Security in Clinical Information Systems

Dr Ross J Anderson

Computer Laboratory
University of Cambridge
Pembroke Street
Cambridge CB2 3QG

4th January 1996

Version 1.1

Contents

1 Introduction

The proposed introduction of a nationwide NHS network has led to concern about security. Doctors and other clinical professionals are worried that making personal health information more widely available may endanger patient confidentiality [ACH95]. The problem is not limited to the NHS; it also concerns clinicians in prisons, immigration services, forensic laboratories and private healthcare. However the NHS network has forced the issues to the fore.

It has been generally agreed that the security of electronic patient records must meet or exceed the standard that should be applied to paper records, yet the absence of clarity on the proper goals of protection has led to confusion. The British Medical Association therefore asked the author to consider the risks, and to prepare a security policy for clinical information systems.

1.1 Scope of the policy

An information security policy says who may access what information; access includes such activities as reading, writing, appending, and deleting data. It is driven by a threat model and in turn drives the more detailed aspects of system design. To be effective, it needs to be written at the right level of abstraction; it must not encumber the reader with unnecessary details of specific equipment. It must tackle the important problems and ignore the distractions.

A potential distraction is the precise meaning of terms such as 'clinician', 'patient' and 'system'. One could dwell at length on what might happen when the clinician delegates a task to a student, or when the patient is a minor or deceased. These questions can be difficult but are, for our purposes, unimportant; so we shall clarify them here rather than in the body of the policy.

1.2 Definitions

By '**personal health information**', or equivalently '**identifiable clinical information**', we mean information that concerns a person's health, medical history or medical treatment (whether past or future) in a form that enables the person to be identified by a person other than the treating clinician [RAC+93].

By a '**clinician**', or equivalently '**clinical professional**' or '**healthcare professional**', we mean a licensed professional such as a doctor, nurse, dentist, physiotherapist or pharmacist, who has access in the line of duty to personal health information and is bound by a professional obligation of confidentiality. We include doctors working in public health, even though they may not technically be clinicians.

The reader may consult the Access to Health Records Act of 1990 for a legal definition of 'healthcare professional', but should be aware that it is controversial: there is debate about whether psychotherapists, telephone advice line staff, practitioners of complementary medicine and social workers should be brought inside the trust boundary. However the boundary has to be somewhere, and its precise location has little effect on our policy. Social workers, students, charity workers and receptionists may of course access personal health information under the supervision of a healthcare professional; but the professional remains responsible for their conduct. To keep things simple, we do not include such delegation in our security policy; but at the level of detailed design, it is wise for system builders to support delegation in intelligent ways.

Our use of '**patient**' will be a shorthand for 'the individual concerned or the individual's representative', in the sense of the draft BMA bill [BMA95]. In most cases this is the actual patient; but where the patient is a young child, it may be a parent or guardian who acts on his behalf. There are rules for patients who are unconscious or who have died, and even more complex rules for patients who are mentally incapacitated. The rules may depend on the previously expressed wishes of the patient, and they vary from one part of the UK to another [Som93]. We shall not discuss this area further.

For economy of expression, we will assume that the clinician is female and the patient male. The feminist versus grammarian issue is traditionally solved in the computer security literature by assigning definite gender roles, with the females being at least as high status as the males. Our choice is not meant to assert that the clinician has higher status than the patient in the therapeutic partnership between them.

By a '**system**' we generally mean the totality of hardware, software, communications and manual procedures which make up a connected information processing system. We are not concerned whether a system is made up of a single large mainframe with thousands of terminals, of thousands of PCs linked by a suite of protocols and distributed applications, or even from thousands of clerks moving pieces of paper around. We are only concerned with the net effect of the information processing; this is also the sense of the recent EU directive on data protection [EU95].

It should be clear from the context whether we are talking about the totality of interconnected clinical systems, or the subsystem which serves the needs of a particular individual or care team.

1.3 Disclaimers

Firstly, this document deals only with the clinical aspects of information security, and not with associated business aspects such as the commercial confidentiality of purchaser and provider contract data. and the legal reliability of electronic records in court. Secondly, we do not deny that there may be security gains in computerising medical records: encrypting records in transit can provide much stronger confidentiality than the postal service; intrusion detection systems can log accesses and analyse them for suspicious patterns; and offsite data backup can provide effective and economic protection against fire and flood. However we need to understand our protection priorities before these techniques can be applied effectively, and a security policy is an important step in creating and clarifying this understanding.

2 Threats and Vulnerabilities

In this section we discuss the threats to the security of personal health information that arise from computerisation and in particular from connecting together the many practice and hospital computers on which clinical records are currently stored. Firstly we review the security goals, then we consider what is likely to go wrong, and finally we set out our protection priorities.

2.1 The ethical basis of clinical confidentiality

The Hippocratic oath incorporated the principle of medical confidentiality into doctors' professional ethics. A modern statement can be found in the booklet 'Good Medical Practice' [GMC1] issued by the General Medical Council:

> Patients have a right to expect that you will not pass on any personal information which you learn in the course of your professional duties, unless they agree.

This is expanded in the GMC booklet 'Confidentiality' [GMC2] which stipulates that doctors who record or who are the custodians of confidential information must ensure that it is effectively protected against improper disclosure. Still more detailed guidance can be found in books published by the BMA [Som93] and HMSO [DGMW94].

Both the government and the healthcare unions are agreed that electronic health records must be at least as well protected as paper ones; the Data Protection Act makes GPs and others responsible for the security of personal health information that they collect; and a recent EU Directive obliges the government to prohibit the processing of health data except where the data subject has given his explicit consent, and in certain other circumstances [EU95].

The basic ethical principle, as stated by both the GMC and the EU, is that the patient must consent to data sharing. Confidentiality is the privilege of the patient, so only he way waive it [DGMW94]; and the consent must be informed, voluntary and competent [Som93]. Thus, for example, patients must be made aware that information may be shared between members of a care team, such as a general practice or a hospital department.

A number of exceptions to this rule have developed over time, and include both statutory requirements and exemptions claimed on pragmatic grounds; they pertain to the notification of abortions, births, some deaths, certain diseases, adverse drug reactions, non-accidental injuries, fitness to drive and disclosure to lawyers in the course of a dispute [DGMW94]. There is controversy over research; the NHSE claims that by seeking treatment, a patient gives implied consent to the use of his records in research, while the healthcare professions do not accept this [Mac94]. However, this debate has no great effect on the security policy set out here.

Finally, there is the issue of the patient's consent to have his record kept on a computer system at all. It is unethical to discriminate against a patient who demands that his records be kept on paper instead; his fears may well be justified if he is a celebrity, or a target for assassination, or for some other reason in danger from capable motivated opponents. Some cases of this kind have been managed using pseudonyms, so that the patient's real identity is never exposed to a computer system.

2.2 Other security requirements for clinical information

In addition to the confidentiality of clinical information, we are concerned with its integrity and availability.

If information is corrupted, clinicians may take incorrect decisions which harm or even kill patients. If information is unreliable, in the sense that it could have been corrupted (even if it has not been), then its value as a basis for clinical decisions is diminished. There is also the medico-legal concern that healthcare professionals called to justify their actions may not be able to rely on computer records in evidence; and there has recently been controversy over whether it is enough to have an electronic record alone, or whether paper or microfiche records should be kept as a backup.

If information systems are unreliable in the simpler sense that information may occasionally be unavailable as a result of system failure or sabotage, then this also diminishes their value and restricts the use which may prudently be made of them.

It is therefore prudent to look for ways to guarantee integrity for certain records, and to prevent attacks which might impact system availability.

2.3 Threats to clinical confidentiality

Many organisations, both public and private, have replaced dispersed manual record keeping systems with centralised or networked computer systems which give better access to data. Their experience is that the main new threat comes from insiders. For example, most of the big UK banks now let any teller access any customer's account; newspapers report that private detectives bribe tellers to get account information which they sell onwards for £100 or so [LB94]. This practice was made illegal in a recent amendment to the Data Protection Act, but there have still been no prosecutions of which we are aware.

The effects of aggregating data into large databases should have been expected. The likelihood that information will be improperly disclosed depends on two things: its value, and the number of people who have access to it. Aggregating records increases both these risk factors at the same time. It may also create a valuable resource which in turn brings political pressure for legalised access by interests claiming a need to know [Smu94].

Health systems are not likely to be different. At present, security depends on the fragmentation and scattering inherent in manual record systems, and these systems are already vulnerable to private detectives ringing up and pretending to be from another healthcare provider. A recent newspaper investigation showed that most people's records could be obtained for as little as £150 [RL95]. There are also some incidents specifically involving computer systems:

- following the theft of a general practice computer, two prominent ladies received blackmail letters threatening to publicise abortions;
- there is continuing abuse of prescription systems [JHC94];
- a Merseyside sex stalker who calls himself 'Dr Jackson' wins the confidence of young women by discussing their family medical history over the telephone and then tries to arrange meetings. Police believe that he is a health worker or a computer hacker [Tho95].

The interim guidelines issued at the same time as this policy give advice on how to make such attacks, on both manual and computer systems, less likely. However, the introduction of networking will change the risk profile, as current UK health networks are limited in scope, whether geographically or by function, and connecting them into a full-function national network will greatly increase the potential for mischief.

Put simply, we may not be much concerned that a GP's receptionist has access to the records of 2,000 patients; but we would be very concerned indeed if 32,000 GPs' receptionists all had access to the records of 56,000,000 patients. The danger of aggregating records, and the likelihood that abuse will result, is confirmed by the experience of the USA, where networking has advanced somewhat more than in Britain:

- a Harris poll on health information privacy showed that 80% of respondents were worried about medical privacy, and a quarter had personal experience of abuse [GTP93];
- forty percent of insurers disclose personal health information to lenders, employers or marketers without customer permission [CR94]; and over half of America's largest 500 companies admitted using health records to make hiring and other personnel decisions [Bru95];
- a banker on a state health commission had access to a list of all the patients in his state who had been diagnosed with cancer. He cross-referenced it with his client list and called in the patients' loans [HRM93];
- a US drug company has gained access to a database of prescriptions for 56 million people by purchasing a health systems company. It now plans to trawl the database for patients whose prescriptions suggest that they might be suffering from depression manifested as several other minor illnesses, such as backaches and sleeplessness, and try to get their doctors to prescribe them Prozac [See95];

- a credit reference agency is building a network to trade health records. It is sponsoring a bill in the US Congress which would facilitate disclosure to interested parties without patient consent, and remove patients' right to sue if unauthorised disclosure results in harm. This is an example of an information resource bringing political pressure for legitimised access, and is being contested by civil liberties' and patients' groups.

The problem was studied by the US government's Office of Technology Assessment. It confirmed that the main threats to privacy in computerised clinical record systems come from insiders rather than outsiders, and that they are exacerbated by the data aggregation which networked computer systems encourage [OTA93]. Other concomitants of data aggregation are growing claims of a need to know and treatment biased towards the interest of the corporate sponsor rather than the patient [Woo95].

The British government admits that wide access to identifiable clinical records has no ethical basis. Not even a clinician (let alone an administrator) may have access to personal health information in the absence of a need to know. In the words of David Bellamy, Principal Medical Officer at the Department of Health:

> It is a commonly held view ... that I as a doctor can discuss with another doctor anything about a patient because a doctor has a duty to maintain confidentiality by reason of his ethical obligations. It is just not true and it no longer holds water. Even if it helps professionals discussing individual patients with their colleagues, they must discuss only on the basis of the information the colleague needs to know [WHC95 p 16].

There are frequent claims by insurers, social workers, policemen and administrators that they have a 'need to know' personal health information. When evaluating such claims, it may be helpful to bear in mind that a surgeon's 'need to know' a patient's HIV status — so that he can take extra care to avoid needlestick injuries — is insufficient to override the patient's right to privacy about this status. A recent court case found that even a doctor's HIV status may not be disclosed: the small risk to patients' health does not outweigh the public interest in maintaining the confidentiality that enables infected persons to seek help [DGMW94].

The BMA does not accept that 'need-to-know' is an acceptable basis for access control decisions. As the EU and GMC documents make clear, it is patient consent that matters. The concept of 'need-to-know' implies and encourages the surreptitious erosion of the patient's privilege for the sake of administrative convenience. In any case, needs do not confer rights: the police's need to know whether a suspect is telling the truth does not give them a right to torture him. It is also useful to bear in mind empirical surveys of patient attitudes that show strong resistance to the sharing of personal health information with NHS administrators, social workers and government statisticians [Haw95].

2.4 Other security threats to clinical information

In addition to the threats to the confidentiality of clinical information, its integrity and availability may also be at risk in computer systems, and often in ways which are not immediately obvious.

- Software bugs and hardware failures occasionally corrupt messages. While mail, fax and telephone systems also fail, their failure modes are more evident than those of computer messaging systems. It is possible, for example, that a software bug could alter the numbers in a laboratory report without changing it so grossly that it would be rejected.

  There are regular press stories of mislaid cervical smear results and of pregnancies terminated in the mistaken belief that the foetus had Down's syndrome. We do not know how many of these involve computer as opposed to manual errors, but experience in other sectors suggests that in the absence of strong integrity controls about one message in 10,000 would be wrong. To a GP, this might mean a wrong test result every few years and a dangerous treatment once in a career. With poorly designed software, the figure could be substantially higher.

- Higher error rates could result from the spreading practice of sending lab results as unstructured electronic mail (email) messages that are sometimes interpreted automatically. A scenario from [Mar95] is plausible: a laboratory technician adds a comment before a numeric result, but the GP's system assumes that the first value it encounters is the result and files this in the patient record, leading to incorrect treatment.

- Viruses have already destroyed clinical information, and a virus could conceivably be written to make malicious alterations to records.

- A malicious attacker might also manipulate messages. Sending email which appears to come from someone else is easy, and with some more effort it is possible to intercept mail between two users and modify it.

- However the majority of malicious attacks will be carried out by insiders [OTA93], with motives such as erasing a record of malpractice [Ald95], supplying an addiction, or committing straightforward theft or fraud. Prescription fraud already happens with manual systems, and in the absence of improved controls it can be expected to continue.

- Attacks on system integrity could be made more likely by an erosion of confidentiality. If clinical records became widely available and were used for purposes such as hiring and credit decisions (as in the USA [Woo95]), then there would be strong motives to alter them.

- An erosion of public trust would also degrade the quality of input, as some patients would suppress sensitive facts. Public concern in America has now reached such a level that a national newspaper has warned its readers to be careful about disclosing sensitive health information [USA95].

- We might see similar effects if some system components have or acquire purposes other than healthcare. For example, if a health card came to be used as an identity card [DPR95], then both criminals and civil libertarians might try to break its security, and patients would assume that the police had access regardless of any assurances from government.

For all these reasons, the confidentiality and integrity properties of clinical systems should not be considered in isolation from each other.

2.5 Protection priorities

A common mistake in computer security is to focus on 'glamorous' but low probability threats such as the possibility that a foreign intelligence service might use eavesdropping equipment to decode the stray electromagnetic radiation from computer monitors. Although such attacks are possible, they may in practice be disregarded, as a capable motivated opponent would find cheaper and more reliable ways of accessing information (e.g., burglary or bribery).

Another example is the publicity given to occasional hacking attacks on the Internet. It is true that capable attackers can manipulate traffic in various ways and may succeed in logging on to systems by password sniffing and address spoofing techniques. However the incidence of such attacks is low, and competent Internet service providers will provide a firewall to make them hard. A much greater risk is that the computer system will be physically stolen from the surgery; over 10% of general practitioners have experienced computer theft [PK95].

We must therefore draw a distinction between vulnerabilities (things that could go wrong) and threats (things that are likely to go wrong). Note that other writers use these two words with their meanings reversed. However, such disputes are peripheral to our present concerns.

Threats vary in their scope, which we will take to be the number of individuals affected. There are global threats to the privacy, integrity or availability of the personal health information of the whole population, such as the black market in personal health information that already exists; while most threats are local and affect the privacy, integrity or availability of the clinical records kept by a care team. Examples are equipment theft, fire, virus infestations and the disclosure of records to third parties by careless staff.

Local threats can be contained by more or less well understood techniques, such as staff training, offsite backup and regular independent audit; most of the security effort of a general practice or hospital department will be devoted to them. General guidelines have been issued by the Department of Health [NHS95] while the BMA has issued its own guidelines And96] on action that should be taken to counter the most serious threats of which we are aware at this time.

Meanwhile, at the policy level, our priority is to ensure that local attacks do not develop into global ones, or exacerbate existing global threats, by the ill-considered aggregation of data, or by neglecting the principle of consent. The security policy principles that we wish to be enforced by all communicating clinical systems must prioritise issues such as aggregation and consent.

2.6 Examples of aggregation in NHS systems

The aggregation of personal health information may come about in a variety of ways, some doubtless well meaning and others driven by nonclinical pressures. Examples in current and proposed NHS systems include:

- the proposed NHS Clearing Service for in-patient contract data will contain information on hospital treatment of patients throughout the country. Requests by the BMA to review the functional specification of this system have been dismissed with the assertion that this information is not in the public domain;
- the Administrative Registers contain sensitive information such as past registration for contraceptive services and relationships with mental health institutions;
- at least two systems have been developed that enable health authorities to link up item-of-service claims, prescriptions and contract data to create a 'shadow' patient record outside clinical control [AIS95] [DL95];

The above systems have been commissioned despite agreement between the NHS Executive and the clinical unions that electronic patient records shall be at least as secure as paper records, and established guidelines of the GMSC/RCGP Joint Computer Group which state that no patient should be identifiable, other than to the general practitioner, from any data sent to an external organisation without the informed consent of the patient [JCG88].

A strategic goal of the NHSE's Information Management Group is an entirely shared electronic patient record; we understand that the collection of GP data is to be the driving force, and that GP systems will be interrogated by NHS systems. However these goals are in clear conflict with the ethical position of the BMA [Som93] as well as the Joint Computer Group guidelines mentioned above.

Patient consent for the sharing of personal health information with NHS administrators is not present; indeed, a survey shows that most patients are unwilling to share personal health information with them [Haw95]. That this information should be collected into large aggregates that are outside the control even of healthcare professionals is extremely dangerous; as the US experience has shown, the mere existence of such a potentially valuable resource will create strong political pressures for legitimised access by law enforcement agencies, insurance companies and others.

The response of the BMA includes this document. Its primary purpose is to help clinical professionals discharge their ethical and legal responsibilities by selecting suitable systems and operating them safely. It seeks to define what kind of systems may prudently be trusted to receive personal health information, and for that we shall build on the threat model developed in this section to develop a security policy for clinical information systems. This consists of a compact set of principles that if implemented properly will enforce patient consent effectively in communicating computer systems.

3 Security Policy

The principle of consent and the rules used to interpret it are well entrenched — they have evolved over centuries of clinical experience, and are supported by data protection law. In this section, we express them in the form of a security policy — a set of principles governing which subject can access which object in a computer system. They contain nothing that is radically new, but rather restate commonsense principles in the modern language of computer security.

The policy covers clinical systems in general. Some clinicians will have extra requirements, and those that treat more than one identifiable patient at a time (such as pediatric psychiatrists, embryologists and human genome researchers) face particularly subtle dangers. For example, the access rights enjoyed by data subjects might enable one subject to discover information about another; there are also special legal requirements in many cases. Designers of systems supporting such activities should seek further advice.

A note on record structure

There are basically two ways to organise electronic clinical records. The first mirrors the existing paper based system; each clinician keeps a record in her own computer (or manual filing system), and information passes between them in the form of summaries (such as referral and discharge letters). The second assumes that each patient will have a single electronic file which will be opened before birth, closed on autopsy, and contain everything of clinical interest in between.

In what follows, we shall start off by assuming the first paradigm, as it is prevalent in the actual practice of clinical medicine and is much simpler to deal with. Once we have developed a security policy for this case, we will discuss the other approach, which has been called 'patient-based records' but in reality may mean keeping records in some central registry. We will finally look at compromise approaches such as keeping the detailed records in clinicians' systems but compiling a central summary with pointers to them.

3.1 Access control

In a computer system, each subject has access to certain objects. This access information may be stored by subject or by object. In the former case, the access permissions are called capabilities, and might have the form 'Dr Jones may read the records of Farid Abdullahi, James Adams, Wendy Adams, Henry Addenbrooke, ... ' If the permissions are stored by objects, they are called access control lists, and might have the form 'This is Farid Abdullahi's record and it can be read by Dr Jones, Dr Smith and Nurse Young'. The latter approach leads to simpler engineering, as the number of patients per doctor is much larger than the number of doctors per patient.

In the normal course of events, any clinician with access to a record may not only read it, but also add information to it (we will deal with the deletion of information later). Our first principle is therefore

> **Principle 1:** 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.

In many current systems, the access control lists are implicit. If a record is present on the practice database, then all the doctors in that practice may read it and append things to it. However, with the introduction of networking, access control lists need to be made explicit and consistent across a range of systems, and must be enforced by mechanisms that are not only technically effective, but that support practices such as deputising and caseload sharing.

To facilitate this, groups may be used instead of individual names. For example, if Dr Jones, Dr Smith and Nurse Young together staff the Swaffham practice, then the records to which they all have access might simply be marked 'Swaffham'. This idea was inherent in the development of Community Care; the teams involved doctors, nurses and social services staff, and written consent was obtained at the start of the assessment for information to be shared. In this way, the patients knew whom they were signing up to trusting.

However, sometimes the only sensible groups include a large number of people. In large hospitals and community health trusts, there might be hundreds of nurses who could be assigned to duty in a particular ward or service. Some extra restrictions may then be needed in defining groups; for example, the group might be 'any clinical staff on duty in the same ward as the patient'. Such an approach would be the electronic equivalent of a traditional note trolley, but with the added advantage that a record can be kept of who consulted what.

Whenever groups are used — whether simple groups including a few clinicians, or complex ones with location and other constraints — a record must always be kept of which individual read a record or added anything to it. We will discuss attribution more fully below; here, we merely emphasise that groups are not virtual clinicians, but mechanisms that simplify the access mapping between identified clinicians and identified patients. Designers should bear in mind that a given system user may belong to many different groups: she might simultaneously be a patient, a doctor, a trainer, a trainee, a practice manager and a consultant to a health authority. Unless provision is made to manage this complexity, it is unlikely to be managed well; ad hoc methods should be avoided.

It is not acceptable, for example, for a group to be implemented by a password shared by all the staff on a ward, or by leaving a terminal permanently logged on to the consultant's account. Such abuses mean that actions could no longer be attributed to individuals, and they can cause serious harm: we are aware of a case where a psychiatric patient used a ward terminal to alter prescription data with murderous intent.

When a patient registers with a practice or otherwise commences a clinical relationship with a care team, and a record is opened for him, he should be given information on the team's access control policy. He must also be given the opportunity to object and request that his record be restricted to one or more named clinicians. For this reason, role-based systems must still support more restricted access control lists, and in particular lists containing a single named clinician (plus of course the patient).

Such a list may even be the default in the case of highly sensitive data. The actual sensitivity of a record is a decision for the patient or patients concerned. Examples of data which are prima facie highly sensitive include psychiatric records, records of sexually transmitted disease and all information given by or about third parties (for a fuller list see [GC95] p 44). However, an AIDS campaigner might make his HIV status public, while a Jehovah's witness might consider even a blood transfusion to be deeply shameful. So the patient's consent remains paramount, and no-one may be added to the access control list without his being notified. We will discuss notification in greater detail below.

Finally, there are some users, such as auditors and researchers, who have no write access at all to the primary record. We will discuss their special problems below, but for simplicity's sake we will not make separate provisions for read-only access in this policy. We will rather assume that they get full access to a temporary copy of the primary record; and this is in fact a better model of how they actually work.

3.2 Record opening

Rather than trying to deal with objects having multiple access control lists, we will assume that there are multiple records. A patient might for example have:

- a general record open to all the clinicians in the practice;
- a highly sensitive record of a treatment for depression which is only open to his GP;
- a record of heart disease open to all casualty staff, a summary of which might be carried on an emergency medical card.

This is logically equivalent to having a record with three different fields each with its own access control list. However is much simpler for us to deal with.

So the clinician may open a new record when an existing patient wishes to discuss something highly sensitive, or when a new patient registers with her, or when a patient is referred from elsewhere. The access control list on a new record is as follows:

> **Principle 2:** 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.

3.3 Control

Apart from the patient himself, only clinicians may have access to personal health information. The reasons for placing the trust perimeter at the professional boundary are both traditional and practical: the clinical professions do not consider the mechanisms of the civil and criminal law to give adequate protection. If a doctor gave a record to a social worker who then passed it to a third party without consent — or merely kept it in an insecure local council computer system which was hacked — then the doctor could still be liable, and might have no recourse.

In effect, only clinicians are trusted to enforce the principle of informed consent, and control of any identifiable clinical record must lie with the individual clinician who is responsible. This might be a patient's GP, or the consultant in charge of a hospital department.

> **Principle 3:** 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.

Where access has been granted to administrators, as in the USA, the result has been abuse. In the UK, the tension between clinical confidentiality and administrative 'need-to-know' has been assuaged by regulations that purchasing organisations must have 'safe-havens' — protected spaces under the control of an independent clinician — to which copies of records may be sent if there is an administrative dispute [NHS92]. Administrative systems that might handle personal health information must support safe-haven procedures; for example, the clinical parts of patient records might be encrypted in such a way that only the clinician in charge of the safe-haven could decrypt them. Such systems must also abide by the Joint Computer Group guidelines mentioned above [JCG88].

When information is sought by, and may lawfully be provided to, a third party such as a social worker, a lawyer, a police or security service officer, an insurance company or an employer, then the information must be provided on paper. This reflects current practice: in the community care scenario mentioned above, records shared between doctors, nurses and social workers were kept on paper rather than on a database because of security concerns.

It should also be borne in mind that computer records are not usable as evidence unless they come with a paper certificate signed by the system owner or operator; direct electronic access is of little evidential value, and a signed statement on paper can best satisfy a bona fide requirement for evidence.

3.4 Consent and notification

The patient's consent must be sought for other clinicians to be added to the access control list, and he must be notified of every addition. In the normal course of business, a poster or box of leaflets displayed prominently in the surgery or hospital reception may discharge this requirement in respect of the clinician's immediate colleagues, so long as there are effective ways to cope with the few patients who will insist that their records be available only to the treating clinician. Adding other clinicians to the access control list, such as when a patient is referred to hospital, should normally be discussed with the patient beforehand.

However, when information is shared in the absence of consent, such as when a GP shares information with a casualty department under emergency procedures, then a notice must be generated and sent to the patient. This is the GP's responsibility; if she merely assumes that the hospital would notify the patient, then she would be seriously negligent. Illegal information brokers often obtain personal health information by pretending to be involved in emergency treatment of patients; detailed guidance on the design of emergency procedures is in [And96], which lays emphasis on the need to establish the identity of the caller (such as by calling back to a number in the Medical Register), and to always notify the patient.

Notification provides an end-to-end audit that is not vulnerable to management capture of auditors or regulators. For example, a hospital employee might be bribed by an illegal information broker to request access to a patient's record from a general practice by falsely claiming that the patient had been admitted unconscious. The callback control would not be effective in this case, but notifying the patient ensures that the attack can be detected and investigated.

The notification requirement thus flows from the principle of consent. It also helps control fraud in private practice, as benefits may be cash limited and patients with expensive treatment needs may impersonate other patients when their budget runs out.

There are no exceptions to it. Even where a clinical professional is under a legal duty to pass some information to a third party, the patient must still be notified. In the event of law enforcement access or the discussion of suspected child abuse with social services, the notification may be delayed if there are reasonable grounds for belief that it would cause the suspect to flee, tamper with evidence or intimidate witnesses. However the patient must still eventually be notified.

> **Principle 4:** 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.

There is also the question of how often to notify. The feeling among clinicians consulted is that notification should be annually by letter, unless a violation or a suspicious pattern of activity has been detected. However, it is not quite straightforward. Recently, GPs were asked to notify women using certain contraceptives; this raised issues of how to deal with young girls who were taking contraceptives without their parents' knowledge, and women whose spouses had had a vasectomy and were taking the pill in a new extramarital relationship [Gil95]. The solution, which is already practised in STD clinics, is for the clinician to ask the patient at the outset of the relationship how notices should be sent.

A more difficult problem arises when the patient-clinician relationship ceases to exist. This may happen when a private practice is dissolved, or a patient dies or goes abroad. Concerns have been raised about the OPCS garnering emigration data from records returned by GPs to FHSAs for storage under current arrangements; it has been suggested that the Data Protection Registrar have custody of all 'dead' electronic records. However this raises the question of who would watch the watchman.

Finally, there needs to be an effective complaints procedure which results in offenders being punished, whether by dismissal, by professional disciplinary action, or by criminal prosecution. When a patient observes from his annual notification letter that someone he never consulted has read his record, what should he do? Should he go to his GP in the first instance, or take the matter up with the General Medical Council, some kind of ombudsman, the Data Protection Registrar, his MP, the press, or even the police? A resolution of this may depend on the success of the BMA's campaign for a bill to enshrine the confidentiality of personal health information in statute [BMA95].

3.5 Persistence

There are rules on how long records must be kept. Most primary records must be kept for eight years, but cancer records must be kept for the patient's lifetime, and records of genetic diseases may be kept even longer. In any case, prudence dictates maintaining access to records until after a lawsuit for malpractice could possibly be brought. So our next principle is:

> **Principle 5:** No-one shall have the ability to delete clinical information until the appropriate time period has expired.

However, these rules are still not fully worked out, and so our use of the word 'appropriate' covers a number of outstanding issues:

- our formulation allows the destruction of old records, but does not mandate it; there are many cases (such as chronic illness) in which it is appropriate to keep records for longer than the law requires;
- the sixth principle of the Data Protection Act [DPA84] states that personal information 'shall not be held for longer than is necessary'. This may mean that once a clinician is no longer the primary record holder (e.g., if the patient has moved) then the record should be destroyed. However, before doing this, she may wish some assurance that it can be made available if necessary (e.g., in the event of a lawsuit);

- patient consent is not immutable, but rather a continuing dialogue between the patient and the clinician [Som93]. It is therefore quite possible that a patient might withdraw consent and insist that a record be destroyed. No case has come to our attention yet; perhaps such cases might be dealt with by transferring the primary record to a clinician of the patient's choice for the rest of the statutory period;
- with temporary copies of records, the appropriate time period will be shorter. For example, where a general practice grants access to a night-time deputising service, it is typically a condition that all copies of records be deleted within a set period of time. Similar considerations apply to copies of records held by a safehaven, an auditor or a researcher; for example, consent to record sharing for research should be renewed every five years [Som93], so copies of records made by researchers should persist no longer than that (and should normally be destroyed much sooner). The design and enforcement of such volatility requirements has an impact on aggregation control, which is discussed below.

Preserving records is not completely straightforward; we do not want information that has been identified as inaccurate, such as simple errors and subsequently revised diagnoses, to be mistakenly acted on. However, we do not want to facilitate the traceless erasure of mistakes, as this would destroy the record's evidential value. So (as with many financial systems) information should be updated by appending rather than by deleting, and the most recent versions brought first to the clinician's attention. Deletion should be reserved for records that are time expired.

An equivalent expression of the above principle may be found in the current requirements for accreditation of GP systems which state that 'the system must not allow records ... to be altered or deleted unless a secure mechanism is provided to reconstruct these records as they were on any specified day in the past' [RFA93].

3.6 Attribution

We next must ensure that all record accesses (whether reads, appends or deletions) are correctly attributable.

> **Principle 6:** 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.

Systems developed under the present requirements for accreditation will typically record all write accesses; even if material is removed from the main record, there is an audit trail which enables the state of the record as it was at any time to be reconstructed and all changes to be attributed [RFA93]. If implemented properly, this will have an equivalent effect to restricting write access to append-only and marking all append operations with the clinician's name. The new requirements are that read accesses be logged, so that breaches of confidence can be traced and punished; and that deletions be logged so that the deliberate destruction of incriminating material can be attributed.

Some applications have particularly stringent attribution requirements. For example, a 'Do-Not-Resuscitate' notice on the record of a patient in hospital must be signed by the consultant in charge, and also requires consent if the patient is competent to give it [Som93]. When such life critical functions are automated, the mechanisms — including those for supporting attribution — must be engineered with the same care and to the same standards that are expected in life support systems.

There are also attribution requirements that are rarely invoked. For example, with only a few exceptions, patients have read access to all their records and may append objections if they have any. These requests are rare, and so they are typically supported with manual mechanisms. A common procedure is for the clinician to print out any records to which access is requested, and in the event of objections to enter the patient's comment and hand him a copy of the updated record for confirmation. We have no objection to these procedures. We do not insist that security be all in software; we are concerned with the net effect of all processing, both automated and manual.

3.7 Information flow

Where two records with different access control lists correspond to the same patient, then the only information flow permissible without further consent is from the less to the more sensitive record:

> **Principle 7:** 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.

The technical mechanisms needed to in enforce such a principle are described in standard computer security texts such as Amoroso [Amo94]: a process's access control list should be set to the intersection of the access control lists of the records it has read, and it should only be able to write to a record whose access control list is included in its own.

Where two records with different access control lists correspond to the same patient, the hard question is whether the existence of the sensitive record will be flagged in the other one. This is one of the continuing dilemmas on which there is no consensus yet [GC95]. If the existence of hidden information is flagged, whether explicitly or by the conspicuous absence of parts of the record, then inferences can be drawn. For example, doctors in the Netherlands removed health records from computer systems whenever the patient was diagnosed with cancer. The result was that whenever insurers and pension funds saw a blank record, they knew that with high probability the subject was a cancer sufferer [Cae95]. Visible flags have also led to a UK case that is currently subjudice.

In the absence of flags, other problems arise. Suppose for example that a psychiatric outpatient goes for an AIDS test and requests that the result be kept secret. Before the result is known, the stress causes a breakdown and his psychiatrist marks him as no longer competent to see his records. However, the psychiatrist is unaware of the test and so does not tell the STD clinic of the patient's new status. It is not possible to solve this problem by having a world readable register of which patients are currently not competent, as mental incapacity is both confidential and a function of circumstance. Another consequence of not flagging hidden data is that sufferers from Munchhausen's syndrome could be harder to detect and manage.

We expect that clinicians will decide in favour of discrete flags that indicate only the presence of hidden information. These will prompt the clinician to ask 'is there anything else which you could tell me that might be relevant?' once some trust has been established.

In any case, system developers should give careful consideration to the propagation of sensitivity properties through dependent records, and to the effects of this on system integrity.

Finally, there needs to be a mechanism for dealing with the release of data that have been made anonymous. As with the downgrading of information in multilevel systems, we will not incorporate this within the security policy model itself. We recommend however that releasing a record believed to be anonymous should require a deliberate act by the responsible clinician and should be logged.

3.8 Aggregation control

The use of access control lists and strong notification are helpful against aggregation threats but are not quite enough to prevent them. The clinician in charge of a safe-haven might be added to the access control lists of millions of hospital patients, making her vulnerable to inducements or threats from illegal information brokers.

> **Principle 8:** 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.

Some hospitals' systems contain personal health information on a million or more patients, with all users having access. The typical control at present is a declaration that unjustified access will result in dismissal; but enforcement is often sporadic, and incidents such as the Jackson case continue to be reported. In general, hospital systems generally tend to be old and poorly administered [AC95a] [AC95b].

Hospital systems which give all clinicians access to all data should not be connected to networks. Having 2,000 staff accessing a million records is bad enough; but the prospect of 200 such hospitals connected together, giving 400,000 staff access to the hospital records of most of the population, is unacceptable.

However, there will inevitably be mechanisms for clinicians to access records from outside their own care team, even if these are manual ones. These mechanisms need careful design. As noted above, a corrupt member of staff might falsely claim that a patient has self-referred while on holiday, and ask for a copy of the record to be sent. Even a simple electronic mail system could enable such enquiries to be repeated on an industrial scale.

The primary control on such threats is notification. However an important secondary control is to keep a count somewhere of who has accessed what record outside their own team. Users who access many records, or a number of records outside the usual pattern, may just be lazy or careless, but they could still be exposing themselves and their colleagues' patients to harm.

Given the tension between clinicians and administrators on privacy issues, both the location of this count and the choice of the persons responsible for acting on it should be chosen carefully: it might for example involve the clinical disciplinary bodies or healthcare unions. It would also make sense to deal with reports of other computer abuse at the same place. The involvement of the clinical unions may help prevent the central security function being captured by bureaucratic interests and thus preserve the principle of consent.

There are applications in which some aggregation may be unavoidable, such as childhood immunisation programmes. Systems to support them will have to be designed intelligently.

As mentioned above, records may be aggregated for research and audit purposes provided that they are made sufficiently anonymous. It has been suggested that records can be made anonymous by replacing names with NHS numbers and diagnoses with Read codes [RSM92], and a number of systems appear to have been specified on the assumption that this is acceptable. It is not; as noted above, the existing GMSC/RCGP guidelines stipulate that no patient should be identifiable, other than to the general practitioner, from any data sent to an external organisation without the informed consent of the patient [JCG88]

Making data anonymous is hard, especially if it contains linkable information: if an attacker can submit database queries such as 'show me the records of all females aged 35 with two daughters aged 13 and 15 both of whom suffer from eczema', then he can identify individuals. The limits of linkage, and techniques for preventing inference, are known as 'statistical security' and have have been researched in detail in the context of census information [Den82]. Where purely statistical research is proposed, then these techniques may be used; where they are impractical, researchers might be granted access to linkable data within protected space [Boe93].

3.9 The Trusted Computing Base

Finally, we must ensure that the security mechanisms are effective in practice as well as in theory. This leads to issues of evaluation and accreditation.

In computer security terminology, the 'trusted computing base' is the set of all hardware, software and procedural components that enforce the security policy. This means that in order to break security, an attacker must subvert one or more of them.

At this point we will clarify what we mean by 'trust'. In the commonplace use of language, when we say that we trust someone we mean that we rely on that person to do — or not to do — certain things. For example, a patient when sharing confidential information with a clinician expects that this information will not be shared with third parties without his consent and relies on this expectation being fulfilled.

A way of looking at such relationships, that has been found to be valuable in system design, is that a trusted component is one which can break security. Thus a clinician who has obtained confidential information from a patient is now in a position to harm him by revealing it, and he depends on her not to. There will be parts of any computer system on which we similarly depend. If they are subverted, or contain bugs, then the security policy can be circumvented.

The trusted computing base of a clinical information system may include computer security mechanisms to enforce user authentication and access control, communications security mechanisms to restrict access to information in transit across a network, statistical security mechanisms to ensure that records used in research and audit do not possess sufficient residual information for patients to be identified, and availability mechanisms such as backup procedures to ensure that records are not deleted by fire or theft.

The detailed design of these mechanisms is discussed in the next section. For now, we will remark that it is not sufficient to rely on the assurances of equipment salesmen that their products are 'secure' — these claims must be checked by a competent third party.

> **Principle 9:** 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.

The need for independent evaluation is shown by long experience, and there is now a European scheme, ITSEC [EU91], under which national computer security agencies (in Britain's case CESG/GCHQ) license commercial laboratories to carry out security evaluations. Independent evaluation is also a requirement in other countries such as Australia [Aus95], Canada [TCP93] and the USA [TCS85]. As schemes such as ITSEC are oriented towards military systems and evaluations under them may be expensive, some industries run their own approved schemes. For example, the security of burglar alarm signaling is evaluated by the underwriters' laboratories of the Loss Prevention Council. Similar industry-wide arrangements may in due course be made for clinical systems.

3.10 Clinical records or patient records?

As noted above, most clinical information systems mirror clinical practice in that each care team has a record keeping system, and information flows between them in the form of summaries (referral letters, discharge letters, opinions, test results and so on). The whole record may be copied to another team if the patient is transferred, but otherwise the records are clinician-based rather than patient-based, and only summary information flows between them.

As mentioned above, there has been interest recently in a different model, the 'unified electronic patient record', which accumulates all the clinical notes and data in a patient's lifetime [MRI94]. But securing a unified record is complicated, for a number of reasons:

- if the records are held by the patient on an optical card or diskette, then how will we recover from lost records? But if the records (or backups) are held on a central database, then how would aggregation be controlled?
- birth records contain the mother's personal health information as well. Surely the patient will not obtain unrestricted access to them?
- how would one deal with large files such as CAT scans and the records of long chronic illnesses?
- how would clinicians be guaranteed access to former patients' records to evaluate the care they gave and to defend themselves from lawsuits?
- suppose that I walk into a hospital and claim that my demons are particularly troublesome. When asked my name I reply 'John Major'. May the psychiatrist access the prime minister's record and append a diagnosis of schizophrenia? In other words, does a patient-based record force us to authenticate patients much more carefully, and if so, what are the implications for emergency care, for patients who wish to be treated anonymously (such as fourteen year old girls seeking post-coital contraception), and indeed for civil liberties?
- if a patient receives treatment in prison, then this fact may not be recorded elsewhere once his conviction has expired under the applicable rehabilitation rules. So prison records cannot realistically be held elsewhere, and neither can highly sensitive records restricted to a single clinician. What then is the gain of a centralised system if local records must still exist?
- a lifetime record would promote data retention because of the accretion of links between episodes, and make sensitive records (or markers indicating their absence) visible to the hundreds of health care staff who would get

access at some time in the patient's life. How could these vulnerabilities be controlled without expensive manual editing?

The above list is by no means exhaustive. For a discussion of the security complexities of patient-based record systems, see Griew and Currell [GC95]. As their paper makes clear, the use of unified electronic patient records would force us to add quite a few principles to our list.

There are also trials with hybrid systems. Rather than putting all a patient's health information in a single file, one might have a central summary containing pointers to detailed files kept in clinicians' systems. There are currently at least two UK hospitals doing trials of systems based on this model, both of which apparently allow all users to access all records; but even with proper access control, one might ask what is wrong with the traditional GP record. Although 'doctor-based', it is the closest we have to a lifelong patient record.

In any case, the onus is on proposers of 'patient-based' record systems to provide a clear statement of the expected health gains and analyse the threats, the cost of added countermeasures and the likely effects of the residual risk.

4 Security Architecture Options

The security policy set out in the section above applies to systems in general. Our goal was not to encumber it with the details of specific equipment, but to produce a policy that is just as capable of implementation on a mainframe with a number of terminals as it is on a heterogeneous distributed system consisting of a number of systems linked together by communications protocols — or even for that matter using rooms full of clerks with quill pens.

However the case of heterogeneous distributed systems is the main one of interest in the UK, and in this section, we consider some technical options for implementing it. This section is indicative rather than normative; it is up to individual equipment suppliers to design their own systems and have them evaluated for compliance with the security policy. Everything required by the policy can be achieved with well understood technology. However the following notes may be helpful, especially to vendors who are not familiar with modern computer security techniques.

4.1 Compusec

Compusec, or computer security, measures include the access control mechanisms built into operating system and applications software. They typically comprise an authentication mechanism such as passwords, an access control mechanism which decides which subject can access which object, and an audit trail which tells who did what. A standard textbook on compusec is Amoroso [Amo94].

Our policy principle describe the functional requirements of the access control mechanism in some detail. As for the authentication mechanism, the strength we require will depend on whether outside access is possible. With a network that is completely within protected space, passwords may suffice. However, if a system supports dial access or Internet access, then it may need the more complex controls discussed in the next section.

This leads to the more general problem of where the access controls are located in the system. It is possible, but expensive, to implement them in each application program; it will usually be cheaper at a lower level in the system. Access control lists are supported by many operating systems, such as Unix, whose group and individual permissions may be used to make records accessible to all team members and to individuals respectively. If a database management system is used, then access controls at the granularity of individual patient records may have to be implemented in the database. In a heterogeneous distributed system that used cryptography as its primary control, then the access control might be largely embedded in the key management mechanism.

The automatic enforcement of principle 7 is very important. When a program derives data from an identifiable clinical record, then the derivative data shall have the same access control list as the original data, or a subset of it. A summary of a record is just as sensitive as the original. One of the benefits of this mechanism is to help prevent accidental as well as deliberate security breaches. For example, it is quite common to post personal messages to a mailing list or newsgroup by mistake. The system should prevent a clinician leaking personal health information in this way.

Finally, where records are made anonymous for audit or research purposes, it is the responsibility of the clinician to ensure that the anonymising process is effective in the context, and for this reason it should take a deliberate action of the clinician to release the data. As the Joint Computer Group guidance makes clear, it is not acceptable for records to be sent to a health authority or drug company on the promise that they will be made anonymous once there.

4.2 Comsec

The main purpose of comsec, or communications security, measures is to ensure that access controls are not circumvented when a record is transmitted from one computer to another. This might happen, for example, if clear data are transmitted to a system which corrupts its access control list, or which does not enforce the principle of informed consent. It might also happen if clear data were intercepted by wiretapping, or if clinical information in an electronic mail message were sent by mistake to a mailing list or newsgroup.

The secondary purpose of comsec mechanisms is to protect the integrity of data in transit through a network. Some messages, such as pathology reports, are life critical; and there is also controversy on whether clear electronic records are adequate for legal purposes. It is therefore desirable in many applications to add an integrity check to messages.

Clinicians should not assume that a network can be trusted, unless it is under their direct control and enclosed in protected space, as may be the case for a local area network joining computers in a surgery. Wide area networks such as the Internet and the NHS wide network may not be trusted. Remember that for a network to be trusted is equivalent to saying that it can break system security. To expose patient confidences to a system component which is not under clinical control, or under the effective control of a trustworthy third party, is imprudent to the point of being unethical.

A convenient means of protecting information in a network is provided by cryptography. Modern cryptographic systems allow users to have separate keys for encryption and decryption, and the encryption key can be published while the decryption key is kept secret. Similarly, a user will have separate keys for signature and signature verification; the signature key will be kept secret while the signature verification key is published so that anyone may verify a signed message. A standard textbook on cryptography is Schneier [Sch95].

Digital signatures allow the creation of trust structures. For example, the General Medical Council might certify all doctors by signing their keys, and other clinical professionals could be similarly certified by their own regulatory bodies. This is the approach favoured by the government of France [AD94]. An alternative would be to build a web of trust from the ground up by users signing each others' keys. A half-way house between these two approaches might involve key certification by a senior clinician in each natural community.

All of these options possess strengths and weaknesses, and are the subject of current discussion. The centralisers' strongest argument appears to be that even if certification were substantially local, one would still need a central service for cross-domain traffic. They may also argue that this central service should be computerised, since if one merely had a key fingerprint next to each clinician's name in the appropriate professional register, it would not enable clinicians to verify signatures on enclosed objects.

However, a single certification authority would be a single point of failure, and electronic trust structures should also reflect the actual nature of trust and authority in the application area [Ros95]. In medicine, authority is hierarchical, but tends to be local and collegiate rather than centralised and bureaucratic. If this reality is not respected, then the management and security domains could get out of kilter, and one could end up with a security system which clinicians considered to be a central imposition rather than something trustworthy under professional ownership and control.

Most published key management and certification standards relate to banking, but clinical systems have additional requirements; one might for example want a count of the total number of patients' records accessed by the clinician outside her team during a certain period of time, and this might well be enforced through the certification mechanism.

In any case, once each clinician has acquired suitably certified key material, the integrity of access control lists and other information on a network can be enforced by means of a set of rules such as:

1. personal health information may not leave a clinical system unless it is encrypted with a key which is reasonably believed to belong to a clinician on its access control list;
2. life critical information that has been transmitted across a network should be treated with caution unless it has been signed using a key which is reasonably believed to belong to an appropriate clinician;
3. reasonable belief in the above contexts means that ownership of the key has been authenticated by personal contact, by introduction, or by other trustworthy means;
4. decrypted information must be stored in a trusted system with an access control list containing only the names of the patient, the clinician whose key decrypted it, and the clinicians (if any) who signed it.

Careful consideration must be given to the circumstances in which acts of decryption and signature may be carried out. If the system can execute a signature without the signer's presence, then it may have no force in law [Wri91]. This ties in with the principle that when working cross-domain, records must be given rather than snatched; access requests should never be granted automatically but need a deliberate action by a clinician.

Comsec techniques may be applicable in more restricted applications. The guidelines issued with this document cover prudent practice for dialback protection of links to branch surgeries. Another example might be where a clinician wished to use a portable computer with a mobile telephone to view records while on night visits. Some mobile phones (particularly those using GSM) provide a level of security which may be acceptable, while others are easy to monitor. If an insecure medium is used then it would be prudent to protect the data by application level mechanisms such as encryption.

Encryption and dialback are not the only comsec options. Another is to make data anonymous, in those applications where this is straightforward. For example, a system for delivering laboratory reports to GPs might replace the patient's name with a one-time serial number containing enough redundancy to make accidental error unlikely. The test results might then be transmitted in clear (with suitable integrity checks).

The most important factor in getting security that works is not so much the choice of mechanisms but the care taken to ensure that they work well together to control the actual threats.

4.3 Evaluation and accreditation

The trusted computing base is the sum total of all hardware, software and procedural components which, singly or in combination, could break the security policy. Its design is a matter for the system supplier, but experience shows that the smaller it is, the better. Small security systems are cheaper to evaluate, and reduce the likelihood of bugs that compromise security.

Procedural mechanisms such as password administration, configuration management and backup are an integral part, and when assessing a system the evaluator must ask whether it is likely to be operated securely by a clinician whose computer skills and administrative tidyness are less than average. Lazy and careless clinicians exist, so if it is more convenient to run the system insecurely, a positive evaluation may not be issued. Evaluators should also take into account human design issues such as the quality of manuals and training, and the use of integrity checks on manual data entry.

The level of evaluation should depend on the exposure. We suggest ITSEC level E2 for up to 50,000 patient records, and E4 for 50,000 — 1,000,000 patient records. Systems which contain personal health information on significantly more than 1,000,000 people should not be built.

Finally, when a system is being installed by a purchaser, the responsible clinicians must ensure that all relevant training has been completed and any necessary plans, procedures and materials — from a disaster recovery plan through informative leaflets to patient consent forms — have been drawn up and tested before patient identifiable clinical information is input to the system. The decision to expose the information in this way should be a conscious professional decision to accept the residual risk, and it should be noted in writing by the responsible clinicians. Only once this accreditation exercise has been completed should a system be furnished with the key material needed to communicate with other systems.

4.4 European and global standardisation

The policy and guidelines set out in this document are as far as the author is aware broadly consistent with European and other standards work. We understand that a European standardisation group for Security and Privacy of Medical Informatics (CEN TC 251/WG6) is working on a draft that mandates the encryption of personal health information in large networks; encryption has been required by the data protection authorities in Sweden for several years, and a number of countries are building trusted certification authorities which will sign healthcare professionals' keys [SPR95].

The use of digital signatures is also discussed in a report to the Ontario Ministry of Health [Smu94]. The Australian standard on health information privacy [Aus95], the Royal Australian College of General Practitioners Interim Code of Practice for Computerised Medical Records in General Practice [RAC+93], the New Zealand Health Information Privacy Code [NZ94], and the US Office of Technology Assessment report [OTA93] may also be referred to. They each contribute in different ways to our understanding of the threats, of the principle of consent, of the technical options, and of pragmatic standards of best practice in other countries.

Suppliers are also encouraged to adopt best European practice, which may be very important once European data protection law comes to be enforced in British courts. This will if anything increase the emphasis on patient consent.

5 Conclusions

We have described the threats to the confidentiality, integrity and availability of personal health information in the light of experience in the UK and overseas, and proposed a clinical information security policy that enables the principle of patient consent to be enforced in the kind of heterogeneous distributed system currently under construction in the UK.

Clinicians making purchasing decisions are encouraged to favour systems which have been evaluated for compliance with this policy. Where no evaluated system is yet available, purchasers should take into account the extent to which available products support the principles set out here, and whether the supplier will undertake to provide an upgrade path to an evaluated system.

Where none of the available products provides an acceptable level of computer and communications security, the advice of the British Medical Association to its members is that exposing unprotected patient identifiable clinical information to the NHS wide network (or indeed to any other insecure network), or even sending it in encrypted form to an untrustworthy system, is imprudent to the point of being unethical.

Acknowledgements: Valuable input was received during the preparation of this document from a number of healthcare professionals, including Fleur Fisher, Tony Griew, Simon Jenkins, Grant Kelly, Stuart Horner, Hilary Curtis, Simon Fradd, John Williams, Iain Anderson, William Anderson, Roger Sewell, Mary Hawking, Ian Purves, Paul Steventon, Steve Hajioff, Stan Shepherd, Jeremy Wright and David Watts; from a number of computer scientists including Stewart Lee, Roger Needham, Mark Lomas, Bruce Christianson, Ian Jackson, Mike Roe, Jeremy Thorp, Roy Dainty and Ian Keith; and from philosophers including Beverly Woodward, Ann Somerville and Keith Tayler.

References

[Ald95] "Nurse sacked for altering records after baby's death", K Alderson, *The Times* 29 November 95 p 6

[Amo94] *'Fundamentals of Computer Security Technology'*, E Amoroso, Prentice Hall 1994

[And96] "Medical System Security — Interim Guidelines", RJ Anderson, to appear in *British Medical Journal* 13th January 1996

[Aus95] *'Australian Standard 4400: Personal privacy protection in health care information systems'*, Standards Australia, 1995

[AC95a] *'Setting the Records Straight — A Study of Hospital Medical Records'*, Audit Commission,, June 1995

[AC95b] *'For Your Information — A Study of Information Management and Systems in the Acute Hospital'*, Audit Commission,, July 1995

[ACH95] *'Keeping Information Confidential'*, Association of Community Health Councils for England and Wales, May 1995

[AD94] "Security of Health Information Systems in France: what we do will no longer be different from what we tell", FA Albert, L Duserre, International Journal of Biomedical Computing v 35 (supplement, 1994) pp 201–204

[AIS95] *'AIS — Advanced Information System'*, FHS Computer Unit, 1995

[Boy94] *'Draft guidance for the NHS on the confidentiality, use and disclosure of personal health information'*, N Boyd, Department of Health, 10 August 1994

[Bru95] "Is your health history anyone's business?" McCall's Magazine 4/95 p 54, reported by M Bruce on Usenet newsgroup comp.society.privacy, 22 Mar 1995

[BMA95] *'A Bill Governing Collection, Use and Disclosure of Personal Health Information'*, British Medical Association 1995

[Cae95] WJ Caelli, *personal communication, July 1995*

[CR94] "Who's reading your medical records?" Consumer Reports, Oct 94 pp 628–632

[DGMW94] *'How to Keep a Clinical Confidence'*, B Darley, A Griew, K MsLoughlin, J Williams, HMSO 1994

[DL95] Data Logic product information at `http://www.datlog.co.uk/`

[DPA84] *'Data Protection Act'*, 1984

[DPR95] *'Identity Cards: A Consultation Document CM2879 — Response of the Data Protection Registrar'*, October 1995

[EU91] *'Information Technology Security Evaluation Criteria'*, EU document COM(90) 314 (June 1991)

[EU95] *'On the protection of individuals with regard to the processing of personal data and on the free movement of such data (final)'*, Directive of the European Parliament and the Council, adopted by the Council on 24 July 1995

[Gil95] "MDU Muddle re Death Pills", C Gilbert, gp-uk mailing list, 23rd October 1995

[GC95] *'A Strategy for Security of the Electronic Patient Record'*, A Griew, R Currell, Institute for Health Informatics, University of Wales, Aberystwyth, 14th March 1995

[GMC1] *'Good medical practice'*, General Medical Council, 178–202 Great Portland Street, London W1N 6JE

[GMC2] *'Confidentiality*, General Medical Council, 178–202 Great Portland Street, London W1N 6JE

[GTP93] "Privacy and Security of Personal Information in a New Health Care System", LO Gostin, J Turek-Brezina, M Powers et al., Journal of the American Medical Association v 20 (24/11/93) pp 2487–2493

[Haw95] "Confidentiality of personal information: a patient survey", A Hawker, Journal of Informatics in Primary Care, 1995 (March) pp 16–19

[HRM93] "RMs need to safeguard computerised patient records to protect hospitals", Hospital Risk Management 1993 v 9 (September) pp 129–140

[JCG88] "GMSC and RCGP guidelines for the extraction and use of data from general practitioner computer systems by organisations external to the practice", Appendix III in 'Committee on Standards of Data Extraction from General Practice Guidelines' Joint Computer Group of the GMSC and RCGP, 1988

[JHC94] "Nurse Jailed for Hacking into Computerised Prescription System", British Journal of Healthcare Computing and Information Management v 1 (94) p 7

[LB94] "Your Secrets for Sale", N Luck, J Burns, The Daily Express, 16/2/94 pp 32–33

[MRI94] "Integrated Health Delivery Needs Integrated Health Record Systems", Medical Records Institute newsletter v 3 no 5 (December 94) pp 1–9

[Mac94] Letter from AW Macara to JS Metters, 31 October 1994, on 'Draft guidance for the NHS on the confidentiality, use and disclosure of personal health information'

[Mar95] "Fear of Flowing", DC Markwell, Proceedings of the 1995 Annual Conference of The Primary Health Care Specialist Group of the British Computer Society, pp 36–42

[NHS92] *'Handling confidential patient information in contracting: A Code of Practice'*, NHS Information Management Group EL(92)60, catalogue number 2009(c), news info 132

[NHS95] *'The Handbook of Information Security — Information Security within General Practice'*, NHS Executive Information Management Group E5209 (May 1995)

[NZ94] *'Health Information Privacy Code 1994'*, New Zealand Privacy Commissioner, 1994/1/1

[OTA93] *'Protecting Privacy in Computerized Medical Information'*, Office of Technology Assessment, US Government Printing Office, 1993

[PK95] "GP Practice computer security survey", RA Pitchford, S Kay, Journal of Informatics in Primary Care, September 95, pp 6–12

[Ros95] "Institutionell-organisatorische Gestaltung informationstechnischer Sicherungsinfrostrukturen", A Roßnagel, Datenschutz und Datensicherung (5/95) pp 259–269

[RAC+93] *'Interim Code of Practice for Computerised Medical Records in General Practice'*, Royal Australian College of General Practitioners, February 93

[RFA93] *'Requirements for accreditation, general medical practice computer systems'*, NHS management executive 1993

[RL95] "For Sale: your secret medical records for £150", L Rogers, D Leppard, Sunday Times 26/11/95 pp 1–2

[RSM92] *'Computers in Medical Audit'*, second edition, M Rigby, A McBride, C Shields, Royal Society of Medicine, London, 1992

[Sch95] *'Applied Cryptography'*, B Schneier, second edition, Wiley 1995

[See95] "Marketing use of medical DB", M Seecof, Usenet newsgroup comp.risks 17.12

[Smu94] *'Health Care Information: Access and Protection'*, RH Smuckler, Institute for Primary Care Informatics, 1994

[Som93] *'Medical Ethics Today — Its Practice and Philosophy'*, A Sommerville, BMA 1993

[Tho95] "Sex Stalker Plays Doctor to Trick Victims", M Thomas, PA newswire no 1236, 7/7/95

[TCP+93] *'The Canadian trusted Computer Product Evaluation Criteria'*, Communications Security Establishment, Government of Canada, January 1993

[TCS+85] *'Trusted Computer System Evaluation Criteria'*, US Department of Defense document 5200.28-STD, December 1985

[USA95] "Online medical records raise privacy fears", USA Today, 22/3/95 pp 1A–2A

[Woo95] "The computer-based patient record and confidentiality", B Woodward, New England Journal of Medicine v 333 no 21 (95) pp 1419–1422

[Wri91] *'The Law of Electronic Commerce: EDI, Fax and Email'*, B Wright, Little, Brown (fourth edition with supplement) 1994

[WHC95] *'Workshop on Health Care — Confidentiality: discussing current initiatives'*, held at the BMA on 4th April 1995; transcript supplied by RH Pyne