much of the 1990s, the debate was dominated by *key escrow*—the view (held by the governments of the United States, France, Russia, and, after 1996, Britain) that copies of encryption keys should be given to the government for the convenience of law enforcement and national intelligence agencies. The opposing position— held by the governments of countries such as Germany and Ireland, Britain until 1996, and almost all of the IT industry—was that it was better to leave the development of the Internet to technological and market forces.

I'll delve into some of the arguments shortly. However, at the beginning of the twenty-first century, the places where government policy meets information security are multiplying. How are government services, from welfare payments through the court system to passports and tax collection, to be organized in a society that increasingly expects everything to be available at once and online? How can government avoid deepening the social exclusion of the poor, the old, and ethnic minorities—who may be the last to go online? Should elections be conducted online, and if so what should we do to make fraud, corruption, and coercion at least as hard as they are now? When automating government, will we replace an inefficient tiresome mess with an automated inefficient tiresome mess?

Government departments, like businesses, are struggling to stake out territory in cyberspace. Key escrow was just one of the earliest land grabs, and was unsurprising given that intelligence agencies are among the most technically sophisticated public-sector organizations. An unfortunate side effect is that, in many countries, the debate over who should have access to cryptographic keys has not only soured government relations with local IT communities, but has also enabled the agencies to take a dominant position in IT policy. In Britain, for example, the mantle of "national technical authority" for such matters is worn jealously by CESG, a department of the signals intelligence agency GCHQ. So it's policy that all state sector keys be escrowed. This will raise serious concerns about any online system for elections. Of course the agencies will want to know who votes for Sinn Féin in Northern Ireland; but if it's too easy for them to find out, then the legitimacy of the province's government will be undermined, and this is likely to cause more deaths than any tactical intelligence failure. Many people consider that letting espionage agencies set national computer security policy amounts to putting the fox in charge of the henhouse. But what's the alternative? What sort of political control should be exercised over the agencies, and how are they to be held accountable? Where and how are alternative public-sector centers of expertise in information security (and IT generally) to be built?

So the first policy issue we need to look at is the whole question of how wiretaps, traffic analysis, and cryptography are to be regulated.

21.2 Cryptography Policy

Millions of words have been written in the last few years on cryptography policy and related issues. In this section, all I can reasonably try to do is to place the debate in context, sketch the broad outlines, and provide pointers to primary sources.

Although restrictions on cryptography had existed for years and greatly irritated civilian users such as the banking industry, they shot to the headlines in 1993 when the new administration of Bill Clinton astonished the IT industry with the *Escrowed En-*

cryption Standard (EES), more popularly known as the *Clipper chip*. This was a proposed replacement for DES, with a built-in back-door key that enabled government agencies to decipher any traffic. (I explained the technical aspects in 14.5.3.) However, Clipper is even more important as the issue that politicized cryptography and information security generally.

U.S. opinion polarized with the government taking the view that since cryptography is about keeping messages secret, it could be used by criminals to prevent the police gathering evidence from wiretaps; the IT industry (with a few exceptions) took the conflicting view that cryptography was the only means of protecting electronic commerce, and was thus vital to the future development of the Net. Civil liberties groups lined up with the industry, and claimed that cryptography would be the critical technology for privacy. By 1994, the NSA had concluded that it faced a war with Microsoft, which it would lose, so it handed off the policy lead to the FBI, while continuing to direct matters from behind the scenes.

The debate rapidiy became tangled up with export controls on weapons, the means by which cryptography was traditionally controlled. U.S. software firms were not allowed to export products containing cryptography that was seen as too hard to break (usually interpreted as meaning a keylength of over 40 bits). A.U.S. software author, Phil Zimmermann, was hauled up before a grand jury for arms trafficking after a program he wrote—PGP—"escaped" on to the Internet. He immediately became a folk hero and made a fortune as his product grabbed market leadership. The conflict became international: the U.S. State Department invested significant effort in persuading other countries to control cryptography too.

The results were mixed. Some countries that had oppressive regimes within living memory, such as Germany and Japan, resisted American blandishments. Others, such as Russia, seized the excuse to pass harsh crypto control laws. France relaxed a traditional prohibition on non-government use of crypto; while Britain went from a liberal, laissezfaire policy under John Major in the mid-1990s to a draconian law under Tony Blair in 2000—the *Regulation of Investigatory Powers* (RIP) Act.

Throughout this process, the means of compulsion applied by governments (outside the Russia/Zimbabwe end of the spectrum) have become progressively more subtle. Outright criminalization has given way to a grab-bag of economic and legal incentives. But, overall, the popular view of the crypto policy struggle has been one in which the Forces of Light (privacy advocates and IT companies) have slowly overcome the Forces of Darkness (policemen and spies) in a Manichean struggle for the Soul of the Internet.

Reality is, as always, a bit more complicated. It may be useful to step back and try to place the debate in its historical context.

21.2.1 The History of Police Wiretapping

Since the earliest states arose, their rulers have tried to control communications. In classical times, this was done by checks on couriers at customs posts. From the Middle Ages, many kings either granted a monopoly of postal services to a trusted nobleman or made them the property of the state. The letter-opening and code-breaking facilities of early modern states, the so-called *Black Chambers*, are described in Kahn [428].

The invention of electronic communications brought forth a defensive and indeed atavistic response, one very reminiscent of the recent crypto policy debate. In most of Europe, the telegraph service was set up as part of the Post Office and was always owned by the government. Even where it wasn't, regulation was usually so tight that the industry's growth was severely hampered, leaving America with a clear competitive advantage. A profusion of national rules, which sometimes clashed with each other, so exasperated Europeans that the *International Telegraph Union* (ITU) was set up in 1865 [729]. This didn't satisfy everyone. In Britain, the telegraph industry was nationalized by Gladstone in 1869. (This experience was so traumatic for both government and business that the next significant nationalizations in Britain were not until after 1945.)

The invention of the telephone further increased government interest in surveillance. Resistance, both legal and technical, has a long history. In the United States, the Supreme Court ruled in 1928 in *Olmstead vs. United States* that wiretapping didn't violate the Fourth Amendment provisions on search and seizure, as there was no physical breach of a dwelling; Judge Brandeis famously dissented. In 1967, the Court reversed itself in *Katz vs. United States*, ruling that the amendment protects people, not places. The following year, Congress legalized Federal wiretapping (in Title III of the Omnibus Crime Control and Safe Streets Act) following testimony on the scale of organized crime in the United States. In 1978. following an investigation into the Nixon administration's abuses, Congress passed the Federal Intelligence Surveillance Act (FISA), which places controls on wiretapping for national security. In 1986, the Electronic Communications Protection Act (ECPA) relaxed the Title III warrant provisions. By the early 1990s, the spread of deregulated services, from mobile phones to call forwarding, had started to undermine the authorities' ability to implement wiretaps, as did technical developments such as out-of-band signalling and adaptive echo cancellation in modems. By 1994, the Communications Assistance for Law Enforcement Act (CALEA) required all communications companies to make their networks tappable in ways approved by the FBI. By 1999, over 2,450,000 telephone conversations were legally tapped following 1,350 court orders [272, 533]. The relevant law is 18 USC (US Code) 2510-2521 [759] for telco services. (The Cable Act of 1984 regulates wiretaps for cable modems and is much more restrictive—so the administration wants it watered down [439].)

It must by noted that, according to some serious analysts, there are at least as many unauthorized wiretaps as authorized ones [250]. In some countries the figures can be distorted by wiretapping being uncontrolled if one of the equipment owners consents—so that calls from phone boxes are free to market.

But even if the official figures have to be doubled or tripled, it's still clear that democratic regimes make very much less use of wiretapping than authoritarian ones. For example, lawful wiretapping amounted to 63,243 line-days in the United States in 1999, or an average of just over 173 taps in operation on an average day. The former East Germany had some 25,000 telephone taps in place, despite having a fraction of the U.S. population [295]. There was also extensive use of technical surveillance measures, such as room bugs and body wires. (It's hardly surprising that nudist resorts became extremely popular in that country.)

It's also worth noting that the incidence of wiretapping is highly variable in the developed democracies. In the United States, for example, only about half the states use

it; and for many years, the bulk of the taps were in the "Mafia" states of New York, New Jersey, and Florida (though recently, Pennsylvania and California have caught up) [372]. There is similar variation in Europe. Wiretaps are very common in the Netherlands, despite Dutch liberalism on other issues [147]: they have up to 1,000 taps on the go at once, with a tenth of America's population. In a homicide investigation there, for example, it's routine to tap everyone in the victim's address book for a week to monitor how they react to the news of the death. In Britain, wiretaps are supposed to need a ministerial warrant, and are rarer; but police use bugs and similar techniques quite a lot in serious cases. To some extent, the technologies are interchangeable.

The cost of wiretapping is a serious issue. Before CALEA was introduced, in 1993, U.S. police agencies spent only $51.7 million on wiretaps—perhaps a good estimate of their value before the issue became politicized [372]. The implementation of CALEA has supposedly cost over $500 million, even though it doesn't cover ISPs. This raises some obvious policy questions. Is it worth it? Should agencies cut back on wiretapping, and spend the money on more cops instead? Or will they try to expand its use to amortize their costs? Once you start molding an infrastructure to meet requirements other than cost and efficiency, someone has to pay; and as the infrastructure gets more complex, the bills keep on mounting.

21.2.2 The History of Traffic Analysis

However, the bulk of police communications intelligence in developed democratic countries does not come from the surveillance of content, but from the analysis of telephone toll records and other communications data. I examined in the chapter on telecomms security how criminals go to great lengths to bury their signals in innocuous traffic using techniques such as prepaid mobile phones and PBX hacking, and the techniques used by the police to trace networks of criminal contexts nonetheless.

Again, this is nothing new. Rulers have long used their control over postal services to track the correspondents of potential subversives, even when the letters weren't opened. The introduction of postage stamps in 1840 was an advance for privacy as it made it much easier to send a letter anonymously. Some countries got so worried about the threat of sedition and libel that they passed laws requiring a return address to be written on the back of the envelope. The development of the telegraph, on the other hand, was an advance for surveillance; messages were logged by sender, receiver and word count, so traffic totals could be compiled, and were found to be an effective indicator of economic activity [729]. World War I brought home to the combatants the value of the intelligence that could be gleaned from listening to the volume of enemy radio traffic, even when it couldn't conveniently be deciphered [428, 569]. Later conflicts reinforced this.

By the late twentieth century, traffic analysis provided the bulk of police communications intelligence. For example, in the United States, there were 1,329 wiretap applications approved in 1998 (the last year for which comparable statistics were available at the time of writing), while there were 4,886 warrants (plus 4,621 extensions) for *pen registers* (devices that record all the numbers dialed from a particular phone line) and 2,437 warrants (plus 2,770 extensions) for *trap-and-trace* devices (which record the calling-line ID of incoming calls, even if the caller tries to block it). In other words, there were 11 times as many warrants for communications data as for content. This pattern has been stable for years, and across many countries. Why should this be?

Wiretaps are so expensive to listen to and transcribe that most police forces with restricted budgets use them only as a weapon of last resort; in contrast, the numbers a suspect calls, and that call him, give a rapid overview of his pattern of contacts. Also, while wiretaps usually have fairly strict warrantry requirements, most countries impose few or no restrictions on the police use of communications data. In the United States, no warrants were required until ECPA. Even after that, they have been easy to get: under 18 USC 3123 [759], the investigative officer merely has to certify to the court "that the information likely to be obtained by such installation and use is relevant to an ongoing criminal investigation." This can be any crime—felony or misdemeanor—and under either federal or state law. Unlike with wiretaps, the court has no power to deny a warrant once a formally correct application has been made, and there is no court supervision once the order has been granted. Since the passage of CALEA, warrants are still required for such communications data as the addresses to which the subscriber has sent e-mail messages, but basic toll records can be obtained under subpoena—and the subscriber need not be notified. So the above figures for pen register and trap-and-trace warrants almost certainly understate the extent of law enforcement traffic analysis. In any case, both phone and computer service records can be provided to bodies other than law enforcement agencies under 18 USC 2703(c); thus, for example, we find Virginia and Maryland planning to use mobile phone tracking data to monitor congestion on the Capital Beltway [710]. Toll data use for marketing purposes was also expressly envisioned by Congress when this law was passed.

In Britain, files of telephone toll tickets were provided by the phone company to the police without any control whatsoever until European law forced the government to regulate the practice in the RIP Act in 2000. Since then, comms data requires only a notice from a senior police officer to the phone company or ISP, not a warrant.

The issue of controlling access to communications data is gradually becoming a live one. The major problem is that comms data and content are becoming more and more intermixed, as what's content at one level of abstraction is often comms data at the next. A good example comes from Web URLs. On the face of it, a URL is just the address of a Web page to be fetched, but a URL such as `http: //www.google.com/search?q=marijuana+cultivation+UK` contains the terms entered into a search engine as well as the search engine's name. Clearly there are many policemen who would like a list of everyone who submitted such an enquiry. Equally clearly, giving this sort of data to the police on anything like a large scale would have a chilling effect on online discourse. It would most likely be found unconstitutional in many jurisdictions.

In fact, when the U.K. government was pushing the RIP bill through Parliament, it tried to entrench a definition that would include URLs (while disclaiming that this was the intention). The news that the police would have unrestricted access to the URLs each user enters—their *clickstream*—caused a public outcry against "Big Browser," so the definition of communications data was trimmed. For general Internet traffic, it now means IP addresses, but it also includes email addresses and the location of mobile phones. All this can be demanded with only a notice from a senior police officer.

Other countries will use different definitions. For example, the U.S. Court of Appeals recently ruled that the cell in which a mobile is located is sufficient, and that to require triangulation on the device (an interpretation the police had wanted) would invade privacy [760]. Also, even cell-granularity location information would not be available under the lower standards applied to pen register warrants. Pen register warrants were also found insufficient for *post-cut-through* dialed digits, as there is no way

to distinguish in advance between digits dialed to route calls and digits dialed to access or to give information. In practice, this means that if a target of investigation in the United States goes to a convenience store and buys a phone card for a few dollars, the police can't get a list of whom he calls unless they obtain a full wiretap warrant. They are entitled only to the digits the suspect dials to contact the phone card operator, not the digits he dials afterward to be connected.

The proliferation of different national standards of what is content and what is communications data may have significant effects on politics and on engineering.

Finally, the analysis of call data is only one aspect of a much wider issue: law enforcement *data matching*, which means the processing of data from numerous sources. The earliest serious use of multiple source data appears to have been in Germany in the late 1970s, to track down safe houses used by the Baader Meinhof terrorist group. Investigators looked for rented apartments with irregular peaks in utility usage, and for which the rent and electricity bills were paid by remote credit transfer from a series of different locations. This worked: it yielded a list of several hundred apartments, among which were several safe houses. The tools to do this kind of analysis are now shipped with a number of the products used for traffic analysis and for the management of major police investigations. The extent to which they're used depends on the local regulatory climate; there have been debates in Britain over police access to databases of the prescriptions filled by pharmacists for the National Health Service, while in America, doctors are alarmed at the frequency with which personal health information is subpoenaed from health insurance companies by investigators. There are also practical limits imposed by the cost of understanding the many proprietary data formats used by commercial and government data processors. But it's common for police to have access at least to utility data such as electricity bills (which get trawled to find marijuana growers); and in the long term, absolutely anything that gets monitored and logged is potentially liable to be subpoenaed. In both Britain and America, regulations being proposed or introduced at the beginning of 2001 will give the police much increased power to demand personal data electronically.

21.2.3 Communications Intelligence on Foreign Targets

I covered the technical aspects of signals intelligence in Chapter 16; now is the time to look briefly at the political and organizational aspects.

The bulk of communications intelligence, whether involving wiretaps, traffic analysis, or other techniques, is not conducted for law enforcement purposes but for foreign intelligence. In the United States, the main agency responsible for this is the National Security Agency, the NSA, whose budget (though classified) is certainly in the billions, given its huge facilities and its tens of thousands of employees. The NSA completely dwarfs law enforcement's 150–200 active wiretaps. The situation is similar in other countries; Britain's Government Communications Headquarters (GCHQ) has thousands of employees and an acknowledged budget of £650 million (about a billion dollars), while for many years one single police officer at New Scotland Yard handled the administration of all the police wiretaps in London (and ran the computer crime squad, too).

Information has steadily trickled out about the scale and effectiveness of modern signals intelligence operations. Kahn's influential history of cryptography laid the groundwork, by describing much of what happened prior to the start of World War II[428]; an anonymous former NSA analyst, later identified as Perry Fellwock, revealed the scale of NSA operations in 1972 [288]. "Information gathering by NSA is complete," he wrote. "It covers what foreign governments are doing, planning to do, have done in the past: what armies are moving where and against whom; what air forces are moving where, and what their capabilities are. There really aren't any limits on NSA. Its mission goes all the way from calling in the B-52s in Vietnam to monitoring every aspect of the Soviet space program."

While Fellwock's motive was opposition to Vietnam, the next major whistleblower was a British wartime codebreaker, Frederick Winterbotham, who wanted to write a memoir of his wartime achievements and, as he was dying, was not bothered about prosecution. In 1974, he revealed the Allies' success in breaking German and Japanese cipher systems during that war [806], which led to many further books on World War II sigint [188, 429, 800]. Thereafter there was a slow drip of revelations by investigative journalists, quite of few of whose sources were concerned about corruption or abuse of the facilities by officials monitoring targets they should not have, such as domestic political groups. For example, whistleblower Peg Newsham revealed that the NSA had illegally tapped a phone call made by Senator Strom Thurmond [157,158]. James Bamford pieced together a fair amount of information on the NSA from open sources and by talking to former employees [70]. But the most substantial source on the organization and methods of the signals intelligence of the United States and allies was put together by New Zealand journalist Nicky Hager [368] following the New Zealand intelligence community's failure to obey an order from its Prime Minister to downgrade intelligence cooperation with the NSA.

The end of the Cold War meant that the agencies had to find new reasons to justify their budgets, and a common theme was developing economic intelligence operations against competitor countries. This has accelerated the flow of information about sources and methods. The most high-profile exposé of U.S. economic espionage was made in a report to the European parliament [278], which is concerned that now the USSR has evaporated, and intelligence is acquiring an economic focus, European Union member nations are now the main targets [160].

The picture that emerges from these sources is of a worldwide signals intelligence collection system, known as *Echelon*, and run jointly by the WASP countries (the United States, Britain, Canada, Australia, and New Zealand). Data, faxes, and phone calls get collected at a large number of nodes, including international communications cables that land in member countries (or are tapped clandestinely underwater), observation of traffic to and from commercial communications satellites, special sigint satellites that collect traffic over potentially hostile countries, and listening posts in member states' embassies [278]. The collected traffic is searched in real time by computers known as *dictionaries* according to criteria such as the phone numbers or IP addresses of the sender or receiver, and keyword searches on the contents of email. These search criteria are entered by member countries' intelligence analysts; the dictionaries then collect the traffic satisfying them and ship it back to the analyst. Echelon appears to work very much like a Web search engine, except that instead of searching Web pages it searches through the world's phone and data network traffic in real time.

A number of points here are worth bearing in mind.

- First, modern military operations would be much more difficult without signals intelligence, and in many cases they would be suicidal. The combatant with the better understanding of the other side's radar and communications has a decisive advantage when it comes to jamming and deception. Without an ability to conduct electronic warfare, a state will be unlikely to be competitive in air or naval warfare or in tank battles on the ground. Even guerilla warfare is less likely to be effective if the occupation forces can deny the guerilla the use of modern communications. So it's not surprising that most of the personnel at NSA are military, and its director has always been a serving general. A large proportion of its work concerns the identification and analysis of the radars, telemetry, weapons guidance, electronic countermeasures, and other such resources of countries that are hostile or potentially so.
- Second, the proliferation of cordless phones, radio LANs and other radio-based technologies, plus the fact that everything is going online, present the agencies with a cornucopia of new information sources [560]. Times have never been so good—regardless of the outcome of policy debates over cryptography.
- Third, even with a budget of billions of dollars a year and tens of thousands of staff, not even the NSA can collect all the electronic communications everywhere in the world. The world described by Fellwock is no more. Sprint's budget is bigger than the NSA's, and is largely spent on low-cost commercial products rather than high-cost classified ones, so it can put in lines much faster than the NSA can tap them. And even if the NSA were only interested in, say, the U.K. university system—and could manage to tap the network access point of every British university—it still couldn't ship all the bits across the Atlantic to Fort Meade, as there just isn't enough transatlantic bandwidth. The task of tapping all the data streams of all the corporations in Japan would be an order of magnitude harder. Thus, the central problem facing intelligence agencies is the same as that facing the police: traffic selection. Although in the old days it was possible to record all telephone and data traffic across the Atlantic, even this would be too expensive nowadays, because communications bandwidth is growing in scale and falling in cost much more rapidly than data storage capacity. The critical question then is whether traffic selection can be done in real time [490].
- Fourth, although other countries may complain about U.S. sigint collection, for them to moralize about it is hypocritical. Other countries also run intelligence operations, and are often much more aggressive in conducting economic and other nonmilitary espionage. The real difference between the WASP countries and the others is that no-one else has built this "system-of-systems." Indeed, there appear to be network effects at work in the economics of sigint as in so many other online activities. The value of a network grows faster than its size, and intelligence networks appear to be no different from phone networks, banking networks, or the Internet itself. The more you tap, the cheaper it gets. There have thus been moves to construct a "European Echelon" involving the police and intelligence agencies of continental European countries [269, 280].

Signals intelligence is necessary for a nation's survival, but potentially dangerous—just like the armed forces it serves. An army can be a good servant, but is likely

to be an intolerable master. The issue is not whether such resources should exist, but how they are to be held accountable. In the United States, hearings by Senator Church in 1975 detailed a number of abuses, such as the illegal monitoring of U.S. citizens [185]. Foreign intelligence gathering is now regulated by U.S. law in the form of 50 USC 1801–1811 [759]. This isn't perfect; its requirements are much more lax than those on domestic wiretapping, and in many cases the president can simply authorize collection rather than getting a warrant. Also, there are known loopholes. One is collaboration with friendly services overseas. When Margaret Thatcher wanted to spy on one of her cabinet ministers, she got the work done by the Canadians [322]; and if the U.S. president really wanted to wiretap a senator there's no doubt he could simply ask Britain's GCHQ to do the job—for them, it would be a perfectly legal foreign intelligence task. And Americans are lucky: in most countries, the oversight of intelligence isn't even discussed.

However, there's a much more serious consequence of poor control and accountability than the occasional political abuse. This is the proliferation of intelligence bureaucracies that turn out to be largely useless once the shooting starts. It became a commonplace in Washington during the Cold War that the agencies hated each other much more than they hated the Russians. In Britain, one of the most vicious intelligence battles was not against the IRA, but between the police and MI5 over who would take the lead in the fight against the IRA. There are numerous accounts of intelligence inefficiency and infighting by well-placed insiders, such as Jones [425]. It is in this context of bureaucratic turf wars that we should approach the whole question of key escrow.

21.2.4 The History of Crypto Policy

Many countries made laws in the mid-nineteenth century banning the use of cryptography in telegraph messages, and some even forbade the use of languages other than those on an approved list. Prussia went as far as to require telegraph operators to keep copies of the plaintext of all messages [729]. Sometimes, the excuse was law enforcement—preventing people obtaining horse race results or stock prices in advance of the "official" transmissions—but the real reason was concern about national security. This pattern was to repeat itself again in the twentieth century.

After the immense success that the Allies had during World War II with cryptanalysis and signals intelligence in general, the U.K. and U.S. governments made an agreement to continue intelligence cooperation. This is known as the UKUSA agreement, although the other WASP countries quickly joined it. Although made in 1947, its existence was acknowledged only in 1999. Throughout much of this period, the member nations operated a crypto policy whose main goal was to prevent the proliferation of cryptographic equipment and know-how. Its outlines were vaguely visible to those of us who worked in industries such as banking; more recently, articles written by former insiders have fleshed out the details.

21.2.4.1 Export Control

Until the 1980s, almost the only makers of cryptographic equipment were companies selling into government markets. They could, by and large, be trusted not to sell anything overseas that would upset their major customers at home. This was reinforced by export controls, which were operated "in as covert a way as possible, with the mini-

mum of open guidance to anyone wanting, for example, an export licence. Most things were done in behind-the-scenes negotiation between the officials and a trusted representative of the would-be exporter" [82].

In these negotiations, the authorities would try to steer applicants toward using weak cryptography where possible; and where confronted with a more sophisticated user, would try to see to it that systems had a "back door" (known in the trade as a *red thread*) that would give access to traffic. Anyone who tried to sell decent crypto domestically could be dissuaded by various means. A large company would be threatened with loss of government contracts; a small one, could be strangled with red tape as it tried to get telecoms and other product approvals. The problem encompassed more than cryptography, as controls designed for mainframes were overtaken by technology. By the mid-1980s, the computers that kids had in their bedrooms were considered to be munitions, and manufacturers ended up doing lots of paperwork for export orders. This pleased the bureaucrats, as it gave them jobs and power. Of course, the power was often abused. In one case, an export order for a large number of British-made home computers to the school system in Yugoslavia was blocked at the insistence of the U.S. authorities, on the grounds that it contained a U.S. microprocessor; a U.S. firm was promptly granted a license to export into this market. Although incidents like this brought the system into disrepute, it persists to this day.

By the early 1970s, the development of ATMs and other electronic banking applications created a significant market for standardized, reasonable-quality cryptographic protection. Part of the solution was to run crypto policy along the same lines as controls on missile technology exports—to let just enough out to prevent companies in other countries developing viable markets. Whenever crypto controls got so onerous that banks in somewhere like Brazil or South Africa started having crypto equipment custom-built by local electronics firms, export licensing would ease up until the threat had passed.

The other part of the solution lay in control of standards for banking crypto. A problem that worried the NSA in the 1970s was that many countries were still using cipher machines that could be broken using the techniques developed in World War II (and these weren't just poor countries: the South Africans used rotor machines up till the mid-1980s and the Swiss till the early 1990s). How could a decent cipher be provided for the banking industry, not just in America but overseas, without its being adopted by foreign governments and thus adding hugely to the costs of intelligence collection?

21.2.4.2 DES and Crypto Research

The solution was the Data Encryption Standard (DES). At the time, as I mentioned in 5.4.3.2, there was a good deal of controversy about whether 56 bits were enough. We now know that this was deliberate. The NSA did not at the time have the machinery needed to do DES keysearch; that came later. But by giving the impression that it did, it managed to stop most foreign governments adopting it. The rotor machines continued in service, in many cases re-implemented using microcontrollers, and the traffic continued to be harvested. Intelligence targets who encrypted their important data with such ciphers merely solved the NSA's traffic selection problem.

A second initiative was to undermine academic research in cryptology. In the 1970s, this was done directly by harassing the people involved; by the 1980s, it had evolved into the subtler strategy of claiming that published research work was all old hat. The agencies opposed crypto research funding, essentially by saying, "We did all that stuff 30 years ago; why should the taxpayer pay for it twice?" The insinuation that DES may have had a trapdoor inserted into it fitted well with this play. (A side effect we still live with is that the crypto and computer security communities got separated from each other in the early 1980s, as the NSA worked to suppress one and build up the other. This has significant costs today for all players, including the NSA. Another cost is that, whenever the NSA makes a mistake, as with the design of Clipper, it gets more harshly judged. What goes around, comes around.)

By the mid-1990s, this line had become exhausted. Agency blunders in the design of various key escrow systems showed that they have no special expertise in cryptology compared with the open research community, and as attempts to influence the direction of academic research by interfering with funding have become less effective, they have become much less common.

21.2.4.3 Clipper

Crypto policy came into the open in 1993 with the launch of the Clipper chip. The immediate stimulus for Clipper was the proposed introduction by AT&T to the U.S. domestic market of a high-grade encrypting telephone that would have used Diffie-Hellman key exchange and triple-DES to protect traffic. The government's response was that it could use its huge buying power to ensure the success of a different standard in which spare keys would be available to the agencies to decrypt traffic. This led to a public outcry, and Clipper was withdrawn.

Several more attempts were made to promote the use of cryptography with government access to keys in various guises. Key escrow acquired various new names, such as *key recovery*; certification authorities that kept copies of their clients' private decryption keys became known as *Trusted Third Parties* (TTPs)—somewhat emphasizing the NSA definition of a trusted component as one that can break security. Much of the policy leverage had to do with export licensing; as the typical U.S. software firm exports most of its product, and as maintaining a separate product line for export is expensive, many firms could be dissuaded from offering strong cryptography by prohibiting its export. Products with "approved" key escrow functionality were then granted preferential U.S. export license treatment. (The history of this struggle is still to be fully written, but a first draft is available from Diffie and Landau [250]; and many of the U.S. source documents, obtained under FOIA, have been published in [684].)

One of the engineering lessons from this whole process is that doing key escrow properly is hard. Making two-party security protocols into three-party protocols increases the complexity and the risk of serious design errors; and centralizing the escrow databases creates huge targets [3]. Where escrow is required, it's usually better done with simple local mechanisms. In one army, the elegant solution is that every officer must write down her passphrase on a piece of paper, put it into an envelope, stamp it "Secret" and hand it to her commanding officer, who puts it in the office safe. That way, the keys are kept in the same place as the documents whose electronic ver-

sions they protect, and there's no central database for an airplane to bomb or a spy to steal. (If you have been following the key escrow debate, you may have been conditioned to object, "But a soldier could deposit a false key and then desert and try to sell back the right one." I posed this question to my informant, and he looked at me as if I was crazy. I now believe this objection is indeed crazy, or at best clutching at straws. Anyone, soldier or programmer, can take paper documents and try to ransom them. In practice, it's so rare an event that nobody bothers about it.)

21.2.4.4 European Initiatives

In Europe, things have been somewhat more confused. Here's a brief summary (there is an extensive survey at [472]). International arms control agreements (COCOM and Wassenaar) bind most governments to implement export controls on cryptographic equipment; and countries that are member states of the European Union are also bound by an EU regulation on the export of *dual-use goods*—goods that have both civilian and military uses. But European bodies have been cool toward crypto control, and national implementations vary. U.K. law doesn't control export of intangibles, so crypto software could be exported electronically; the Belgian government would grant licenses for almost anything; and Switzerland remained a major exporter of crypto equipment. Domestic controls also varied. The French government started from a position of prohibiting most civilian cryptography, and moved to almost complete liberalization, while Britain went the other way.

In 1996, one of the last acts of the outgoing Major government in Britain was to propose that key escrow be mandatory. The opposition Labour party made a ringing denunciation of this: "Attempts to control the use of encryption technology are wrong in principle, unworkable in practice, and damaging to the long-term economic value of the information networks" [197]. Once in power, though, their view changed rapidly and the new RIP Act allows a policeman to demand any crypto key that's been in your possession. If you refuse you can get two years, and if you tell anyone that it's been seized you can get five. One intended effect was 'escrow by intimidation'—to bully companies into using key escrow to ensure they could comply with law enforcement demands for keys. However an attempt to make company directors liable to go to prison if keys couldn't be produced was defeated by industry lobbying. For the history of the RIP bill, see [304].

Another thread running through European crypto policy initiatives has been the attempt to link key escrow to other initiatives and standards. For example, the European Electronic Signature Directive forces member states to grant higher-quality recognition of digital signatures made using approved products; in at least one country, it was proposed that this would mean products supporting escrow. And, as noted in the chapter on telecomms fraud, law enforcement access was built into the standards for third generation mobile services.

21.2.4.5 Red Threading and the Crypto AG Case

Quite often, key escrow has been implemented without the knowledge of the users. The Swedish government got upset when it learned that the "export version" of Lotus Notes, which it used widely in public service, had its cryptography deliberately weakened to allow NSA access; and at least one (U.S. export approved) cipher machine has

broadcast its plaintext in the clear in the VHF band. But the most notorious example was the Bühler case.

Hans Bühler worked as a salesman for the Swiss firm Crypto AG, which was a leading supplier of cryptographic equipment to governments without the technical capability to build their own. He was arrested in Iran in 1992, and the authorities accused him of selling them cipher machines that had been tampered with so that the Great Satan could get at the plaintext. After he had spent some time in prison, Crypto AG paid 1.44 billion Rials—about $1 million U.S.—to bail him out; then he was fired after he got back to Switzerland. Bühler later alleged on Swiss radio and TV that the firm was secretly controlled by the German intelligence services, and that it had been involved in intelligence work for years [143]. The interpretation commonly put on this was that ultimate control resided with the NSA (the founder of Crypto, Boris Hagelin, had been a lifelong friend of William Friedman, the NSA's chief scientist) and that equipment was routinely red-threaded [517]. A competing interpretation is that these allegations were concocted by the NSA to undermine the company, as it was one of the third world's few sources of cryptographic equipment. Bühler's story is told in [740].

What should an ordinary security engineer—one not involved in the intelligence business—make of all this?

21.2.5 Discussion

When the key escrow debate got going in Britain in 1994–1995, I took a line that was unpopular at the time with both the pro-escrow and the anti-escrow lobbies. The pro-escrow people said that because crypto provided confidentiality, and confidentiality could help criminals, there had to be some way to defeat it. The anti-escrow lobby said that because crypto was necessary for privacy, there must not be a way to defeat it. I argued in [21] that essentially all the premises behind these arguments were wrong. Most crypto applications (in the real world, as opposed to academia) are about authentication, rather than confidentiality; they help the police rather than hinder them. As for criminals, they require unobtrusive communications—and encrypting a phone call is a good way to bring yourself to the attention of the agencies. As for privacy, most violations result from abuse of authorized access by insiders. Finally, a much more severe problem for police or auditors investigating electronic crimes is to find acceptable evidence, for which decent authentication can be helpful.

Events since have largely borne out this initially contrarian view. For most of the 1990s, I helped organize an annual conference on white-collar crime in Cambridge, and organized regular sessions and workshops on key escrow and related issues. These turned out to be of almost no interest to the policemen and prosecutors who formed the bulk of our audience; they headed off to the bar as soon as the session on wiretaps and crypto got going. Most police forces took an interest in the subject only once they were told to. In many countries, including the United States and Britain, the lead agency on crypto policy is a law enforcement one (the FBI and the National Criminal Intelligence Service, respectively), but this is simply a front for the intelligence community—as was admitted in an unguarded moment in 1996 by the U.K. representative on the European body responsible for crypto policy [378].

21.2.5.1 Law Enforcement or Intelligence?

The use of law enforcement as a cover is a source of continuing problems. The aims and objectives of policemen and spies are not quite identical, and confusing them has clouded matters. It is perhaps an oversimplification that the former try to prevent crimes at home, while the latter try to commit them abroad; but such aphorisms bring out some of the underlying tension. For example, policemen want to preserve evidence, while spies like to be able to forge or repudiate documents at will. During the discussions on a European policy toward key escrow ("Euroclipper") that led up to the Electronic Signature Directive, the German government demanded that only confidentiality keys should be escrowed, not signature keys, whereas Britain wanted signature keys to be escrowed as well. The British view followed the military doctrine that deception is at least as important as eavesdropping, while the Germans supported the police doctrine of avoiding investigative techniques that undermine the value of any evidence subsequently seized.

Key escrow can also, like the system for classifying official documents, help provide plausible deniability for official wrongdoing. The key management system used in the U.K. civil service distributes signature keys to end users encrypted under escrowed confidentiality keys [50]. So if an embarrassing electronic document is leaked to the press, the government can claim that it was forged by the departmental security officer—the person responsible for preventing leaks, who is also the person with access to escrowed keys. Depending on your point of view, this is either a brilliant piece of security engineering, whose inventor should get a medal, or a wicked and perverted design whose inventor should get jail time for undermining public accountability and the principles of freedom of information.

Quite apart from signing key issues, the intelligence community appears to be the main beneficiary of crypto control. It's not just that wiretaps are the most economic way to keep an eye on guys like Saddam Hussein. If a significant proportion of data traffic were encrypted, then the automated keyword searching done by systems such as Echelon would be largely frustrated. Spooks are also aware that large numbers of new network infrastructures are built each year, and if cryptography isn't built in at the start, it may well be too expensive to retrofit it later. Therefore, each year that the NSA can hold the line on crypto controls means hundreds of networks that will be open to surveillance for decades in the future. Whether this will work for the long-term benefit of the United States and Europe, leave us terribly exposed in twenty years' time once China starts to compete as a superpower, or even lead to destabilizing conflicts on economic espionage between the United States and Europe, is a question that doesn't get debated much.

This is not to say that the police have no use for wiretaps. Although many police forces get by quite happily without them, and many of the figures put forward by the prowiretap lobby are dishonest [250], there are some occasions when wiretapping can be economic as an investigative tool. The Walsh report—by a senior Australian intelligence officer—gives a unusually balanced examination of the issues [787]. Walsh compared the operational merits of wiretaps, bugs, and physical surveillance, and pointed out that wiretaps were either the cheapest or the only investigative technique in some circumstances; but he still felt that compelling disclosure of crypto key material to the government was likely to be ineffective. "The invocation of the principle of non-self-incrimination may well represent the polite end of the possible range of responses," he drily remarked. Among his findings were that there is "no compelling rea-

son or virtue to move early on regulation or legislation concerning cryptography." But he did recommend that police and intelligence agencies be allowed to hack into target computers to obtain access or evidence.[1] Although there will be some policing costs associated with technological advances, there will also be opportunities: for example, to infect a suspect's computer with software that will turn it into a listening device. In general, the police—like the intelligence services—are reaping a rich harvest from modern technology.

Overall, the net effect on law enforcement of the key escrow debate has been negative; it has eroded both public trust and operational effectiveness. In the intelligence community, too, many officers deeply regret having launched the Clipper initiative. Before it, cryptography was largely unknown: a few mathematicians studied it academically, and it was used in cash machines and pay-TV decoders, but public awareness of communications security was low. (When I first wrote some email encryption software in 1985, there was almost no interest.) That has now changed. Not only do many more criminals use anonymous communications channels, such as prepaid mobiles, but many countries that previously bought weak or red-threaded cipher machines for their military and diplomats have now started to develop local expertise and products. However, as the saying goes, "Policy has no reverse gear."

21.2.5.2 Carnivore

As of summer 2000, the direction of policy on wiretaps, traffic analysis, and crypto control is acquiring two main features, The first is the blurring of the line between intelligence and law enforcement. There has always been some overlap, especially in counterespionage and terrorism cases. In some countries, such as the United States, there are agencies explicitly endowed with both functions (the FBI—though note that in 1998, for example, only 45 of its 12,730 convictions involved what the Justice Department classified as internal security or terrorism matters [751]). In others, there have been huge turf fights. I mentioned the one in Britain over whether the police or MI5 should deal with the IRA; since the Northern Ireland peace treaty, the same fight has been repeated over computer crime. The end of the Cold War, and of many regional insurgencies, has left a lot of well-connected agencies desperately looking for new lines of business.

The second thread is more intrusive surveillance at ISPs. Tapping data traffic is harder than tapping voice used to be; modern modems use adaptive echo cancellation that makes passive interception of the local loop more difficult, while interception elsewhere faces several obstacles such as transient IP addresses given to dial-up customers and the increasingly distributed nature of packetized traffic. Both Russia and

[1] The Walsh report has an interesting publishing history. Originally released in 1997 as an unclassified document, it was withdrawn three weeks later after people asked why it wasn't yet on sale in the shops. It was then republished in redacted form. But in 1998, researchers found unexpurgated copies in a number of public and university libraries, which had received legal deposit copies and had been insufficiently diligent in finding and returning them. These were published on the Web, and the redacted parts drew attention at once to the issues the government considered sensitive. As late as 1999, the Australian government was still trying to suppress the report [787].

Britain have introduced laws requiring ISPs to attach black boxes to their networks for surveillance purposes, while in the United States the FBI has a device called Carnivore that performs this function. So-called because it's supposed to "get the meat" out of a digital wiretap, Carnivore is documented extensively at [717].

The thinking behind Carnivore was that legal solutions were becoming ineffective, as the technology changes too quickly; and that the standard tools used by ISPs to monitor their networks for diagnostic purposes got only parts of the needed information, or too much (which conflicted with legal requirements for minimization). It was preferable to have a technological solution based on a general-purpose platform whose software could be upgraded as needed. In fact, Carnivore can be configured remotely, which some ISPs don't like. What's more, the operator is completely trusted; pressing a single button causes all TCP traffic to be collected, and the device lacks the audit trails needed for establishing individual accountability. There are a number of serious problems, such as dealing with non-standard ISP equipment and with services layered on top of other services, such as webmail. No doubt Carnivore and its foreign equivalents will continue to evolve, and the growing complexity of the ISP business will keep their maintainers busy.

At least in the United States, the better legal supervision of wiretaps means that Carnivore isn't preplaced but is installed only after a court has granted a warrant; and the number of deployments each month can still be counted on the fingers of one hand. In the U.K. and the Netherlands, it looks like similar devices will be installed at all major ISPs to monitor traffic continuously [147]; in Russia, they already have been.

21.2.5.3 Underlying Policy Problems

What are we to make of a huge effort to build a capability that's used only rarely? I am afraid that many of the disputes in e-policy involve what Freudians might call a *displacement activity:* inability to solve a hard problem causes frustration, which is vented by energetically solving an irrelevant but easier one. In England, for example, it has been notorious since at least the time of Queen Elizabeth the First that rich, successful criminals are almost never prosecuted. They usually get caught only when their businesses collapse, as with the Barings and Maxwell cases discussed in Section 9.2.3 (and even then, Leeson was prosecuted in Singapore rather than London, while Maxwell's crimes were detected only after his suicide.) In my own professional practice, I have long since given up reporting crooked bankers to the police: there has been no prosecution of a senior banker that anyone can remember. In the United States, about a thousand bankers at the grade of vice president and up get prosecuted every year, and over a third get jail time. This isn't a matter of British virtue, or American vice, but has to do with how the two law enforcement systems are organized. U.S. police officers get promoted if they win high-profile convictions, so the relevant U.S. agencies such as the FBI, the Secret Service and local DAs' offices compete to put bent bankers in jail. In contrast, their British counterparts depend for promotion on establishment patronage; and raiding a prominent person for anything short of murder is career death. Thus, U.K. agencies compete to pass the buck and look the other way. Now as high-value crimes by smart crooks are precisely the minority of crimes in which wiretaps are often economic, the U.K. government's public arguments about police surveillance powers seem even thinner.

Displacement activity isn't limited to communications intelligence issues. A lot of noise has been made about Internet-based child sex offenses, and especially kiddie-porn. Yet the number of such cases is small; and even in a high-profile case involving what the judge called the "very worst possible type" of material—against former pop idol Gary Glitter—the court thought a four-month sentence appropriate. Most offenders get away with fines or community service [12]. So it's hardly the most serious of crimes, and as the use of computers by child porn networks goes back to the 1980s, it's hardly a new one either. What's more, when you talk to people involved in child protection, it becomes clear that there are thousands of really serious cases of child abuse every year in Britain, usually involving abuse by family members, abuse of young persons with learning difficulties, abuse of children in local authority care, and under-age prostitution. For various political reasons, the police don't always find it convenient to crack down on these crimes; and as for the charities, the end of the orphanage system has left them dependent on local government for permission to place vulnerable children in care. Still, children's organizations spend their charitable funds campaigning against the evils of the Net [168], rather than lobbying for the respectable middle-class customers of 13-year-old prostitutes to be sent down for child rape [528]. (There are some interesting reflections on attitudes to sex offenders, and the transference mechanisms involved, at [215]. It appears that, just as the end of universal belief in God left, a surveillance vacuum which governments have rushed to fill, so also the death of the devil has left a vacancy. The greatest hysteria about child sex abuse is whipped up in the very neighborhoods where the abuse of girls by their stepfathers or stepbrothers is routine. People transfer to 'the devil' their own darkest fears and childhood traumas.)

The implication of all this for the security engineer is that you have to think hard about the risk that your product or service will become the target of hysterical abuse by ineffective or corrupt public servants, or by ignorant and hypocritical self-publicists. You can't ignore the social and political context of what you're trying to build.

21.3 Copyright

In Chapter 20, I suggested that the 1990s debate on crypto policy is likely to be a test run for an even bigger battle, which will be over anonymity, censorship, and copyright. I looked at some of the technical aspects in that chapter, and discussed a number of the business and political aspects that were integral to that story. The context is not just copyright though. Mechanisms such as anonymous remailers and highly distributed file stores allow people to exercise their right to anonymous political speech, and also let them publish material that is defamatory or seditious with a decreased likelihood of being caught and punished. Thus, the copyright enforcement lobby has some powerful potential allies.

There are geopolitical aspects, too. In most countries, there is no right to free speech (let alone anonymous political speech), as enjoyed in the United States; and even in European countries, the laws on defamation and sedition can be savage. There have also been high-profile cases in which courts in countries with laws against hate speech, such as France and Germany, have looked for ways to censor U.S. online services. As I write, a court in Paris has just given Yahoo three months to prevent its French subscribers having access to auctions of Nazi memorabilia, which are illegal in France. And at the Global Internet Project conference in Berlin, in November 2000, I heard the

German federal justice minister proclaim that her greatest achievement in office was stopping Books Online shipping copies of "Mein Kampf" to addresses in Germany. Speaking through an interpreter, she assured us that she would not rest until they stopped shipping it in Arizona, too. (Given that the copyright of "Mein Kampf" is owned by the government of the state of Bavaria, they may have the right to end its publication by boring, old-fashioned legal means; perhaps denouncing the Evils of the Internet is thought more attractive to the voters.)

If, as Leslie Lamport said, you know you have a distributed system when the crash of a computer you've never heard of stops you from getting any work done, then you also know you're living in the global village when a judge in a country you've never heard of can try to close down your business—or at least dictate onerous conditions on how you conduct it in your own country. (Being based in the United States, which is isolated from international enforcement of court judgments, gives you some protection—as discussed in Section 19.9.) I don't think anyone has ever considered the distributed systems aspects of international law, but there could be an interesting PhD thesis in it.

Of course, it cuts both ways: third world despots and Asian strongmen denounce the freedom of speech on the Internet as "neo-imperialism." And most European countries have a more liberal view of pornography than most Americans are comfortable with. It remains to be seen whether the Internet of 2020 will have U.S. rules on freedom of speech and European rules on porn, or the other way round.

Because of the lack of consensus on issues such as obscenity and sedition, the most likely way for controls to be introduced will be copyright. Chapter 20 described how successive technologies such as audiocassettes and videocassettes arrived on the market, caused panic among copyright owners, but turned into profitable lines of business once Hollywood had learned to manage them. PC software followed exactly the same model, only more rapidly. Pay-TV was slightly different, as the use of tamper-resistant subscriber tokens, plus aggressive legal pursuit of token forgers, enabled piracy to be kept in the single percentage figures. Hollywood is now trying to get DVD to follow the pay-TV path; however, thanks to design and other errors, this looks to be slipping from their grasp. It now looks like DVDs will follow the same model as PC software or videocassettes, and will be the future distribution medium for both.

The issue is by no means a straight fight between copyright and privacy. As noted in Chapter 20, the doctrine of fair use allows people to copy parts of a work for purposes ranging from scholarship to ridicule. The possible abolition of fair use has alarmed universities and libraries. Pamela Samuelson expresses a common sentiment: "Why would the Clinton administration want to transform the emerging information superhighway into a publisher-dominated toll road" [667]?

21.3.1 DMCA

Following heavy lobbying, a treaty was made in Geneva in 1996 under the auspices of the *World Intellectual Property Oragnization* (WIPO), with signatory states obliged to harmonize the treatment of digital copyright. The implementation in the United States was the *Digital Millennium Copyright Act* (DMCA) of 1998. This prohibits the alteration of any electronic *copyright management information* (CMI) bundled with digital content, such as details of ownership and licensing; and outlaws the manufacture, importation, sale or offering for sale of anything primarily designed to circumvent copyright protection technology. There are specific exemptions for people engaged in

encryption research, for libraries, and to detect and disable the kind of snitchware discussed in Section 18.3.2.4.

DMCA also provides some limited protection for ISPs that unknowingly host copyright material on Web sites that has been posted by their clients or other third parties. A condition is "Notice and Take Down": when a copyright owner notifies an ISP of a violation, the offending material must be removed. To prevent this being abused, there is also a provision for "Notice and Put Back": if the subscriber files a proper "counter notice," attesting to its lawful use of the material, then the ISP must promptly notify the copyright owner and restore the material within 14 business days, unless the matter has been referred to a court.

The exemptions for libraries are the focus of continuing debate, as digital access will mean only limited access, unless you own a copy of the work; and libraries will have to negotiate terms for fair access for all reasonably expected purposes, in the face of initial licensing conditions and fees that that may be far too high. (These issues are discussed at [513].) Particularly intractable problems are raised by legal deposit libraries, such as the Library of Congress and its counterparts elsewhere. Traditionally, the grant of copyright in many countries was conditional on the copyright owner's depositing one or more copies of the work in a national archive. This serves a number of purposes ranging from helping the courts resolve copyright disputes, through access by future generations of scholars, to the supply of obscure books through library loan schemes. But how can we preserve digital works that may use proprietary platforms, have copyprotection schemes, and may even require occasional online access to a license server? However, the first big test of DMCA looks to be the reverse-engineering of DVD CSS, which is currently making its way through the U.S. courts.

21.3.2 The Forthcoming European Directive and UCITA

In Europe, there is already a Conditional Access Directive that obliges EU member states to outlaw devices that enable unauthorized access to services such as pay-TV and Internet subscription sites. As with DMCA, there are exemptions for bona fide research. A difference is that the Conditional Access Directive protects measures that control access to a service, not those that control access to a work.

The protection of works should arrive in the form of a Copyright Directive, which at the time of writing is still the subject of debate. There has been vigorous lobbying on the one hand from Hollywood and on the other from libraries. It looks like the directive will have broadly the same effect as DMCA, though the details will be up to member countries. (For a discussion of the proposed directive and a comparison with DMCA, see [468].) The sort of argument being considered by the European Commission is that if a rights holder could object to circumvention in cases where a claim based upon copyright law would not succeed, the effective "reach" of copyright holders would expand. What's more, some companies are already packaging copyright control mechanisms with other kinds of protection, such as accessory control in the computer games industry. It seems unreasonable to just grant copyright holders and gaming console vendors any right for which they can devise a protection mechanism, and to criminalize defeats of such mechanisms, however ineffective or inappropriate they are.

Also, existing European law allows reverse-engineering for interoperability—to ensure that the program can work with other programs—while DMCA adds the restriction that there must not be a readily available commercial alternative for that purpose. The general European view, following the reverse-engineering of the DVD CSS (which was necessary for the Linux community), is that the DMCA provisions are too tightly drawn. (This means that even if Hollywood wins the U.S. case, there should be a safe haven for Linux developers in Europe.)

Another issue in the United States is the *Uniform Computer Information Transactions Act* (UCITA), a model law sponsored by the National Council of Commissioners on Uniform State Laws (NCCUSL), which will be introduced in state legislatures nation wide. UCITA will update the U.S. Uniform Commercial Code to cover digital trade, and will govern contracts between manufacturers and consumers regarding nearly all "transactions in information." This means everything from stories, computer programs, images, music, and Web pages to online databases and interactive games. UCITA will significantly extend federal copyright law. Many states are unhappy with it; it will replace copyright law with contract law in many cases, undercut fair use, outlaw reverse-engineering even for interoperability, and enable shrink-wrap/click-on licenses to bind users to contracts before they can even read the conditions. It could have severe consequences for everyone from open source software authors to universities [608].

It's by now inevitable that there will be important differences between the laws in the United States and Europe, which may have a significant effect on security designs. Also, at the level of fine structure, the issues look set to be fought out in dozens of individual state and national legislatures. This might work in Hollywood's favor, as it has the money and organizational resources to lobby in dozens of places at once. But it also increases the likelihood that a spectacular loss somewhere will create a haven, just as Ireland became a safe haven for pay-TV smartcard cloning in the early 1990s.

21.4 Data Protection

Data protection is a term used in Europe to mean the protection of personal information from inappropriate use. Personal information generally means any data kept on an identifiable human being, or *data subject*, such as bank account details and credit card purchasing patterns. It corresponds roughly to the U.S. term *computer privacy*. The difference in terminology is accompanied by a major difference in law and in attitudes. In fact, this may become one of the thorniest problems in e-policy in the first decade of the twenty-first century, as well as a serious complication for people setting up e-businesses.

European law gives data subjects the right to inspect personal data held on them, have them changed if inaccurate, understand how they're processed, and in many cases prevent them being passed on to other organizations without their consent. This means, for example, that people who have been refused credit can see not just their files but also the credit-scoring algorithms used to make the decision; and if a U.S. bank doesn't

like that, tough. There are exemptions for national security, but not for all police data. Most commercial data are covered, and there are particularly stringent controls on data relating to intimate matters such as health, religion, race, sexual life, and political affiliations. Finally, recent law prescribes that personal data may not be sent to organizations in countries whose laws do not provide comparable protection. In practice, that means America, where legal protections on privacy are fragmentary.

The implication for the engineer designing an e-commerce application is that once the relevant European law has been tested all the way to the European court—which might be in 2004 or 2005—it may be illegal for you to process data about your European customers in a facility on U.S. soil. One solution may be to put your servers in a European country with lax enforcement, such as Britain or Iceland; but there is a growing body of case law that constrains European governments' freedom to turn a blind eye. If your business model involves collecting large amounts of personal information about buying habits, news-reading patterns, and so on, then you could be in trouble even in London or Reykjavik.

Another solution favored by many business is *coercive consent*, which means you insist that customers agree to their personal data being shared before doing business with them. This tends to work at present (it's how U.S. medical insurers get away with their abuses) but isn't guaranteed for ever. The click-on no-privacy agreement on your Web site might be deemed an unfair contract term by a court; and in some countries, it will be invalid if the customer is a minor or the information relates to an intimate matter such as health.

European privacy law didn't spring full-formed from the brow of Zeus, though, and it may be helpful to look at its origins.

21.4.1 European Data Protection: History

Technofear isn't a late twentieth-century invention. As early as 1890, Warren and Brandeis warned of the threat to privacy posed by "recent inventions and business methods," specifically photography and investigative journalism [792]. Years later, after large retail businesses started using computers in the 1950s and banks followed in the early 1960s, people started to worry about the social implications if all a citizen's transactions could be collected, consolidated, and analyzed. In Europe, big business escaped censure by making the case that only government could afford enough computers to be a serious privacy threat. It was realized that it was possible, economic and rational for government to extend its grasp by using the personal data of all citizens as a basis for prognosis; and given the recent memory of the Gestapo in most European countries, this became a human rights issue. A patchwork of data protection laws began to appear, starting with the German state of Hesse in 1969. Because of the rate at which technology changes, the successful laws have been technology-neutral. Their common theme was a regulator (whether at national or state level), to whom users of personal data had to report and who could instruct them to cease and desist from inappropriate processing The practical effect was usually that the general law became expressed through a plethora of domain-specific codes of practice.

Over time, processing by multinational businesses became an issue, and it became clear that purely local or national initiatives were likely to be ineffective against them.

Following a voluntary code of conduct promulgated by the OECD in 1980 [598], data protection was entrenched by a Council of Europe convention in January 1981, which entered into force in October 1985 [206]. Although, strictly speaking, this convention was voluntary, many states signed on for fear of losing access to data-processing markets. It was founded in the European Convention on Human Rights, and required signatory states to pass domestic legislation to implement at least certain minimum safeguards. Data had to be obtained lawfully, and processed fairly, and states had to ensure that legal remedies were available when breaches occurred.

The quality of implementation varied widely. In Britain, for example, Margaret Thatcher unashamedly did the least possible to comply with European law. A data protection body was established, but starved of funds and technical expertise; and many exemptions were provided for favored constituencies. Though not for journalists; if you kept notes on your laptop which identified people, you were formally liable to give copies of this information to the data subjects on demand. In hard-line privacy countries such as Germany the data protection bodies became serious law enforcement agencies. Many non-European-union countries, such as Australia, Canada, Iceland and Switzerland, passed comparable privacy laws in the 1980s and early 1990s. Some, like Switzerland, went for the German model, while others, like Iceland, followed the British one.

By the early 1990s, it was clear that the difference between national implementations, exacerbated by the accretion of case law, was erecting barriers to trade. For many businesses, the solution was to avoid controls altogether by moving (or outsourcing) their data processing to the United States. The growing tensions led in 1995 [279] to a new Data Protection Directive. This sets higher minimum standards than most countries had required before, with particularly stringent controls on highly sensitive data such as health, religion, race, and political affiliation. It also prevents personal information being shipped to "data havens" such as the United States, unless there are comparable controls in place. The directive could prove to be a serious headache for new business models, such as application rental [182].

21.4.2 Differences between Europe and the United States

The history in the United States is, basically, that business managed to persuade government to leave privacy largely to "Self-regulation" (for more on U.S. history on this topic, see [572]). Although there is a patchwork of state and federal laws, they are application-specific and highly fragmented. In general, privacy in federal government records and in communications is fairly heavily regulated, while health and business data are largely uncontrolled. One or two islands of regulation, do exist, such as the Fair Credit Reporting Act of 1970, which governs disclosure of credit information, and is broadly similar to European rules; and the Video Privacy Protection Act or "Bork Bill," enacted after a Washington newspaper published Judge Robert Bork's video rental history following his nomination to the U.S. Supreme Court.

Attitudes also differ. According to Westin, about twenty-five percent of Americans are "privacy fundamentalists," favoring legislative standards; twenty percent are unconcerned, and will readily pass on their personal information for minor benefits; while the majority, fifty-five percent, are pragmatists who take privacy decisions case

by case. But there is a growing feeling that people have lost control of the uses to which their personal information is put. This still lags behind Europe, where privacy is seen as a fundamental human right that requires vigorous legislative support [802].

Clearly, the stage is set for a major conflict between Europe and the United States on data protection. U.S. policymakers have failed to appreciate the severity of the problem; a common view on Capitol Hill is that, "It's just a spiteful retaliation for the Helms-Burton Act, and we can negotiate some deal on it." Their current hope for a deal is the *safe haven* concept, that U.S. data processors can simply enter into a contract with their European customer or subsidiary to the effect that data will be processed in accordance with European law. Some firms have already done this, led by Citibank which uses such an arrangement to process German cardholder data in South Dakota. But this creates severe practical enforcement problems for EU citizens who feel that their rights have been violated, and may well fail when tested in court. For a discussion, see [802].

21.4.3 Current Trends

The European regulatory drive toward data thrift is counter to the direction in which commerce is developing. Quite apart from the law enforcement surveillance techniques discussed in the first section of this chapter, e-businesses are developing all sorts of customer tracking and marketing tools from cookies to clicktrails, wallets to IPR enforcement tools, and snitchware that enables software vendors to monitor customers' hard drives remotely. The information flow is one-way, in the sense that you retain essentially, no rights over personal information once surrendered; yet businesses will only license their software to you rather than sell it. Some writers have expressed the fear that, regardless of any regulatory efforts, technology will land us in a world in which there is no place to hide [323].

An extreme version of this view is taken by David Brin [139]. He argues that pervasive surveillance technologies will inevitably be available to the authorities, and the only real question is whether they will be available to the rest of us, too. He paints a choice between two futures—one in which the citizens live in fear of an East German-style police force, and one in which officials are held to account by public scrutiny. The cameras will exist: will they be surveillance cams or Web cams?

There are some successful experiments in openness. The U.S. Freedom of Information Act may be the most conspicuous, but there are others, such as the practice (in Iceland and in some Swiss cantons) of publishing tax returns—a practice that greatly cuts evasion as rich men fear the loss of social status that an artificially low declared income would bring.

Underlying such considerations is a growing understanding of the economics of privacy. The basic problem is that for the data subject, the value of personal data is its marginal cost, while for the collector it's the average cost. Thus collectors are going to pay more to get data than most users will pay to deny it to them. Another economic aspect is that, if privacy is left to technology, it will be a cost that falls largely on the data subject; but if it's done by regulation, it will fall more on the collector [323]. One ray of hope is that the data that people want to keep private and the data that marketers want to collect are often not the same commodity. Personal secrets tend to be long term (such as a treatment for alcoholism ten years ago), while marketing data is short term (how much can I increase the probability of selling this person an airplane seat today if I cut the price 20%?)

Perhaps part of the solution will come from tools such as online auctions. But there are many places where Web cams will probably always be considered unacceptable, such as corporate research and development labs, attorneys' offices and doctors' consulting rooms. Defining the boundary will no doubt involve a lot of pushing and shoving.

The evolution of this issue over the next few years will be of great interest to security engineers. The issue will not be limited to the collection of data, but also to its collation. For example, while U.S. felony convictions remain on the record forever, many European countries have laws governing rehabilitation of offenders, under which most convictions disappear after a period of time that depends on the severity of the offence. But how can such laws be enforced now that Web search engines exist? The German response is that if you want to cite a criminal case, you're supposed to get an officially deidentified transcript from the court. But if electronic newspaper archives are searchable online, what good will this do—unless the identities of all offenders are blocked from electronic reporting? Recently, for example, there has been much debate over the monitoring of former child sex offenders, with laws in some states requiring that offender registers be publicly available. Riots occurred in England following the naming of some former offenders by a Sunday newspaper. There's a long list of similar issues, from the permissible uses of electoral rolls and lists of people who have been naturalized to whether it is permissible to index certain types of publicly available information. The upshot is that even if data is public, its use can still cause offenses under European privacy law.

This causes peculiar difficulties in the United States, where courts have consistently interpreted the First Amendment to mean that you can't stop the repetition of true statements in peacetime except in a small number of cases, of which the classic example is a regulated profession such as securities trading. Perhaps marketing will end up a regulated profession; or perhaps the penalties for the repetition of untrue statements can be made high enough to cause people to take care. Neither seems likely at present in mass markets, although the rich and famous can extract substantial damages for libel in many countries' courts. I await with interest the first case in which someone bankrupts a search engine operator for bringing to public attention an expired conviction for drug use.

It's possible that America will enact privacy legislation that's sufficiently mid-Atlantic to prevent a trade war on the issue; Al Gore promised an "Electronic Bill of Rights" to protect people against the misuse of computerized personal information of all types. It's conceivable that, like the Internet, privacy intrusions will suddenly reach a critical mass, and public opinion in the United States will compel politicians to override business interests and pass European-style data protection laws. It's also conceivable that Europeans will come to share the view of American privacy pragmatists—that though a few unlucky people may have terrible experiences, the worst that will happen to the average family is an armful of junk mail each week with which they can light the barbecue. But it's also possible that Europe might become more fundamentalist still, perhaps in reaction to U.S. e-commerce practices. Thus, although the two markets might converge, there is a real risk that they won't; and neither is small enough to ignore.

In the meantime, it is prudent for the e-commerce designer to ensure that business processes and systems can comply with the European way of doing things as well as the American one.

21.5 Evidential Issues

I mentioned the European Electronic Signature Directive, which forces member states to grant higher-quality recognition of digital signatures made using approved products; that there were attempts to link this approval to approved key escrow mechanisms; and that there were attempts to force the escrow of signature keys as well, which could have undermined the value of digital evidence.

But these are neither the beginning nor the end of the evidential issues confronting the security engineer. Designing a system whose functions include the production of evidence is a lot harder that it seems at first.

21.5.1 Admissibility of Evidence

When courts were first confronted with computer evidence in the 1960s, there were various concerns about the reliability, both in a technical sense and in the legal sense of whether it was inadmissible on the grounds that it was hearsay. Different legislatures tackled this differently. In some, computer evidence is deemed to be admissible, but can be challenged in court by the other side; in others, it can't even be presented unless accompanied by a certificate stating that the computer was working properly. (This can cause problems when the evidence comes from a machine that has been hacked.) In the United States, most of the law is found in the Federal Rules of Evidence, while in Britain it's in the Police and Criminal Evidence Act 1984 and the Civil Evidence Act 1995.

In many cases, evidence can be derived only from the operation of a machine as it's operated in the normal course of business, and this can cause problems if a requirement for evidence hasn't been anticipated by the engineer. For example, in one case in my own experience, a woman was accused of stealing a debit card from the mail, and the police wanted to ascertain whether a torn-off corner of a PIN mailer found in her purse would activate the stolen card. They got the branch manager to put the card into a statement printer in the branch office, entered the PIN, and the card was confiscated. The manager testified that the way the card was confiscated showed that it was because the account had been closed rather than because the PIN was wrong. The court ruled this evidence to be inadmissible. The law on this subject changes regularly, though.

21.5.2 Reliability of Evidence

Even where the local formalities can be observed, computer forensics pose complex and nontrivial engineering problems. Even to the experienced systems administrator, securing evidence of an intrusion in a timely and nondestructive manner is hard. As operating systems get ever more complex, they become less deterministic, and their logging and other features more opaque. The response of the law enforcement community has been tools that will take a mirror image copy of a hard disk for subsequent examination. This isn't the end of the story, though, because of the complexity and quantity of data, and the multiple interpretations that are often possible. Application file formats usually aren't adequately documented, and may contain bugs or features which their creators are unwilling to discuss, because they would embarrass or even incriminate them. New gadgets, such as palmtop computers with closed operating systems, and SIM cards for which the suspect won't divulge the password, can force the

practitioner to resort to the kind of reverse-engineering tricks described in Chapter 14. Things are made worse by the technical incompetence of judges and other lawyers; the common result is that arguments (and judgments) confuse fact, conjecture, assumption, inference and opinion.

The signal-to-noise ratio of the court system is especially low when a case hinges on a technical matter. Often, the only safeguard against injustice lies in the adversarial system itself. Recall the Munden case described in 9.4.3. A man was falsely accused and wrongly convicted of attempted fraud after he complained of unauthorized withdrawals from his bank account. Rational argument having failed, the way in which the appeal was won was tactical—getting an order requiring the bank to open its systems to the defense expert, as it had done for the prosecution. When the bank refused, the defendant's bank statements were ruled inadmissible, and the prosecution case collapsed. Thus, if a system is to be useful as a source of evidence, then it must be designed to withstand examination by hostile experts. I'll have more to say on this in Chapter 23.

The hostile expert problem isn't something we can expect to go away anytime soon. In countries where experts are appointed by the court, the risk is that they will be from the developer community, and so may have an interest in defending the system that they are supposed to be examining dispassionately. In general, we can expect computer forensics to remain a hard problem.

21.5.3 Electronic Signatures

In this generally unsatisfactory environment, many people hope that things can be simplified by gee-whiz technologies such as *electronic signatures*. This term encompasses (among other things) cryptographic digital signatures and alternative technologies such as tablets on which users scribble copies of their manuscript signatures to record assent to a document. In some cases, such as the U.S. Electronic Signatures in Global and National Commerce Act, the objective is to give legal force to any "sound, symbol, or process" by which a consumer assents to something. By pressing a telephone keypad ("Press 0 to agree or 9 to terminate this transaction"), clicking a hyperlink to enter a Web site, or clicking "Continue" on a software installer, the consumer consents to be bound to an electronic contract [709].

In many jurisdictions, this is already the case. In both the United States and England, the defining attribute of a signature is the signer's intent, and a plaintext name at the bottom of an email message has legal force. It may be easy to forge, but then so are the manuscript signatures that have been used for centuries [810, 811].

However, as I discussed in the section on handwritten signatures (13.2) there are many specific requirements that particular types of transaction—real estate, patent, copyright—be in writing, and these can hold up the adoption of online systems. Some countries, like Australia, have simply passed laws stating that electronic writing is OK wherever manuscript writing was required in the past; others, like Britain, have passed laws giving the government the power to issue regulations causing this to happen; still others, such as Germany, have made laws giving effect to digital signatures provided they meet laid-down technical standards. Such laws often suffer from multiple objec-

tives. Britain, for example, wants to promote the use of software and systems that support key escrow, while Germany wants to support its smartcard industry.

The laws passed in various American states are less tainted with ulterior motives, but still create a confusing and contradictory mosaic. Sometimes, digital signatures are enabled for general use, and sometimes for limited purposes such as communicating with the state government. Sometimes they're technology specific and sometimes they're not. For surveys of digital signature laws, see [68, 335].

Efforts are now underway to sort out the mess. The European Union issued an Electronic Signature Directive, which came into force in January 2000, that requires member states to introduce compatible legislation to recognize digital signatures as the legal equivalent of manuscript signatures. The directive sets out two different standards: an *electronic signature* means data attached to or logically associated with other electronic data and that serve as a method of authentication, while an *advanced electronic signature* must also:

- Be uniquely linked to the signatory.
- Be capable of identifying the signatory.
- Be created using means that the signatory can maintain under his sole control.
- Be linked to the data to which it relates in such a manner that any subsequent change of the data is detectable.

The basic idea is that an electronic signature includes a name typed at the bottom of an email, or a push of a Web page button to assent to a deal, while the advanced variety means use of a digital signature or biometric device. Lawmakers and people writing contracts should therefore be able to distinguish, using terms that are uniform across Europe, between weak and strong signature mechanisms.

One embarrassing problem is that the third of these requirements can't be met by currently available consumer electronics technology. Given the large number of ways in which a PC can be subverted, it would be very imprudent to have a signing key on your PC that could bind you for more than a small sum of money. Smartcards don't help; a villain who can write a virus to infect your PC and sign messages with a key in your browser software can just as easily infect the device driver of your smartcard reader to get the bogus message signed next time you insert the card. Also, if the card can be used in a parking meter as well as to mortgage your house, then you are extending to the parking meter the level of trust you'd normally restrict to your spouse or your lawyer. In the absence of secure platforms, some protection can be got from the traditional practice of having separate cards or other tokens for different types of transaction, so that the customer can keep the valuable ones under lock and key. (But personally I don't see any benefit in having an electronic means of performing a transaction I do at intervals of many years such as mortgaging my house.)

In the words of Bohm, Brown and Gladman, "It is of course no fatal reproach to the Directive that it should thus deliver thunder with no lightning; and it could be excused on the basis that the law will for once be ahead of events" [124]. However, there's enough wriggle room for countries to tack on interpretations and regulations that will deem products from their own smartcard or other suppliers to be adequate; and the current indications are that a moderately good smartcard will do, even if it's used with an insecure PC. (I'll discuss this in more detail in Section 23.3.3.1 below.) Businesses in other countries will then have to accept the resulting "advanced" signatures as valid.

So what sort of risks will people run once we have digital signatures that are considered by judges to be totally secure even although they aren't?

21.5.4 Burden of Proof

There is an even deeper problem with most digital signature laws (including those of many U.S. states). This is that they create a presumption that a digital signature meeting certain criteria (authorized type of smartcard, public key certified by licensed TTP, whatever) is valid. This flies in the face of traditional business practice, in which the risk that a signature is forged falls on the party who relies on it rather than on the party who made it.

If a bank debits your account with payment of a check that you did not sign, it has no authority for the debit and must credit the money back to you. In general, if someone wishes to enforce a document against you on the basis that you signed it, and you deny that you signed it, then it is for them to prove that the signature was made or authorized by you. This means that banks and merchants can decide for themselves how much care to take when verifying signatures; if they decide to verify signatures only for amounts over $1,000, or even $10,000, that is their concern, and has nothing to do with the customer. I discussed the error rates of handwritten signatures at 13.2; in practice the associated risks are manageable. In Chapter 19, I explained that essentially the same happens with credit cards, although there the customer typically bears the first $50 of the risk and in return gets the ability to pursue a claim against the card issuer if the merchant goes bust or otherwise fails to deliver.

It is understandable that banks and merchants would like to offload their exposure, and digital signature laws have been held out as a means of doing this. As described in Chapter 19, VISA and MasterCard went as far as to design the SET protocol to support credit card payment via digital signatures; and a number of governments dangled the bait of a presumption of validity of digital signatures as a way to get key escrow adopted [132].

Clearly, this is a bad thing from the customer's point of view. What's less obvious is that any temptation for the banks to use new technical security measures to dump risks on the customer should also be resisted in the wider interest of public confidence in electronic commerce and of the banking industry itself. This isn't just a digital signature issue. In the U.K., when it turned out that people who'd accessed electronic services at Barclays Bank via a public terminal could be hacked by the next user pressing the Back button on the browser, the bank tried to blame customers for not clearing their Web caches [747]. If opposing that in court, I'd have great fun finding out how many of Barclays' branch managers knew what a cache is, and the precise date on which the bank's directors had it brought to their attention that such knowledge is now essential to the proper conduct of retail banking business.

It is predictable that such risk dumping will reduce the motivation banks have to build secure systems, and will in time lead to injustices that neither the courts nor public opinion will tolerate. Recall that banks in some countries dumped the risks of ATM systems on customers, and claimed that any customers who complained of "phantom withdrawals" were mistaken or lying, then were greatly inconvenienced when the courts destroyed their fiction by sending ATM fraudsters to jail. The banks seem to be slow learners, and the ATM mistakes are likely to be repeated on a very much grander scale if digital signatures made by customers start being accepted as gospel in business-to-consumer e-commerce transactions.

Even in advance of the deployment of digital signatures, a number of banks have adopted electronic banking terms and conditions under which their records of a transaction are definitive; they are already getting into trouble under consumer law and truth-in-advertising regulations. These issues are discussed in detail in [124].

Business-to-business is a different matter, and as discussed in 19.5.4 there have been systems fielded for some years that use digital signatures in applications such as interbank funds transfer, registration of securities, and bills of lading. This appears likely to be the main application of digital signature technology, at least in the short term. One might assume that large businesses either have the expertise to secure the systems that they use to generate signatures or to pay others to do so. But disputes will still arise, especially with small businesses that don't have these resources. The liability for a forged digital signature could be particularly difficult for the courts to pin down, given the refusal of most software companies to accept any liability at all for security failures and even just plain bugs in their products. Therefore, the prudent thing for an e-commerce system designer to do is to set out in the subscribers' contract a procedure for dispute resolution, which should be sufficiently fair to withstand furious legal challenges once the first frauds occur.

21.6 Other Public Sector Issues

A whole grab-bag of other public sector information security issues are appearing. They vary from one country to another; I'll just give a few examples.

21.6.1 Service Delivery

A typical government department, such as a welfare agency or a passport office, has the operation and maintenance of a large distributed system as its core business function. Yet governments have usually been bad at conceiving and implementing large IT projects. Many of the reasons are well known. The civil service doesn't pay very well, so can't usually compete for the brightest IT staff; many government departments have traditional ways of doing things that don't automate well; planning and purchasing cycles are long compared with technology cycles; the managerial culture is more risk-averse than is ideal; and outside a few specialized functions, it isn't easy to set up a dozen competing organizations and just let the market sort them out. Many of the things I've seen go wrong with public sector projects have at their heart the culture clash between the computer business and the civil service. This problem cuts both ways, of course: civil servants tend to see computer people as impossibly ambitious and pushy people who want to disturb time-honored political fudges and eliminate discretionary powers in the interests of automation.

This clash will worsen as the growth of the Net places more severe strains on civil service administrative capabilities. The political leadership expects that government services will be delivered online, and that service levels will rise to somewhere near those of the private sector. The voters expect no less. Yet often automation makes

problems worse. In Britain, for example, the National Health Service suppresses demand for healthcare in a number of ways that doctors have evolved over time, such as by making it difficult to get appointments to see a specialist. Recently, ministers have started to insist that patients be able to book a specialist appointment over the phone, and the predictable result is a sudden rise in demand with no corresponding increase in supply. In the absence of a working price mechanism, this is a recipe for chaos. Already, we see signs of specialist doctors heading for early retirement as a response to rising pressure to treat more patients.

The relevance for the security engineer is that many things that are claimed to be impossible on "security" or "privacy" grounds are really demand-suppression issues. A lack of sensitivity to this can make a sale of your "solution" unlikely, or its side effects unpleasant. So you should always try to dig beneath the surface excuses and find out what your prospective clients' real concerns are.

21.6.2 Social Exclusion and Discrimination

A separate set of issues cluster around the delivery of government services to the poor and the old, in the belief that they are much less likely to be online and therefore face a reduced quality of support. The British government, for example, wants public-access Internet terminals made available in libraries and post offices [758]. In effect, this will provide subsidized public-sector competition for Internet cafes.

We'll just have to wait and see whether this catches on. But while Internet use has tended in the past to be the preserve of young affluent white males, it's not altogether clear that this will continue. Women and seniors are among the fastest-growing sectors of Net usage, and the integration of mail and browser facilities into satellite TV is bringing the Net to Joe Sixpack too. Perhaps the interesting question for the security engineer is the extent to which public terminals open up interesting new attacks. We saw in 21.5.4 how systems can be attacked using information kept in caches and the like; there are many pitfalls here.

Another security engineering issue related to equality of access is that many of the assumptions embedded in protection mechanisms can discriminate in ways that may be illegal or at least undesirable. Section 13.8 described how many biometric authentication systems may be regressive, in that the elderly and manual workers can suffer higher error rates with fingerprint readers, and that disabled people with no fingers, or no eyes, risk exclusion if fingerprint or iris scanning systems become widespread. Blind people are already seriously prejudiced in their use of the Web by many of the tricks used by website designers to prevent their pages being scanned by comparison shopping bots—which from the site owner's viewpoint are security measures.

Intrusion detection systems are another contentious area; as discussed in 18.5.2, automatic systems that detect fraud, or bogus insurance claims, or airline passengers likely to be terrorists, often end up discriminating against some ethnic or social group. Another issue is that systems designed for, and by, college-educated computer scientists are often too hard for less educated people to use. The attitude that users are a nuisance must be vigorously resisted; secure systems, like any other systems, need to be designed for the people who will actually use them. Replacing the word "user" with "customer" or "citizen" is a small step in the right direction.

21.6.3 Revenue Protection

One of the most high-profile concerns is that a combination of anonymous remailers, digital cash and offshore tax havens might make the task of collecting taxes impossible, leading to the breakdown of the system of nation states. This is perhaps most cogently expressed by Neal Stephenson [736] but has found echoes in much other commentary. This tends to ignore the fact that many countries get most of their revenue from sales taxes and customs duties; European readers used to paying over $5 a gallon for gas and $20 for a bottle of whisky will be much more sceptical about this vision.

21.6.4 Elections

Finally, the most fundamental process in any democracy is the conduct of elections. If this is undermined, the whole structure may collapse. I sincerely hope that the election of security chief Vladimir Putin as the president of Russia had nothing to do with the fact that the national electoral reporting system is run by FAPSI, a Russian signals intelligence agency formed in 1991 as the successor to the KGB's 8th and 16th directorates. Its head, General Starovoitov, was reported to be an old KGB type; his agency reported directly to President Yeltsin, who chose Putin as his successor [327, 430].

I would certainly be concerned if Britain were to introduce an electronic election system, and if CESG, the part of GCHQ that is our "national technical authority" for information protection, had anything to do with its design or audit. I mentioned in the introduction to this chapter that the U.K. policy of escrowing all public sector keys could cause serious problems here: even if the agencies don't actually manipulate the result, they will be sorely tempted to find out who voted for parties such as Sinn Féin. But where are the alternative centers of expertise?

The situation in the United States is perhaps not so worrying, because control over elections is very widely distributed, with accreditation state by state and hundreds of legacy systems in the field. But complacency isn't advisable. The sheer cost of obtaining accreditation in fifty states (over $100,000 a state to have design and source code checked by an independent expert) will limit the number of companies that can make a serious bid to provide the online successors to the current local systems. The disputes over the 2000 election may also drive state legislators to embrace "modern" online systems without stopping to think. If one or two companies end up controlling voting in all or most of the states, they will bear close watching.

21.7 Summary

Governments and public policy concerns generally are intruding more and more into the work of the security engineer. The legal controls on cryptography in many countries are just the most obvious example. Although misguided, these controls have a number of pernicious and unobvious effects, of which the worst may be the erosion of the boundary between law enforcement and intelligence. Other boundaries whose ero-

sion could threaten civil liberties include those between traffic analysis and wiretaps for content, between copyright and censorship, and between mechanisms that enforce copyright and those that do other things as well, such as accessory control. Working for increased transparency might be more strategic than taking issue with particular technologies.

There are many other issues though. The engineer must pay attention to the protection of personal data, the quality of evidence a system produces, copyright law issues, social exclusion and discrimination. There are also some mechanisms that we really must get right, such as the integrity of the systems used to record and tally votes in elections.

Research Problems

Technopolicy issues tend to involve a complex interplay between science, engineering, applied psychology, law and economics. There is altogether too little serious cross-disciplinary research; the apothegm at the head of this chapter captures well the problem that people from these different disciplines often talk at cross purposes. Debates on issues such as key escrow are slowly building a body of people with experience in talking to both computer scientists and lawyers; and electronic commerce leads computer scientists to talk to economists. Initiatives that speed up this process are almost certainly a good thing; bringing in psychologists, historians of science, and others would also be positive. What's not clear is how to do this within the current structures of academic and industrial research organizations.

Further Reading

It's extraordinarily easy for technopolicy arguments to get detached at one or more corners from reality; and many of the nightmares conjured up to get attention and money (such as credit card transactions being intercepted on the Internet) are really the modern equivalent of the monsters that appeared on medieval maps to cover up the cartographer's ignorance. An engineer who wants to build things that work and last has a duty not to get carried away. For this reason, it's particularly important to dig out primary sources—material written by experienced insiders such as R.V. Jones [425] and Gerard Walsh [787], books by people with a long involvement in the policy process such as Whitfield Diffie and Susan Landau [250], government reports that were influential in policy formation such as the NRC study on cryptography policy [580], and compilations of primary materials, such as [684].

There's also useful material at the Web sites of organizations such as EPIC [266], EFF [264], FIPR [304], CDT [173], the Privacy Exchange [628], and on mailing lists such as politech [619] and ukcrypto [755].

The best book I know on computer forensics is by Tony Sammes and Brian Jenkinson [664]; and there's a nice article by Peter Sommer on the forensics and evidential issues that arose when prosecuting some U.K. youngsters who hacked the USAF Rome airbase [722]. The Department of Justice's "Guidelines for Searching and Seizing Computers" also bear some attention [245]. For collections of computer crime case histories, see Peter Neumann [590], Dorothy Denning [235], and Donn Parker [602].

On the topic of data protection, there is a huge literature, but no concise guide that I know of. [802] Alan Westin provides a good historical overview, with a perspective on the coming collision between Europe and the United States. Simson Garfinkel [330] and Michael Froomkin [323] survey U.S. privacy and surveillance issues.

There's now quite a literature on electronic voting. The issues are largely the same as with voting by mail or by phone, but not quite. An influential survey of the requirements, and of the things that can go wrong, is by Mike Shamos [693]; while Roy Saltman (for many years the authority at NIST) discusses things that have gone wrong in the United States and various NIST recommendations, in [663]. There's a report on the feasibility of Internet voting from the State of California at [152]. Finally, Lorrie Cranor has a useful link farm on electronic voting at [209].