CHAPTER

8

Multilateral Security

Privacy is a transient notion. It started when people stopped believing that God could see everything and stopped when governments realized there was a vacancy to be filled.

—ROGER NEEDHAM

You have zero privacy anyway. Get over it.

—SCOTT MCNEALY

8.1 Introduction

Often, our goal is not to prevent information flowing “down” a hierarchy but to prevent it flowing “across,” between departments. Relevant applications range from healthcare to national intelligence, and include most applications where the privacy of individual customers’, citizens’ or patients’ data is at stake. They account for a significant proportion of information processing systems, but their protection is often poorly designed and implemented. This has led to a number of expensive fiascos.

In such systems, instead of the information flow-control boundaries being horizontal, as in the Bell-LaPadula model (Figure 8.1) we instead need the boundaries to be mostly vertical, as shown in Figure 8.2.

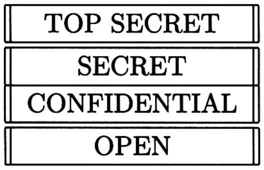


Figure 8.1 Multilevel security.

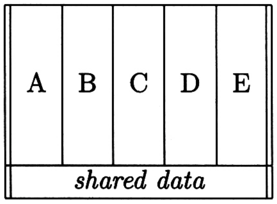


Figure 8.2 Multilateral security.

These lateral information flow controls may be organizational, as in an intelligence organization that wants to keep the names of agents working in one foreign country secret from the department responsible for spying on another. They may be privilege-based, as in a law firm where different clients' affairs, and the clients of different partners, must be kept separate. They may even be a mixture of the two, as in medicine where patient confidentiality is based in law on the rights of the patient, but usually enforced by limiting medical record access to a particular hospital department.

The control of lateral information flows is a very general problem, of which I'll use medical systems as a clear and well-studied example. The problems of these systems are readily understandable by the nonspecialist, and have considerable economic and social importance. Much of what we have to say about them goes across with little or no change to the practice of other professions and to government applications where access to particular kinds of classified data are restricted to particular teams or departments.

One minor problem is that of terminology. Information flow controls of the type we're interested in are known by a number of different names; in the U.S. intelligence community, for example, they are known as *compartmented security* or *compartmentation.* I will use the European term *multilateral security*, as the healthcare application is bigger than intelligence, and the latter term also covers the use of techniques such as anonymity—the classic case being de-identified research databases of medical records. This is an important part of multilateral security. As well as preventing overt information flows, we also have to prevent information leakage through, for example, statistical and billing data that get released.

The use of de-identified data has wider applicability. Another example is the processing of census data. In general, the relevant protection techniques are known as *inference control.* Despite occasional differences in terminology, however, the problems facing the operators of census databases and medical research databases are very much the same.

8.2 Compartmentation, the Chinese Wall, and the BMA Model

There are (at least) three different models of how to implement access controls and information flow controls in a multilateral security model. These are *compartmentation*, used by the intelligence community; the *Chinese Wall* model, which describes the mechanisms used to prevent conflicts of interest in professional practice; and the *BMA model*, developed by the British Medical Association to describe the information flows permitted by medical ethics. Each of these has potential applications outside its field of origin.

8.2.1 Compartmentation and the Lattice Model

For many years, it has been standard practice in the United States and allied governments to restrict access to information by the use of codewords as well as classifications. The best-documented example is the codeword *Ultra* used during World War II, to refer to British and American decrypts of German messages enciphered using the

Enigma machine. The fact that the Enigma had been broken was so important that it was worth protecting at almost any cost. So Ultra clearances were given to only a small number of people (in addition to the cryptanalysts and their support staff, the list included the Allied leaders, their senior generals, and handpicked analysts.) No one who had ever held an Ultra clearance could be placed at risk of capture; and the intelligence could never be used in such a way as to let Hitler suspect that his principal cipher had been broken. Thus, when Ultra told of a target, such as an Italian convoy to North Africa, the Allies would send over a plane to "spot" it and report its position by radio an hour or so before the attack. This policy was enforced by special handling rules; for example, Churchill got his Ultra summaries in a special dispatch box, to which he had a key but his staff did not. Because such special rules may apply, access to a codeword is sometimes referred to as an *indoctrination*, rather than simply a clearance. (Ultra security is described by David Kahn [429] and Gordon Welchman [800].)

Much the same precautions are in place today to protect information whose compromise could expose intelligence sources or methods, such as agent names, cryptanalytic successes, the capabilities of equipment used for electronic eavesdropping, and the performance of surveillance satellites. The proliferation of codewords results in a large number of compartments, especially at classification levels above Top Secret.

One reason for this is that classifications are inherited by derived work; so a report written using sources from 'Secret Desert Storm' and 'Top Secret Umbra' can in theory only be read by someone with a clearance of 'Top Secret' and membership of the groups 'Umbra' and 'Desert Storm'. Each combination of codewords gives a compartment, and some intelligence agencies have over a million active compartments. Managing them is a significant problem. Other agencies let people with high-level clearances have relatively wide access. But when the control mechanisms fail, the result can be disastrous; in the Aldrich Ames case, a CIA officer who had accumulated access to a large number of compartments by virtue of long service and seniority, and because he worked in counterintelligence, was able to betray almost the entire U.S. agent network in Russia.

Codewords are, in effect, a pre-computer way of expressing access control groups, and can be dealt with using a variant of Bell-LaPadula, called the *lattice model.* Classifications together with codewords form a lattice a mathematical structure in which any two objects A and B can be in a dominance relation $A > B$ or $B > A$. They don't have to be: A and B could simply be incomparable (but in this case, for the structure to be a lattice, they will have a least upper bound and a greatest lower bound). As an illustration, suppose we have a codeword, say, 'Crypto'. Someone cleared to 'Top Secret' would be entitled to read files classified 'Top Secret' and 'Secret', but would have no access to files classified 'Secret Crypto' unless he or she also had a crypto clearance. This can be expressed as shown in Figure 8.3.

In order for information systems to support this, we need to distill the essence of classifications, clearances, and labels into a security policy that we can then use to drive security targets, implementation, and evaluation. As it happens, the Bell-LaPadula model appears to go across more or less unchanged. We still have information flows between High and Low as before, where High is a compartment that dominates Low. If two nodes in a lattice are incompatible—as with 'Top Secret' and 'Secret Crypto' in Figure 8.3—then there should be no information flow between them at all.

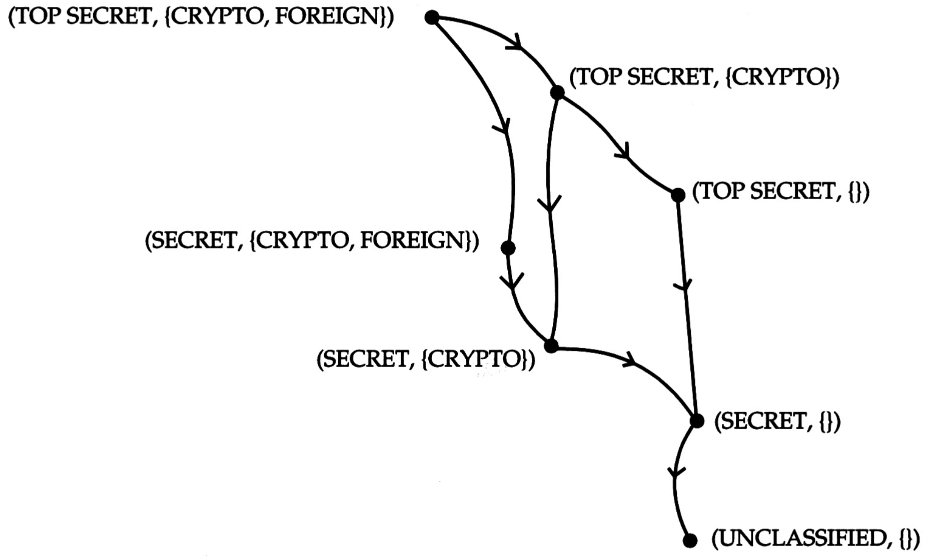


Figure 8.3 A lattice of security labels.

In fact, the lattice and Bell-LaPadula models are essentially equivalent, and were developed at the same time.

- Roger Schell, Peter Downey, and Gerald Popek of the U.S. Air Force produced an early lattice model in 1972 [675].
- A Cambridge PhD thesis by Jeffrey Fenton included a representation in which labels were managed using a matrix [289].
- About this time, the Pentagon's World Wide Military Command and Control System (WWMCCS) used a primitive lattice model, but without the *-property. As I noted in Chapter 7, the demonstration that a fielded, critical, system handling Top Secret data was vulnerable to attack by Trojan caused some consternation [674]. It meant that all users had to be cleared to the highest level of data in the machine.
- Kenneth Walter, Walter Ogden, William Rounds, Frank Bradshaw, Stan Ames, and David Shumway of Case Western University produced a more advanced lattice model, as well as working out a lot of the problems with file and directory attributes, which they fed to Bell and LaPadula [788, 789].[1]
- Finally, the lattice model was systematized and popularized by Dorothy Denning [233].

[1] Walter and his colleagues deserve more credit than history has given them. They had the main results first [788], but Bell and LaPadula had their work heavily promoted by the U.S. Air Force. Fenton has also been largely ignored, not being an American.

Most products built for the multilevel secure market can be reused in compartmented mode. But, in practice, these products are not as effective as one might like. It is easy to use a multilevel operating system to keep data in different compartments separate—just give them incompatible labels ('Secret Tulip', 'Secret Daffodil', 'Secret Crocus', etc.). But the operating system then becomes an isolation mechanism, rather than a sharing mechanism; the real problem is how to control information sharing.

One solution is to impose least upper bounds in the lattice using some algorithm. An example comes from the system used by the government of Saudi Arabia to manage the Haj, the annual pilgrimage to Mecca [385]. While most compartments are by default Confidential, the combination of data from different compartments is Secret. Thus, 'Haj-visas' and 'Gov-guest' are confidential, but their combination is Secret.

In many intelligence systems, where the users are already operating at the highest level of clearance, data owners don't want a further classification level at which everything is visible. So data derived from two compartments effectively creates a third compartment using the lattice model. The proliferation of millions of compartments is complex to manage and can be intertwined with applications. A more common solution is to use a standard multilevel product, such as a mail guard, to ensure that "untrustworthy" email goes to filters. But now the core of the trusted computing base consists of the filters rather than the guard.

Worse, the guard may lose some of the more important functionality of the underlying operating system. For example, the Standard Mail Guard [715] is built on top of an operating system called LOCK whose basic mechanism is *type enforcement*, which in this context can be thought of as a system of unchangeable access rules for processes and files. Later versions of LOCK support role-based access control, which would be a more appropriate mechanism to manage the relationships between compartments directly [386]. Using it merely as a platform to support BLP is wasteful.

In general, the real problems facing users of intelligence systems have to do with combining data in different compartments, and downgrading it after sanitization. Multilevel and lattice security models offer little help here.

8.2.2 The Chinese Wall

The second model of multilateral security is the Chinese Wall model, developed by Brewer and Nash [137]. Its name comes from the fact that financial services firms such as investment banks have internal rules designed to prevent conflicts of interest, which they call Chinese Walls.

The model's scope is wider than just investment banking. Many professional and services firms have clients who may be in competition with each other: software vendors, advertising agencies, and accountants are other examples. A typical rule is that "a partner who has worked recently for one company in a business sector may not have access to the papers of any other company in that sector." So an advertising copywriter who has worked on, say, the Shell account, will not be allowed to work on any other oil company's account for some fixed period of time.

The Chinese Wall model thus features a mix of free choice and mandatory access control: a partner can choose which oil company to work for, but once that decision is taken their actions in that sector are completely constrained. It also introduces the concept of *separation of duty* into access control; a given user may perform transaction A or transaction B, but not both.

Part of the attraction of the Chinese Wall model to the security research community comes from the fact that it can be expressed in a way that is fairly similar to Bell-LaPadula If we write, for each object c, $y(c)$ for c's company and $x(c)$ for c's conflict-of-interest class, then, like BLP, it can be expressed in two properties:

The simple security property A subject s has access to c if and only if, for all $c\square$ that s can read, either $y(c) \square x(c\square)$ or $y(c) = y(c\square)$.

The *-property A subject s can write to c only if s cannot read any $c\square$ with $x(c\square) \neq \emptyset$ and $y(c) \neq y(c\square)$.

The Chinese Wall model made a seminal contribution to the theory of access control. It also sparked a debate about the extent to which it is consistent with the BLP tranquility properties, and some work on the formal semantics of such systems (see, for example, Foley [300] on the relationship with noninterference). There are also some interesting new questions about covert channels. For example, could an oil company find out whether a competitor that used the same investment bank was planning a bid for a third oil company by asking which specialists were available for consultation and noticing that their number had dropped suddenly?

In practice, however, Chinese Walls still are implemented using manual methods. One large software consultancy has each of its staff maintain an "unclassified" curriculum vitae containing entries that have been sanitized and agreed with the customer. A typical entry might be:

September 97–April 98: Consulted on security requirements for a new branch accounting system for a major U.S. retail bank.

This is not the only control. A consultant's manager should be aware of possible conflicts, and not forward the CV to the client if in doubt; if this fails, the client can spot potential conflicts from the CV; and if this also fails, then the consultant is duty-bound to report any potential conflicts as soon as they appear.

8.2.3 The BMA Model

Perhaps the most important, interesting, and instructive example of multilateral security is found in medical information systems. The healthcare sector spends a much larger share of national income than the military in developed countries; and although hospitals are still less automated, they are catching up fast.

Healthcare safety and (especially) privacy have become hot-button issues in many countries. In the United States, the debate over the privacy regulations being introduced by the Department of Health and Human Services under the Health Insurance Portability and Accountability Act is unsetting doctors, patients, privacy advocates, researchers, and marketers; final regulations are due out by the end of 2000. Austrians are arguing about whether to introduce a smartcard to record health insurance data in a portable way, and Germans (who already have such a smartcard) are deliberating the pros and cons of putting emergency medical information (such as current prescriptions and allergies) on the card, too. The main objection here is that if data currently held on a MedAlert bracelet, such as allergies, are moved to a smartcard, there is a significant risk to patients who fall ill in locations where there is no smartcard reader available, such as on an airplane or in a foreign country. Not all privacy-enhancing technologies are without risk!

Everywhere, people are arguing about whether privacy norms will have to be radically revised as genetic data become widely available. In Iceland, for example, a project to build a national medical database that will incorporate not just medical records but also genetic and genealogical data, so that inherited diseases can be tracked across generations, has caused an uproar.

The protection of medical information is also a model for protecting personal information of other kinds, such as that held on individual customers by banks, insurance companies, and government agencies. In all European countries (and in many others, including Canada and Australia) there are *data protection* laws that restrict the dissemination of such data. I'll discuss data protection law in Part 3; for present purposes, it's enough to note that for some classes of data (affecting health, sexual behavior and preferences, political and trade union activity, and religious beliefs) the *data subject* must either give consent to information sharing or have a right of veto. This raises the issue of how one can construct a security policy in which the access control decisions are taken not by a central authority (as in Bell-LaPadula) or by the system's users (as in discretionary access control) but by the data subjects.

We will look first at the access control aspects.

8.2.3.1 The Threat Model

Currently, the main threat to medical privacy is social engineering (which I mentioned briefly in Chapter 3). The typical attack on medical record privacy comes from a private detective who phones a doctor's office or health insurer with a plausible tale:

> *Hello, this is Dr. Burnett of the cardiology department at the Conquest Hospital in Hastings. Your patient Sam Simmonds has just been admitted here in a coma, and he has a funny-looking ventricular arrythmia. Can you tell me if there's anything relevant in his record?*

This kind of attack is usually so successful that in both the United States and Britain there are people who earn their living doing it [260]. (It's not restricted to health records: in June 2000, millionaire British government minister Lord Levy was acutely embarrassed after someone called the tax office pretending to be him and found out that he'd only paid £5000 in tax the previous year [638]. But the medical context is a good one in which to discuss it.)

In 1996, an experiment was done in England whereby the staff at a health authority (a government-owned insurer that purchases health care for a region or district) were trained to screen out such false pretext telephone calls. The most important element of the advice they were given was that they were to always call back—and not to a number given by the caller, but to the number in the phone book for the hospital or other institution where the caller claimed to work. It turned out that some 30 telephone enquiries a week were bogus. (At that time, there were about 200 health authorities; the advice given is described in [22].)

Training staff in this way is more important than most technical protection measures. But the best staff training in the world won't protect a system in which too many people see too much data. There will always be staff who are careless or even crooked; and the more records they can see, the more harm they can do.

In one high-profile case, a convicted child rapist working as an orthopedic technician at Newton-Wellesley Hospital in Newton, Massachusetts, was caught using a former employee's password to go through the records of 954 patients to get the phone numbers of girls to whom he then made obscene phone calls [136]. He ended up doing jail time. There are many more incidents of a less dramatic nature.

Even where staff behave ethically, a lack of technical understanding can lead to leaks. Old PCs sold on the second-hand market or given to schools often have recoverable data on their hard disks; most people are unaware that the usual delete command does not remove the file, but merely marks the space it occupies as reusable. In a recent headline case, a PC sold on the second-hand market by investment bank Morgan Grenfell Asset Management had recoverable files containing the financial dealings of ex-Beatle Paul McCartney [153]. There have been very similar problems with old health records. Even where staff are honest and conscientious, equipment can still get stolen; some 11 percent of U.K. family doctors have experienced the theft of a practice PC, and in one case two prominent society ladies were blackmailed over terminations of pregnancy following such a theft [23].

The likelihood that a resource will be abused depends on its value and the number of people who have access to it. Aggregating personal information into large databases increases both these risk factors at the same time. Put simply, we can live with a situation in which a doctor's receptionist has access to 2,000 patients' records: there will be abuse from time to time, but at a tolerably low level. However, if the receptionists of the 5,000 family doctors who might work with a large American HMO, or of the 32,000 in Britain's National Health Service, all had access to the records of tens of millions of patients, then abuse would be likely. In a notable recent case, the U.S. Veterans' Administration is being sued in a class action for violating the privacy of its 180,000 employees; their system makes part of their records visible to their colleagues (and to some patients). And privacy issues aren't limited to organizations that treat patients directly; some of the largest collections of personal health information are in the hands of health insurers and research organizations. I discuss their special problems in Section 8.3.

Lateral information flow controls are required even for systems on a much smaller scale. A good illustration comes from a hospital system whose designers believed that for reasons of safety, all staff should have access to all records. This design decision was influenced by lobbying from geriatricians and pediatricians, whose patients are often treated by a number of specialist departments in the hospital; they were frustrated by the incompatibilities between different departmental systems. The system was first fielded in England in Hampshire, where then health minister Gerry Malone had his parliamentary seat. The system made all lab tests performed for local doctors at the hospital's pathology lab visible to most of the hospital's staff. A nurse who had had a test done by her family doctor complained to him after she found the result on the hospital system at Basingstoke where she worked; this caused outrage among local medics, and Malone lost his seat in Parliament at the 1997 election (by two votes) [32].

There are many ad hoc measures that hospitals can take to improve the protection of existing systems. One of the most effective is to keep the records of former patients in a separate archive, and give only a small number of admissions staff the power to move records from there to the main system. Another is to introduce a *honey trap*, a number of bogus records with celebrity names. Reportedly, one Boston hospital uses "medical records" with the names of Kennedy family members for this purpose; staff who browse them can be identified and disciplined. A particularly ingenious proposal, due to Gus Simmons, is to investigate all staff who consult a patient record but do not submit a payment claim to the insurer within 30 days; this aligns the patient's interest in privacy with the hospital's interest in maximizing its income [23].

However, a patchwork of ad hoc measures isn't a good way to secure a system We need a proper access control policy, thought through from first principles and driven by a realistic model of the threats. Which policy is appropriate for healthcare?

8.2.3.2 The Security Policy

This question faced the British Medical Association (BMA) in 1995. The U.K. government had introduced an IT strategy for the National Health Service whose security policy was multilevel. The idea was that AIDS databases would be at a level corresponding to 'Secret'; normal patient records at 'Confidential'; and administrative data, such as drug prescriptions and bills for treatment, at 'Restricted'. It was soon realized that this wasn't going to work. For example, how should a prescription for AZT be classified? It's a drug prescription, so it should be 'Restricted'; it identifies a person as HIV positive, so it must be 'Secret'. So all the 'Secret' AZT prescriptions must be removed from the 'Restricted' file of drug prescriptions. The same goes for most of the other prescriptions, as they identify treatments for named individuals and so should be 'Confidential'. But then what use will the file of prescriptions be to anybody? Pretty well all it will contain will be prescriptions written by doctors for general surgery stocks.

A second problem—and one that's now becoming an issue in the United States—is that the strategy was based on the idea of a single *electronic patient record* (EPR) that would follow the patient around from conception to autopsy, rather than the traditional system of having different records on the same patient at different hospitals and doctors' offices, with information flowing between them in the form of referral and discharge letters. An attempt to devise a security policy for the EPR, which would observe existing ethical norms, became unmanageably complex [355].

In a project for which I was responsible, the BMA developed a security policy to fill the gap. The critical innovation was to define the medical record not as the total of all clinical facts relating to a patient, but as the maximum set of facts relating to a patient and to which the same staff had access. Thus, an individual patient may have more than one record, and this offended the "purist" advocates of the EPR. But multiple records are dictated anyway by law and practice. Depending on the country (and even the state) that you're in, you may have to keep separate medical records for human fertilization, sexually transmitted diseases, prison medical services, and even birth records (as they pertain to the health of the mother as well as the child, and can't simply be released to the child later without violating the mother's confidentiality). This situation is likely to get more complex still as genetic data start being used more widely.

In many countries, including all the members of the European Union, a special status is given to patient consent in law as well as in medical ethics. Records can be shared only with third parties if the patient approves, or in a limited range of statutory exceptions, such as tracing contacts of people with infectious diseases such as TB. Definitions are slightly fluid; in some countries, HIV infection is notifiable, in others it isn't, and in others the data are collected stealthily.

The goals of the BMA security policy were, therefore, to enforce the principle of patient consent, and to prevent too many people getting access to too large databases of identifiable records. It did not try to do anything new, but merely to codify existing best practice. It also sought to express other security features of medical record management such as safety and accountability. For example, it must be possible to reconstruct the contents of the record at any time in the past, so that, for example, if a malpractice suit is brought, the court can determine what information was available to the doctor at the time. (The details of the requirements analysis are in [23].)

The policy consists of nine principles:

1. Access control: each identifiable clinical record shall be marked with an access control list naming the people or groups of people who may read it and append data to it. The system shall prevent anyone not on the access control list from accessing the record in any way
2. Record opening: a clinician may open a record with herself and the patient on the access control list. Where a patient has been referred, she may open a record with herself, the patient and the referring clinician(s) on the access control list
3. Control: One of the clinicians on the access control list must be marked as being responsible. Only she may alter the access control list, and she may only add other health care professionals to it
4. Consent and notification: the responsible clinician must notify the patient of the names on his record's access control list when it is opened, of all subsequent additions, and whenever responsibility is transferred. His consent must also be obtained, except in emergency or in the case of statutory exemptions
5. Persistence: no-one shall have the ability to delete clinical information until the appropriate time period has expired
6. Attribution: all accesses to clinical records shall be marked on the record with the subject's name, as well as the date and time. An audit trail must also be kept of all deletions
7. Information flow: Information derived from record A may be appended to record B if and only if B's access control list is contained in A's
8. Aggregation control: there shall be effective measures to prevent the aggregation of personal health information. In particular, patients must receive special notification if any person whom it is proposed to add to their access control list already has access to personal health information on a large number of people

9. Trusted computing base: computer systems that handle personal health information shall have a subsystem that enforces the above principles in an effective way. Its effectiveness shall be subject to evaluation by independent experts.

This policy may seem to be just common sense, but it is surprisingly comprehensive and radical in technical terms. For example, it is strictly more expressive than the Bell-LaPadula model; it contains a BLP-type information flow control mechanism in principle 7, but also contains state. (A fuller discussion from the point of view of access control, and for a technical audience, can be found at [24].)

Similar policies were developed by other medical bodies, including the Swedish and German medical associations; the Health Informatics Association of Canada, and an EU project (these are surveyed in [469]). However, the BMA model is the most detailed and has been subjected to the most rigorous review; it was adopted by the Union of European Medical Organisations (UEMO) in 1996. (Feedback from public consultation on the policy can be found in [25].)

8.2.3.3 Pilot Implementations

In a top-down approach to security engineering, one should first determine the threat model, then write the policy, and finally test the policy by observing whether it works in real life.

BMA-compliant systems have now been implemented both in general practice [374], and in a hospital system which enforces access rules such as “a ward nurse can see the records of all patients who have, within the previous 90 days, been on her ward.” (The hospital system was initially designed independently of the BMA project. When we learned of each other we were surprised at how much our approaches coincided, and reassured that we had captured the profession’s expectations in a reasonably accurate way.)

One of the lessons learned was the difficulty of constructing a small trusted computing base. The hospital records system has to rely on the patient administrative system to tell it which patients and which nurses are on which ward. A different prototype system at a hospital in Cambridge, England, furnishes staff with certificates in smartcards, which they use to log on; combining the two ideas into authorization certificates for access to the records of patients in particular wards may well be the way forward; the support promised in Win2K for both groups and certificates is promising. As for the longer term, people are now researching ways in which medical privacy policy can be expressed using the formalisms and mechanisms of role-based access control. (Other lessons learned are discussed in [231, 232, 374].)

8.2.4 Comparative Analysis

Which of these three models—lattice, Chinese Wall and BMA—should be used in a given application? The lattice model on its own isn’t enough, as it shows how to isolate compartments but not how to manage information flows between them. Both BMA and Chinese Wall tackle this problem, but BMA is as decentralized as possible, while in Chinese Wall the assignment of access rights is centralized, and the resulting aggregation risk is managed by a more explicit mechanism to prevent any one user getting their hands on too much data.

There is surprisingly little difference in the protection requirements of medical data and intelligence data, or, for that matter, the files of lawyers, investment bankers, or advertising agents. Some will be the target of more capable motivated opponents, and will need stronger protection mechanisms; but strength of mechanisms should never be confused with functionality. In all these cases, the underlying threat model, of careless or dishonest insiders, is the same.

In fact, the fundamental policy decision is whether or not to centralize. Can you cope better with lots of little traitors or with one big traitor? Medics, lawyers, and other professionals prefer the former, while spies seem to prefer the drama of the latter.

8.3 Inference Control

Access control in medical record systems is hard enough in hospitals and other organizations that care for patients directly. It is much harder to assure patient privacy in secondary applications such as databases for research, cost control, and clinical audit. This is one respect in which doctors have a harder time protecting their data than lawyers; lawyers can lock up their confidential client files and never let any outsider see them at all, while doctors are under all sorts of pressures to share data with third parties.

8.3.1 Basic Problems of Inference Control in Medicine

The standard way of protecting such information is to remove patients' names and addresses from their records, and thus make them anonymous. But this is rarely sufficient. If a database allows detailed enough queries, then individuals can still be identified, and this is especially so if information about different clinical episodes can be linked. For example, if I am trying to find out whether a politician born on the June 2, 1946, and treated for a broken collar bone after a college football game on the May 8, 1967, had since been treated for drug or alcohol problems, and I could make an enquiry on those two dates, then I could very probably pull out a single medical record from a national database. Even if the date of birth is replaced by a year of birth, I am still likely to be able to compromise patient privacy if the records are detailed or if records of different individuals can be linked. For example, a query such as "show me the records of all women aged 36 with daughters aged 14 and 16, such that the mother and exactly one daughter have psoriasis" is also likely to narrow down the search to one family out of millions. And, complex queries with lots of conditions are precisely the kind that researchers want to make.

For this reason, the U.S. Healthcare Financing Administration (HCFA), which is responsible for paying doctors and hospitals for treatments provided under the Medicare program, maintains three sets of records. There are complete records, used for billing. There are *beneficiary-encrypted* records, with only patients' names and social security numbers obscured. These records are still considered personal data (as they still have dates of birth, postal codes and so on) and so are only usable by trusted researchers. Finally there are *public-access* records which have been stripped of identifiers down to the level at which patients are identified only in general terms such as 'a white female aged 70–74 living in Vermont.' Nonetheless, researchers have found that many patients can still be identified by cross-correlating the public access records with commercial databases, and following complaints by privacy advocates, a recent report from the General Accounting Office criticized HCFA for lax security [333].

Many other countries have healthcare monitoring systems that use similar technologies. New Zealand has a national database of encrypted-beneficiary medical records, with access restricted to a small number of specially cleared medical statisticians. No query is answered with respect to fewer than six records [584]. Germany has very strict privacy laws, and the fall of the Berlin Wall forced the former East German cancer registries to install protection mechanisms rapidly [118]. In other countries, protection has been less adequate. Britain's National Health Service started out with strict guidelines but then built a number of centralized databases that make personal health information widely available within government, and that have led to confrontation with doctors [32]. Similar systems in Switzerland were replaced at the insistence of local privacy regulators [685]. The most controversial of all has been a genetic database in Iceland, which I'll discuss shortly.

De-identifying personal information is important in many other fields. Under the rubric of *privacy enhancing technology* (PET), it is being promoted actively by regulators in Europe and Canada as a general privacy mechanism (along with smartcards, encryption, and a few other tools). But, as the medical examples show, there can be serious tension between the desire of researchers for detailed data, and the right of patients (or other data subjects) to privacy. It is important to understand what can, and what cannot, be achieved with this technology.

8.3.2 Other Applications of Inference Control

The inference control problem was first seriously studied in the context of census information. A census collects a vast amount of sensitive data about individuals, then makes statistical summaries of it available by geographical (and governmental) units such as regions, districts, and wards. This information is used not just in the general formulation of policy, but also in determining electoral districts and the levels of government funding, for public services for many years. The census problem is somewhat simpler than the medical record problem, as the data are rather restricted and in a standard format (age, sex, race, income, number of children, highest educational attainment, and so on).

There are two broad approaches, depending on whether the data are de-identified before or during processing—or equivalently whether the software that will process the data is untrusted or trusted.

An example of the first kind of processing comes from the treatment of U.S. census data until the 1960s. The procedure was that one record in a thousand was made available on tape—minus names, exact addresses, and other sensitive data. Noise was also added to the data to prevent people with some broader knowledge (such as of the salaries paid by the employer in a company town) from tracing individuals. In addition to the sample records, local averages were given for people selected by various attributes; and records with extreme values (such as very high incomes) were suppressed.

The reason for this might not be immediately obvious. But consider a wealthy family living in a small village. Their income might make a significant difference to the per capita village income, and thus be deduced on the assumption that the per capita income of the other villagers is no different from that in nearby villages. Hence the policy of excluding extreme values before averaging.

In the second type of processing, identifiable data are retained in a database, and privacy protection comes from controls on the kind of queries that may be made. Early attempts at this were not very successful, and various attacks were proposed on the processing used at that time by the U.S. census. The question was whether it was possible to construct a number of inquiries about samples containing a target individual, and work back to obtain supposedly confidential information about that individual.

If our census system allows a wide range of statistical queries, such as "tell me the number of households headed by a man earning between $50,000 and $55,000," "tell me the proportion of households headed by a man aged 40–45 years earning between $50,000 and $55,000," "tell me the proportion of households headed by a man earning between $50,000 and $55,000 whose children have grown up and left home," and so on, then an attacker can quickly home in on an individual. Such queries, in which we add additional circumstantial information to defeat averaging and other controls, are known as *trackers*. They are usually easy to construct.

A problem related to inference is that an opponent who gets hold of a number of unclassified files might deduce sensitive information from them. For example, a New Zealand journalist deduced the identities of many officers in GCSB (that country's equivalent of the NSA) by examining lists of service personnel and looking for patterns of postings over time [368]. Intelligence officers' cover postings might also be blown if an opponent gets hold of the internal phone book for the unit where the officer is supposed to be posted, and doesn't find his name there. The army list might be public, and the phone book 'Restricted', but the fact that a given officer is involved in intelligence work might be 'Secret'. Combining low-level sources to draw a high-level conclusion is known as an *aggregation attack*. It is clearly related to (but not the same as) the increased risk to personal information that arises when databases are aggregated together, thus making more context available to the attacker, and making tracker and other attacks easier. The techniques that can be used to counter aggregation threats are similar to those used for general inference attacks on databases, although there are some particularly difficult problems where we have a multilevel security policy, and the inference or aggregation threats have the potential to subvert it.

8.3.3 The Theory of Inference Control

A theory of inference control was developed by Dorothy Denning and others in late 1970s and early 1980s, largely in response to problems of census bureaux [234]. The developers of many modern privacy systems are unaware of this work, and repeat many of the mistakes of the 1960s. (Inference control is not the only problem in computer security where this happens.) The following is an overview of the most important ideas.

A *characteristic formula* is the expression (in some database query language) that selects a set, known as the *query set*, of records. An example might be "all female employees of the computer laboratory at the grade of professor." The smallest query sets, obtained by the logical AND of all the attributes (or their negations), are known as *elementary sets* or *cells.* The statistics corresponding to query sets may be *sensitive statistics* if they meet criteria which I will discuss below (such as the set size being too small). The objective of inference control is to prevent the disclosure of sensitive statistics.

If we let *D* be the set of statistics that are disclosed, and *P* the set of sensitive statistics that must be protected, then we need *D* □ *P*□for privacy, where *P*□is the complement of *P*. If *D = P'*, then the protection is said to be *precise.* Protection that is not precise will usually carry some cost in terms of the range of queries that the database can answer, and thus its usefulness to its owner.

8.3.3.1 Query Set Size Control

The obvious protection mechanism is simply to specify a minimum query size. As mentioned, New Zealand's National Health Information System databases will reject statistical queries whose answers would be based on fewer than six patients' records. But this is not enough in itself. An obvious tracker attack is to make an enquiry on six patients' records, then on those records plus the target's. Rather than reduce the effectiveness of the database by building in more restrictive query controls, the designers opted to restrict access to a small number of specially cleared medical statisticians.

Even so, one extra control is needed, and is often forgotten. We must prevent the attacker from querying all but one of the records in the database. In general, if there are *N* records, query set size control with a threshold of *t* means that between *t* and *N* – *t* of them must be the subject of a query for it to be allowed.

8.3.3.2 Trackers

Probably the most important attacks on statistical databases come from trackers. There are many simple examples. In our laboratory, only one of the full professors is female, so we can find out her salary with only two queries: "average salary professors?" and "average salary male professors?"

This is an example of an *individual tracker.* There are also *general trackers*, sets of formulae that will enable any sensitive statistic to be revealed. A surprising discovery made about trackers in the late 1970s was that, provided the minimum query set size *n* is less than a quarter of the total number of statistics *N*, and there are no further restrictions on the type of queries that are allowed, we can find formulae specifying sets with more than 2*n* and fewer than *N* – 2*n* statistics, and these provide general trackers. Thus, tracker attacks are easy, unless we place severe restrictions on the query set size or control the allowed queries in some other way.

8.3.3.3 More Sophisticated Query Controls

There are a number of alternatives to simple query set size control. The U.S. census, for example, uses the "*n*-respondent, *k*%-dominance rule": it will not release a statistic of which *k*% or more is contributed by *n* or fewer values. Other techniques include, as mentioned, suppressing data with extreme values. A census bureau may deal with high-net-worth individuals in national statistics, but not in the local figures, while some medical databases do the same for less common diseases. For example, a U.K. prescribing statistics system suppresses sales of the AIDS drug AZT from local statistics.

8.3.3.4 Cell Suppression

The next question is how to deal with the side effects of suppressing certain statistics. Suppose, for example, that a university wants to release average grades for various combinations of courses, so that people can check that the grading is fair across courses. Suppose now that the table in Figure 8.4, contains the number of students studying two science subjects, one as their major subject and one as their minor subject.

Next suppose that our minimum query set size is 3 (if we set it at 2, then either of the two students who studied geology with chemistry could trivially work out the other's grade); then we cannot release the average for geology with chemistry. But if the average for chemistry is known, then this can easily be reconstructed from the averages for biology with chemistry and physics with chemistry. Therefore, we have to suppress at least one other average in the chemistry row; and for similar reasons we need to suppress one in the geology column. But if we suppress geology with biology and physics with chemistry, then we'd also better suppress physics with biology to prevent these values being worked out in turn. The remaining table is shown in Figure 8.5.

Major:	Biology	Physics	Chemistry	Geology
Minor:				
Biology	-	16	17	11
Physics	7	-	32	18
Chemistry	33	41	-	2
Geology	9	13	6	-

Figure 8.4 Table containing data before cell suppression.

Major:	Biology	Physics	Chemistry	Geology
Minor:				
Biology	-	blanked	17	blanked
Physics	7	-	32	18
Chemistry	33	blanked	-	blanked
Geology	9	13	6	-

Figure 8.5 Table after cell suppression.

This process is called *complementary cell suppression.* If there are further attributes in the database schema—for example, if figures are also broken down by race and sex, to show compliance with anti-discrimination laws—then even more information may be lost. Where a database scheme contains *m*-tuples, blanking a single cell generally means suppressing $2^m - 1$ other cells, arranged in a hypercube with the sensitive statistic at one vertex. Clearly, even precise protection can rapidly make the database unusable. (where a database is not homogeneous, things are even worse: there can be many *pivot points*—cells that prevent large numbers of queries having answers.)

Sometimes complementary cell suppression can be avoided, as when large incomes (or rare diseases) are tabulated nationally and excluded from local figures, but it is often necessary when we are publishing microstatistics, as in the preceding tables of exam grades. Where the database is open for online queries, we can get much the same effect by *implied queries control*, whereby we allow a query on *m* attribute values only if all of the 2^m-implied query sets, given by setting the *m* attributes to true or false, have at least *k* records.

8.3.3.5 Maximum Order Control and the Lattice Model

The next thing we might try to make it harder to construct trackers is to limit the type of inquiries that can be made. *Maximum order control* limits the number of attributes that any query can have. However, to be effective, the limit may have to be severe. One study found that of 1,000 medical records, three attributes were safe; with four attributes, one individual record could be found; and with 10 attributes, most records could be isolated. A more thorough approach (where it is feasible) is to reject queries that would partition the sample population into too many sets.

We saw how lattices can be used in compartmented security to define a partial order to control permitted information flows between compartments with combinations of codewords. They can also be used in aslightly different way to systematize query controls in some databases. If we have, for example, three attributes *A*, *B*, and *C* (say, area of residence, birth year, and medical condition), we may find that, while inquiries on any one of these attributes are nonsensitive, as are inquiries on *A* and *B* and on *B* and *C*, the combination of *A* and *C* might be sensitive. It follows, that an inquiry on all three would not be permissible either. Thus, the lattice divides naturally into a top half of prohibited queries and a bottom half of allowable queries, as shown in Figure 8.6.

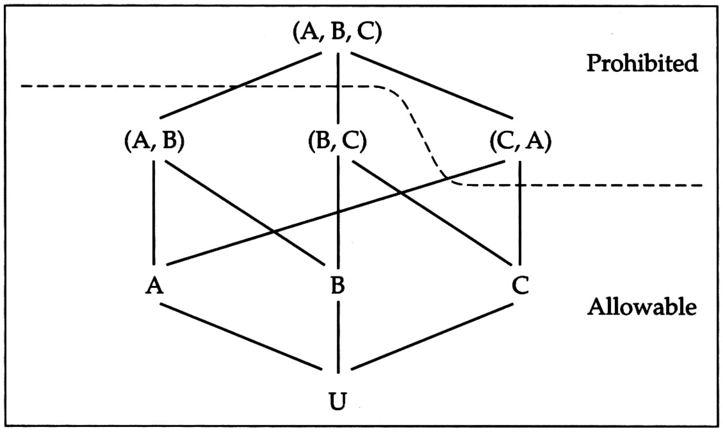


Figure 8.6 Table lattice for a database with three attributes.

8.3.3.6 Audit-Based Control

As mentioned, some people try to get round the limits imposed by static query control by keeping track of who accessed what. Known as *query overlap control*, this involves rejecting any query from a user that, combined with what the user knows already, would disclose a sensitive statistic. This may sound perfect in theory, but in practice it suffers from two usually unsurmountable drawbacks. First, the complexity of the processing increases over time, and often exponentially. Second, it's extremely hard to be sure that your users aren't in collusion, or that one user hasn't registered under two different names. Even if your users are all honest and distinct persons today, it's always possible that one of them will take over another, or that two of them get taken over by a predator, tomorrow.

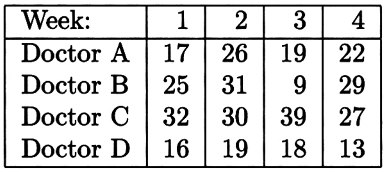
8.3.3.7 Randomization

The cell suppression example shows that if various kinds of query control are the only protection mechanism used in a statistical database, they can often have an unacceptable performance penalty. So query control is often used in conjunction with various kinds of randomization, which are designed to degrade the signal-to-noise ratio from the attacker's point of view while impairing that of the legitimate user as little as possible.

The simplest such technique is *perturbation*, or adding noise with zero mean and a known variance to the data. One way of doing this is to round, or truncate, the data by some deterministic rule; another is to swap some records. Perturbation is often not as effective as one would like, as it tends to damage the legitimate user's results precisely when the sample set sizes are small, and leave them intact when the sample sets are large (where we might have been able to use simple query controls anyway). There is also the worry that suitable averaging techniques might be used to eliminate some of the added noise.

Often, a better randomization technique is to use *random sample queries.* This is another of the methods used by census bureaux. The idea is to make all the query sets the same size, selecting them at random from the available relevant statistics. Thus all the released data are computed from small samples rather than from the whole database. If this random selection is done using a pseudorandom number generator keyed to the input query, then the results will have the virtue of repeatability. Random sample queries are a natural protection mechanism for large medical databases, where the correlations being investigated are often such that a sample of a few hundred is sufficient. For example, when investigating the correlation between a given disease and some aspect of lifestyle, the correlation must be strong before doctors will advise patients to make radical changes to their way of life, which might have undesirable side effects. If a teaching hospital has records on five million patients, and five thousand have the disease being investigated, then a randomly selected sample of two hundred sufferers might be all the researcher could use.

This doesn't work so well where the disease is rare, or where for other reasons there is only a small number of relevant statistics. A possible strategy here is *randomized response*, where we randomly restrict the data we collect (the subjects' responses). For example, if the three variables under investigation are obesity, smoking, and AIDS, we might ask each subject with HIV infection to record whether they smoke or whether they are overweight, but not both. Of course, this can limit the value of the data.



Week:	1	2	3	4
Doctor A	17	26	19	22
Doctor B	25	31	9	29
Doctor C	32	30	39	27
Doctor D	16	19	18	13

Figure 8.7 Sample of de-identified drug-prescribing data.

8.3.4 Limitations of Generic Approaches

As with any protection technology, statistical security can only be evaluated in a particular environment and against a particular threat model. Whether it is adequate or not depends to an even greater extent than usual on the details of the application.

An instructive example is a system used for analyzing trends in drug prescribing. Here, prescriptions are collected (minus patient names) from pharmacies. A further stage of de-identification removes the doctors' identities; the information is then sold to drug company marketing departments. The system has to protect the privacy of doctors as well as of patients (the last thing a busy family doctor wants is to be pestered by a drug rep for prescribing a competitor's brands).

One problem with an early prototype of this system was that it merely replaced the names of doctors in a cell of four or five practices with Doctor A, Doctor B, and so on, as in Figure 8.7. We realized that an alert drug rep could identify doctors from prescribing patterns, by noticing, for example, "Well, Doctor B must be Susan Jones because she went skiing in the third week in January, and look at the fall-off in prescriptions here. And Doctor C is probably her partner Mervyn Smith who would have been covering for her." The fix was to replace absolute numbers of prescriptions with the percentage of each doctor's prescribing that went to each particular drug, and to randomly perturb the timing by shifting the figures backward or forward a few weeks [530].

In general, contextual knowledge is extremely hard to quantify, and is quite likely to grow over time. Latanya Sweeney has shown that even the HCFA's "public-use" files can often be re-identified by cross-correlating them with commercial databases [744]. (Such *data detective* work is an important part of assessing the level of protection that an actual statistical database gives, just as we only have confidence in cryptographic algorithms that have withstood extensive analysis by capable motivated opponents.) And even without cross-correlation, there may be contextual information available internally. Users of medical research databases are often doctors who have normal access to parts of the patient record databases from which the statistical data are drawn.

8.3.4.1 Active Attacks

Active attacks are particularly powerful. These are where users have the ability to insert or delete records into the database. A user might add records to create a group that contains the target's record, plus those of a number of nonexistent subjects created by himself. One (imperfect) countermeasure is to add or delete new records in batches. Taking this to an extreme gives *partitioning*, whereby records are added in groups and any query must be answered with respect to all of them or none. However, this is once more equivalent to publishing tables of microstatistics.

Active attacks are not limited to data, but can also target metadata. A nice example, due to Whit Diffie, is the *chosen drug attack*. Suppose a drug company has access through a statistical system to the amounts of money spent on behalf of various groups of patients, and wants to find out which patients are receiving which drug, in order to direct its marketing better (there was a scandal in Quebec about just such an inference attack). A powerful trick is to set the drug prices in such a way as to make the resulting equations easy to solve.

A prominent case at the moment involves a new medical research database in Iceland, which comprises three linked databases: one with the nation's medical records, one with the genealogy of the whole population, and one with genetic data acquired from sequencing. The rationale is that since Iceland's population is largely descended from a few founding families that settled there about a thousand years ago, there is much less genie variance than in the general human population, and so genes for hereditary illnesses should be much easier to find.

The privacy problem in the Icelandic database is much more acute than in the general case. For example, by linking medical records to genealogies, which are in any case public (genealogy is a common Icelandic hobby), patients can be identified by such factors as the number of their uncles, aunts, great-uncles, great-aunts and so on—in effect, by the shape of their family trees. There was much debate about whether the design could even theoretically meet legal privacy requirements [33], and European privacy officials expressed grave concern about the possible consequences for Europe's system of privacy laws [217]. However, the Icelandic government pressed ahead with it anyway over the strong objections of local doctors. The result was that 11% of the population opted out of the system, including a majority of medical practitioners.

8.3.5 The Value of Imperfect Protection

So doing de-identification right is hard, and the issues can be politically fraught. But it is often worthwhile to make some attempt, even if the protection you can provide is imperfect.

Some kinds of security mechanism may be worse than useless if they can be compromised. Weak encryption is a good example. The main problem facing the world's signals intelligence agencies is how to filter out interesting nuggets from the mass of international phone, fax, email, and other traffic. A principal who helpfully encrypts his important traffic makes this part of his an opponent's job easier. If the encryption used is breakable (or one of the end systems can be hacked), then the net result is worse than if the traffic had been sent in clear.

Statistical security is not generally like this. The main threat to databases of personal information is often *mission creep*. Once an organization has access to data that are potentially valuable, then all sorts of ways of exploiting that value will be developed. Some of these are likely to be highly objectionable; a topical U.S. example is the resale of medical records to banks for use in filtering loan applications. However, even an imperfect de-identification system may destroy the value of medical data to a bank's loan department. If only five percent of the patients can be identified, and then only with effort, the bank may decide that it's simpler to tell loan applicants to take out their own insurance, and let the insurance companies send out medical questionnaires if they wish. So de-identification can help, even if the main effect is prophylaxis against future harm rather than treatment of existing defects.

As well as harming privacy, mission creep can have safety implications. In at least one European country, diabetic registers—databases designed to monitor the quality of diabetes care—are abused to provide a rudimentary means of electronic communication between family doctors and hospital diabetologists, who are frustrated at not having email. But as the diabetes registers were never designed as communications systems, they lack the safety and other mechanisms that they should have if they are to be used for this purpose. Even the most rudimentary form of de-identification would have prevented this abuse.

So in statistical security, the question of whether one should let the best be the enemy of the good can require a finer judgment call than elsewhere.

8.4 The Residual Problem

The two previous sections may have convinced you that the problem of managing medical record privacy in the context of immediate care (such as in a hospital) is reasonably straightforward, while in the context of secondary databases (such as for research, audit, and cost control) there are statistical security techniques that, with care, can solve much of the problem. Somewhat similar techniques are used to manage intelligence information in military organizations and for highly sensitive commercial data such as details of forthcoming mergers and acquisitions in an investment bank. In all cases, the underlying concept is that the really secret material is restricted to a compartment of a small number of identified individuals, and less secret versions of the data are manufactured for wider use. This involves not just suppressing the names of the patients, spies, or target companies, but also controlling any contextual and other information by which they might be re-identified.

But making such systems work well in real life is much harder than it looks. First, determining the sensitivity level of information is fiendishly difficult, and many initial expectations turn out to be wrong. You might expect, for example, that HIV status would be the most sensitive medical data there is; yet many HIV sufferers are quite open about their status. You might also expect that people would rather entrust sensitive personal health information to a healthcare professional such as a doctor or pharmacist rather than to a marketing database; yet many women are so sensitive about purchasing feminine hygiene products that, rather than going into a pharmacy and buying them for cash, they prefer to use an automatic check-out facility in a supermarket, even if this means they have to use their store card and credit card, so that the purchase is linked to their name and stays on the marketing database forever. The actual embarrassment of being seen with a packet of tampons is immediate, and outweighs the potential future embarrassment of being sent discount coupons for baby wear six months after the menopause.

Second, it is extraordinarily difficult to exclude single points of failure, no matter how hard you try to build watertight compartments. The CIA's Soviet assets were compromised by Rick Ames, who, as a senior man in counterintelligence, had access to too many compartments. The KGB's overseas operations were similarly compromised by Vassily Mitrokhin, an officer who had become disillusioned with communism after 1968, yet was sent to work in the archives while waiting for his pension [51].

In medicine, many of the really hard problems lie in the systems that process medical claims for payment. When a patient is treated, and a request for payment is sent to the insurer, it has not just full details of the illness, the treatment, and the cost, but also the patient's name, insurance number, and other details such as date of birth. There have been proposals for payment to be effected using anonymous credit cards [117], but as far as I am aware, none of them has been fielded. Insurers want to know which patients, and which doctors, are the most expensive. This holds whether the insurer is a private insurance company (or employer) or a government-owned health authority, such as HCFA or Britain's National Health Service. And once an insurer possesses large quantities of personal health information, it becomes very reluctant to delete it in case it might be useful or valuable in the future.

In the United States, the retention of copies of medical records by insurers, employers, and others is now widely seen as a serious problem. Writers from such widely different political viewpoints as the communitarian Amitai Etzioni [277] and the libertarian Simson Garfinkel [330] agree on this point, if on little else. Public concern spurred Congress to pass the Health Insurance Portability and Accountability Act (HIPAA), which empowered the Department of Health and Human Services (DHHS) to regulate the security of health data. The debate now is over how the regulations are to be implemented. If the private medical insurance sector were brought up to the standards of HCFA, this would probably be a good thing for most patients. But given the sums involved, one can anticipate a lot of foot-dragging and litigation. Even so, the act only enables the DHHS to regulate health plans, healthcare clearinghouses, and healthcare providers, leaving many organizations that process medical information (such as lawyers, employers, and universities) outside its scope.

What lessons can be drawn from other countries?

As we noted above, Britain's system has been a source of conflict with doctors and with patients' associations. The Swiss system, which was initially similar to Britain's, has now been de-identified much more thoroughly at the insistence of privacy regulators. In Germany, the richer people use private insurers (who are bound by tight data protection laws), while the poor use state health insurers, which are run by doctors, so non-doctors don't have access to records. The most radical solution is in Japan, where cost control is done by regulating fees: doctors are discouraged from performing expensive procedures, such as heart transplants, by pricing them below cost. This mechanism doesn't involve large-scale access to personal health information, and is much more effective than the case-by-case cost control practiced in most other countries. Healthcare takes up some 3 percent of GNP in Japan, versus 7 to 8 percent for the typical developed country, and 15 percent for America. Oh, and Japanese live longer than Europeans, who live longer than Americans. A variant of the Japanese solution was adopted in Oregon in February 1994 and proved popular with Oregonians, but was resisted fiercely by health industry lobbyists as "rationing."

To sum up, the problem of health record privacy is fundamentally a political one. Whether large quantities of medical records ever accumulate in one database depends on how the health care system is organized, and whether these are destroyed—or at least properly de-identified—after payment has been processed is a matter of regulation, not primarily of technology. In such debates, one role of the security engineer is to see to it that policymakers understand the likely consequences of their actions.

Other privacy problems also tend to have a serious political entanglement. Bank customer privacy can be tied up with the bank's internal politics; often the best privacy protection comes from branch managers' reluctance to let other branches learn about their customers. Access to criminal records and intelligence depends on how law enforcement agencies decide to share data with each other, and the choices they make internally about whether access to highly sensitive information about sources and methods should be decentralized (risking occasional losses), or centralized (bringing lower-probability but higher-cost exposure to a traitor at head office).

8.5 Summary

In this chapter, we looked at the problem of assuring the privacy of medical records. This is representative of a number of information security problems, ranging from the protection of national intelligence data through professional practice in general to the protection of census data.

It turns out that with medical records there is an easy problem, a harder problem, and a really hard problem.

The easy problem is setting up systems of access controls so that access to a particular record is limited to a sensible number of staff. Such systems can be designed largely by automating existing working practices. The harder problem is statistical security: how one designs databases of medical records (or census returns) so as to allow researchers to make statistical enquiries without compromising individuals' privacy. The hardest problem is how to manage the interface between the two, and in the specific case of medicine, how to prevent the spread of payment information. The only realistic solution for this lies in regulation.

Research Problems

In the near future, a lot of medical treatment may involve genetic information. Your medical records may involve personal health information about your parents, siblings, cousins, and so on. How can the BMA model be extended to deal with medical records that relate to multiple individuals?

Are there any ways of linking access control policies for privacy with statistical security with (perhaps) digital cash for payment? Can there be such a thing as seamless privacy where everything fits neatly together?

What other ways of writing privacy policies are there? For example, are there useful ways to combine BMA and Chinese Wall? Are there any technical or semi-technical ways of aligning the data subject's interest with others?

Further Reading

The literature on compartmented mode security is somewhat scattered: most of the public domain papers are in the proceedings of the NCSC/NISSC and AISSAC conferences cited in detail at the end of Chapter 7. Standard textbooks such as Amoroso [15] and Gollmann [344] cover the basics of the lattice and Chinese Wall models.

For the BMA model, see the policy document itself—the Blue Book [23], the shorter version at [24], and the proceedings of the conference on the policy [29]. See also the papers on the pilot system at Hastings [231, 232]. For more on Japanese healthcare, see [159]. For a National Research Council study of medical privacy issues in the United States, see [581]; there is also an HHS report on the use of de-identified data in research at [511].

For inference control, Denning's book [234] is the classic reference, and there's an update at [238]. A more modern textbook on database security is the one by Castano, et al. [172] whose chapter on statistical security is a useful update on Denning and whose other chapters also cover some related multilevel security and intrusion detection issues.