A Security Policy Model for Clinical Information Systems

Ross J. Anderson
University of Cambridge Computer Laboratory
Pembroke Street, Cambridge CB2 3QG
ross.anderson@cl.cam.ac.uk

Abstract

The protection of personal health information has become a live issue in a number of countries including the USA, Canada, Britain and Germany. The debate has shown that there is widespread confusion about what should be protected, and why. Designers of military and banking systems can refer to Bell-LaPadula and Clark-Wilson respectively, but there is no comparable security policy model that spells out clear and concise access rules for clinical information systems.

In this article, we present just such a model. It was commissioned by doctors and is driven by medical ethics; it is informed by the actual threats to privacy, and reflects current best clinical practice. Its effect is to restrict both the number of users who can access any record and the maximum number of records accessed by any user. This entails controlling information flows across rather than down and enforcing a strong notification property. We discuss its relationship with existing security policy models, and its possible use in other applications where information exposure must be localised; these range from private banking to the management of intelligence data.

1 Introduction

The introduction of nationwide health information networks has caused concern about security. Doctors are worried that making health information more widely available may endanger patient confidentiality. In the USA, there is controversy over a proposed law on medical privacy [Ben95]. In Ontario, an attempt to give the Minister of Health access to all medical records was defeated after intense pressure by the public and the Ontario Medical Association [Lan95]. In Germany, there has been disquiet about the introduction of a uniform national smartcard system to handle health insurance payments.

In the UK, the government has commissioned a nationwide health information network, and is setting up a number of centralised applications that will use it. One of them will centralise the billing of hospital treatment in a single system that will process large amounts of personal health information, and make various analyses available to administrators. Doctors will remain responsible for the security of clinical information which they originate; yet the doctors' main professional organisation, the British Medical Association (BMA), has been refused information about the security mechanisms that are supposed to protect patient information on the new network and its applications.

It also became clear that there was much confusion about the actual threats, and about the protection measures that it would be prudent to take. For these reasons, the BMA asked the author to study the threats to personal health information [And95] [And96c], and then to draw up a security policy model [And96a] and interim guidelines for prudent practice [And96b]. In this paper, we present the policy model. The presentation is of necessity abbreviated, and readers are urged to obtain a the full document from the BMA or via the web [And96a].

1.1 A note on terminology

We define and discuss the terminology at length in the full policy, so it is merely summarised here. By 'clinician' or 'clinical professional' we mean a licensed professional such as a doctor, nurse, pharmacist, radiologist or dentist who has access in the line of duty to 'personal health information'; by this we mean any information concerning a person's health or treatment that enables them to be identified. By 'patient' we mean the patient or his representative — whoever must give consent and be notified. We ignore delegation of access to persons such as receptionists, as a clinician remains responsible for their actions.

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 crypto 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.

Finally, some authors draw a distinction between 'confidentiality' (which protects the interests of the organisation) and 'privacy' (which protects the autonomy of the individual). We will rather follow the common medical usage in which both words interchangeably mean 'privacy'.

1.2 The ethical basis of confidentiality

The Hippocratic oath says:

> Whatsoever I shall see or hear in the course of my dealings with men, if it be what should not be published abroad, I will never divulge, holding such things to be holy secrets.

Doctors in most countries interpret the words 'should not' in terms of consent. In Britain, for example, the doctors' disciplinary body is the General Medical Council which expresses the duty of confidence as follows [GMC1]:

> 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.

The GMC further stipulates that doctors who record or who are the custodians of confidential information must make sure that it is effectively protected against improper disclosure when it is stored, transmitted, received and disposed of [GMC2]. Other clinicians such as nurses, pharmacists and physiotherapists are under similar professional obligations. Finally, a number of countries have laws on data protection; and from 1998, an EU directive on data protection will compel European countries to make patient consent the paramount principle in the protection of personal health information.

Consent must be informed and voluntary. For example, patients must be made aware that information may be shared between members of a care team (such as a general medical practice[1] or hospital department); and if researchers want access to records which cannot effectively be made anonymous, then every effort must be made to inform the patient and gain his consent, which must be renewed every five years [Som93].

[1]the UK 'general practitioner', or GP, is the primary care physician or 'family doctor'

A number of exceptions to this rule have developed over time. For example, Britain has rules on notifiable diseases, adverse drug reactions, non-accidental injuries and fitness to drive [Boy94]. However, these exceptions are peripheral, as disclosures are rare and are typically made on paper.

1.3 Threats to clinical confidentiality

Many organisations 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 abuse by insiders. For example, most of the big UK banks now let any teller access any account. The effect is that private eyes get hold of information by bribing tellers and sell it for £100 or so [LB94]. The practice was made illegal by 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 should have been expected. The likelihood that information will be improperly disclosed depends on its value, and the number of people who have access to it. Aggregation increases both these risk factors at the same time. It may also create a valuable resource which brings political pressure for legalised access by interests claiming a need to know [Smu94].

Health systems are no different. At present, privacy depends on the fragmentation and scattering inherent in manual systems and standalone computers; removing this without introducing effective compensating controls is unethical. There have been persistent UK press reports of health records being sold by private detectives for as little as £150 [LB94] [RL95]. Perhaps the most serious reported case is that of 'Dr Jackson', a Merseyside sex stalker, who wins the confidence of young women by discussing their family medical history over the telephone, urges them to examine themselves, tries to arrange meetings, and then attempts to abduct them. Police believe that he is a health worker or a computer hacker [ISM95].

The US experience is much worse. This may be partly due to the control exerted by HMOs and insurance companies, and partly because networking has advanced somewhat more than in Britain:

- 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 Harris poll on health information privacy

showed that 80% of respondents were worried about medical record privacy, and a quarter had personal experience of abuse [GTP93];

- Forty percent of insurers disclose medical information to lenders, employers or marketers without customer permission [CR94]; and over half of America's largest 500 companies admitted using medical records to make hiring and other personnel decisions [Bru95].

The problem was studied by the US government's Office of Technology Assessment, which confirmed that the main threats come from insiders, and are exacerbated by the data aggregation that networked computer systems encourage [OTA93]. There is now controversy over a bill introduced into the US Congress which would remove the patient's right to sue should his privacy be breached and harm result [Ben95]. This bill is sponsored by a credit reference agency that is currently building a large network for trading health information.

However, doctors do not accept, and in many countries administrators do not even claim, that the uncontrolled aggregation of data is ethically permissible. In the words of David Bellamy, Principal Medical Officer at the UK 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].

The real political struggle here is over control, and in particular whether access decisions should be taken by the patient (as is required by the GMC) or by administrators (as is implicit in the use of the phrase 'need-to-know'). After all, while it is the patient who gives consent, it is the administrator who decides who needs to know. Recent court cases have eroded the strength of 'need-to-know' arguments: it has been ruled that even a doctor's HIV status may not be disclosed, as the small risk to patients' health does not outweigh the public interest in maintaining the confidentiality that enables infected persons to seek help [DGMW94]. In this context, a recent government attempt to get doctors to disclose details of HIV and AIDS sufferers to assist in 'estimating the need for local community services' is being resisted by the profession.

In addition, the EU directive is about to enforce the principle of consent throughout Europe. So administrators are scrambling to redefine 'consent'.

The UK government's initial position was that a patient gave 'implied consent' to information sharing by the mere act of seeking treatment. More recently, officials have tried to redefine 'informed consent' as the consequence of putting up notices informing patients that their personal health information may be shared with officials. Consent as understood by the layman has been renamed 'explicit consent' and derided as unpractical. The struggle continues.

However, the purpose of this document is normative more than descriptive. Our goal is to describe things as they should be, and as they would be if attention were paid to the ethical rulings of the GMC, the EU directive, and surveys showing that most patients are unwilling to share their personal health information with administrators [Haw94] [CB95].

1.4 Other threats to clinical information

The integrity and availability of medical information are also important, for the obvious safety and medico-legal reasons. While mail, fax and telephone messages are just as prone to failure as computer systems, their failure modes are more evident. Software bugs could alter the numbers in a laboratory report without changing it so grossly that it would be rejected; viruses have already destroyed clinical information; and concern has been expressed that the lack of standards in clinical EDI may lead to data being interpreted differently by different systems, with life-threatening effect [Mar95].

Turning from random to malicious failure, it is clearly possible (in the absence of comsec mechanisms) for outsiders to intercept or modify messages. But most reported attacks on clinical information systems consist of the physical theft of the computer from a surgery, with over 11% of British GPs having suffered this [PK95],. The majority of other attacks on system integrity are likely to be carried out by insiders. In typical cases of which we are aware, attackers have tried to shift liability by altering a record of malpractice [Ald95], to abuse prescription systems [JHC94], or to commit straightforward theft or fraud by changing records of stocks or contracts.

There are also system level effects. For example, attacks on integrity may be made more likely by loss of confidentiality: if medical records become widely used outside of clinical practice for purposes such as hiring and credit decisions (as in the USA), then there will be motives to alter them [Woo95]. The same can happen if system components are shared with systems having purposes other than healthcare. A Spanish healthcard doubles as a bankcard [Bro95], so criminals might try to break it; and if a health card came to be used as an identity card, then civil libertarians might also join in [DPR95]. Health information might also become entangled with civil liberties issues through the use of escrowed cryptography; and there is concern about how electronic records may be made reliable enough to be used as evidence in court.

However, the greatest concern of both clinicians and the courts is that if patients cease to believe that their clinical confidences will be respected, they will suppress relevant information, leading not just to inaccurate records but to poor treatment of individual patients and to an increased risk to others (e.g., from the spread of infectious disease) [DGMW94].

1.5 Protection priorities

For all these reasons, the confidentiality and integrity of medical systems may not be considered in isolation, and have to be considered at two levels.

At the local level, we are concerned with the threats to information held on a single system, such as that of a general practice or hospital department. Examples are theft of the computer and the unauthorised disclosure of information by a dishonest or careless employee. The associated risks can be controlled by more or less well understood techniques, such as staff training, regular backup and audit: the BMA has issued guidelines on this [And96b].

However, in this document, our main concern is the security policy used to control global threats — those threats to the privacy, integrity or availability of the medical records of large numbers of people, which arise from the ill-considered aggregation of systems, the erosion of patient consent, and various other causes. We are not overly concerned that a GP's receptionist can access the records of his 2,000 patients; but we would be extremely concerned if a network gave the receptionists of Britain's 32,000 GPs access to the records of all 56,000,000 residents.

The global and local domains are linked. Where the aggregation threat arises from networking many small systems together, rather than from building large central databases, then most of the global protection mechanisms must be implemented locally. Another example is that local systems may have common failure modes: private detective agencies routinely obtain personal health information by making false pretext telephone calls to the patient's doctor or health authority. Here, too, the global threat can only be countered by local measures, and the BMA recommends the use of callback-based authentication protocols to ensure that personal health information is only shared with clinicians or with suitably accredited clinical systems [And96b].

This brings us back to our central problem, which is to examine what sort of systems might prudently be trusted with personal health information. Before we can evaluate the security of particular systems, we need to know what the security mechanisms are supposed to achieve. This means having a security policy that says who can access what.

2 Security Policy

We will now set out a security policy model for clinical information systems, in a form comparable with the Bell-LaPadula model for military systems [BL73] and the Clark-Wilson model for banking systems [CW87]. Our policy is based on the rules set out by the General Medical Council [GMC1] [GMC2] and the British Medical Association [Som93], which incorporate much clinical experience. It has also informed by extensive discussions with clinical professionals.

As usual with policy models, we will attempt to translate the application requirements into a set of rules that say which subject can access which object. Here a subject may be a computer user (such as a doctor, health administrator or outside hacker) or a computer program acting on behalf of a user; the objects are the information held in the system, and may include both programs and data; and access may include the ability to read, write and execute objects.

We also make a number of simplifying assumptions. These are discussed in the full policy; the most important is that records pertain to only one person at a time. When this assumption breaks down, things get complicated; special rules need to be made for environments such as obstetrics, pediatric psychiatry and genetics where records often contain clinical facts about more than one identifiable person.

2.1 Access control lists

Since a typical patient has fewer doctors than a typical doctor has patients, it is convenient to state the policy in terms of access control lists rather than capabilities.

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 clinicians in that practice may read it and append things to it. Such practices typically keep their few highly sensitive records on paper in a locked drawer. However, patients whose records are kept in this way fall outside many of the safety mechanisms, and with the introduction of networking, access control lists need to be made explicit and consistent across a range of systems.

Groups and roles 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'. If they make frequent use of a locum, then they might add 'locum' to the above list, and assign individuals to the role at appropriate times.

The problem is that 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. Extra restrictions may then be needed, and roles may be preferable to groups; for example, one might use active badges [WHFG92] to limit access to 'any clinical staff on duty in the same ward as the patient'. This would create the electronic equivalent of a traditional note trolley, but with the added advantage that a record can be kept of who consulted what. We will discuss attribution more fully below; here we will merely remark that groups and roles are not virtual clinicians, but mechanisms that simplify the access mapping between identified clinicians and identified patients.

There are clearly some kinds of clinical information that are highly sensitive and should only be available to a restricted access list. The paternalistic approach is to lump into this category all psychiatric records, records of sexually transmitted disease, information given by or about third parties, and records of employees and their families. But the actual sensitivity of a record is always a decision for the patient, and there is little correlation between the above list and patients' actual priorities [CB95]. An AIDS campaigner might consider his HIV status to be public knowledge, while a Jehovah's witness might consider even a blood transfusion to be profoundly shameful [GC95]. For this reason, patients must be informed of a care team's access control policy when they first enrol, and have the opportunity to restrict access further if they wish. Since consent must be voluntary, systems must be designed so that the standard of care received by patients who do not consent to information sharing will be degraded as little as possible.

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. We will rather assume that they get full access to a temporary copy of the primary record; and this is a better model of how they actually work.

2.2 Record opening

Rather than trying to deal with multilevel objects, we will assume that there are multiple records. Thus a patient might 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, but 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.

The reason for this is that it would seem unnatural for a patient who had been referred to hospital for tests to have to give explicit consent at the hospital for the test results to be sent back to his GP.

2.3 Control

Apart from the patient himself, only clinicians may have access to his records. 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, whether for the patient or for the clinician. 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 a local government computer that was hacked — then she could still be liable, and might have no effective recourse.

So only clinicians are trusted to enforce the principle of informed consent, and control of any identifiable clinical record must lie with the 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 health authorities must have 'safe-havens' — protected spaces under the control of an independent clinician — to which copies of records may be sent if there is a dispute [NHS92]. In both Germany and Ontario, medical associations buffer billing information; they have access to detailed item of service claims but pass on only aggregate information to the government agencies that pay for treatment.

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 it should be provided on paper. In the UK, 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.

2.4 Consent and notification

The patient's consent must be sought for other persons such as the clinician's colleagues to be added to the access control list, and he must be notified of every addition. There are some exceptions to consent, as noted above, but even where a doctor is obliged to pass to a third party some information — such as a diagnosis of a notifiable disease — the patient must still be notified of this information sharing. The legislation presently before the US Congress would permit notification to be delayed for 90 days in the case of law enforcement access, but not to be omitted.

These strong notification requirements flow from the principle of consent. They also help control fraud, as medical benefits are cash limited in many countries and patients with expensive treatment needs may impersonate other patients when their budget runs out. A letter to an unsuspecting victim that his records had been opened by a physician of whom he had never heard is often how fraud is detected; and an effective way of identifying abusive access may be to screen for clinicians who read a patient's record without subsequently sending in a bill [Sim96].

Most importantly, notification provides an end-to-end audit mechanism that is not open to capture by governments and healthcare managers.

> **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.

The mechanics of this are not as onerous as they might seem. In most cases, the patient will consent to the default access control list — all the clinicians in the practice — and that will be the end of the matter. When patients are referred to specialists in the normal course of events, there will also be consultations with the GP at which consent and notification can be dealt with. The GP will usually only send a written notification in the case of emergency access (e.g. after an emergency hospital admission), access by police or others under court authority, or following a security failure which we treat as the mistaken addition of an unauthorised person to the access control list.

But even so, notification is not entirely straightforward. Recently, GPs were asked to notify a possible side-effect to women using certain contraceptives; this raised issues of how to deal with young girls who were having sex without their parents' knowledge, and women whose spouses had had a vasectomy and were taking the pill in a new extramarital relationship. The solution, which is already practised in STD clinics, is for the clinician to ask the patient at the outset of the relationship how to send any notices.

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 government garnering emigration data from records returned by GPs to health authorities 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.

2.5 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. Prudence may dictate keeping access to records until after a lawsuit for malpractice could 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.

The rules are still not fully worked out, and so our use of the word 'appropriate' glosses a number of open issues. There are cases (such as chronic illness) in which records must be kept for longer than usual. There are also disputes about whether they could be retained against the patient's wishes to defend possible lawsuits. In some countries (e.g., Germany) clinicians may claim a copyright in records they create, while in others (e.g., Britain) they are routinely transferred to the patient's new doctor.

In general patient consent is not immutable, but rather a continuing dialogue between the patient and the clinician [Som93]. So 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 record to a clinician of the patient's choice for the rest of the statutory period.

Finally, we do not want information that has been identified as inaccurate, such as simple errors and subsequently revised diagnoses, to be mistakenly acted on. But 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.

2.6 Attribution

We must next 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 UK requirements for accreditation will typically record all write accesses; even if material is removed from the main record, the audit trail must enable the state of the record at any time in the past to be reconstructed and all changes to be attributed [RFA93]. If implemented properly, this will have the same effect as restricting write access to append-only and marking all append operations with the clinician's name. Our new requirements are that read accesses be logged, so that breaches of confidence can be traced; 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 by the patient too if he is competent to consent [Som93]. When such life critical functions are automated, the mechanisms — including those for attribution — must be engineered to the standards required in life support systems.

Rarely invoked requirements may be supported by manual mechanisms. For example, in most countries, patients may read their records and append objections if they wish. The common procedure is for the clinician to print out the record for the patient, and then if there are any comments, to append them and print them out too for confirmation.

2.7 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.

This rule naturally gives rise to a lattice [Den76], in which domination is equivalent to the inclusion of access control lists. Information flow can thus be controlled using mechanisms that are well understood from the world of multilevel security [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.

The second-order problems of multilevel secure systems, such as polyinstantiation, have an interesting counterpart in clinical systems. Where two records with different access control lists correspond to the same patient, should the existence of the more sensitive record be flagged in the other one?

This is a known dilemma on which there is still no consensus [GC95]. If the existence of hidden information is flagged, whether explicitly or by the conspicuous absence of information, then inferences can be drawn. For example, doctors in the Netherlands removed health records from computer systems whenever a 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 sub judice.

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.

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.

2.8 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 sporadic, and incidents such as the Jackson case continue to be reported. Networking such systems together could be disastrous. Having 2,000 staff each with access to a million records is bad enough; but the prospect of 200 such hospitals connected together, giving 400,000 staff access to records on most of the population, is profoundly unsettling.

However, even if cross-domain access is restricted to a few trusted staff at each hospital (perhaps an 'officer of the watch' in the emergency room) there must be controls that protect both patients and clinicians.

In this policy model, the primary control is notification, and the secondary control is to keep a list 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. The natural location for the secondary controls might be with a professional disciplinary body such as the GMC.

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; and the same goes for systems that de-identify and aggregate records for research purposes. We shall discuss them below.

2.9 The Trusted Computing Base

Finally, we must ensure that the security mechanisms are effective in practice as well as in theory.

> **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 Bundesamt für Sicherheit in der Informationstechnik has recently recommended that systems which process clinical diagnoses of identifiable persons should be evaluated to E4/E5 [BSI95]. We have recommended that the evaluation level should depend on the number of people whose personal health information was at risk; we suggested E2 for small systems, such as those used in general practice, and E4 for large systems, such as those used in district hospitals where a million patients' records could be on file [And96a].

As schemes such as ITSEC are oriented towards military systems and evaluations under them are expensive, some industries run their own schemes. For example, UK insurers evaluate the security of burglar alarms using the laboratories of the Loss Prevention Council, which they jointly fund. Similar industry-wide arrangements might be made for clinical systems, but would have to enjoy the support of both clincians and patients. Britain's current accreditation system for clinical software is run by the NHS and so does not inspire universal confidence.

As always, the most important factor in achieving a workable security solution is often not so much the choice of mechanisms but the care which is taken to ensure that they work well together, and that the system can be managed by a clinician whose computer literacy and administrative tidiness are less than average. It must be less trouble to manage the system properly, and care should be taken to evaluate systems under realistic assumptions about the skills and discipline of their operators.

3 Protection Mechanisms

The TCB 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 compusec mechanisms used to build a TCB that enforces information flow controls in a single machine are fairly well understood. The more interesting part concerns the comsec mechanisms needed in distributed heterogeneous systems.

3.1 Comsec mechanisms

In our view, the primary purpose of comsec in medicine is to ensure that access controls are not circumvented when a record is sent from one computer to another. This might happen, for example, if an object is sent to a system that corrupts its access control list, or that does not enforce the principle of 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 the wrong doctor or even to a mailing list or newsgroup.

The secondary purpose of comsec mechanisms is to protect the integrity of data sent through a network. Records such as pathology reports might, as discussed above, become accidentally corrupted in ways which are not obvious to the recipient. There is also controversy in some countries on whether electronic records are adequate for legal purposes. For these reasons, it may be desirable to use digital signatures or other strong integrity checks.

3.2 Trust structures

Digital signatures also 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 the trust structure bundled with PGP, in which a web of trust is built 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 doctor in each 'natural community' (a district hospital plus the several dozen general practices that feed patients to it).

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

However, it is vital that electronic trust structures reflect the actual nature of trust and authority in the application area [Ros95]. In the practice of 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 we could end up with a security system which clinicians considered to be a central imposition rather than something trustworthy under professional ownership and control.

It is by no means clear that clinical systems can be accommodated by the certification structures considered in X.509 and X9.3. For example, a doctor might want to have a number of different keys (e.g. where she works in a hospital, a prison and a general practice); some of these will be signed by organisations, and others might not be (e.g. for her private practice). Yet we will need to keep a dependable count of the total number of cross-domain records she accesses, and this might be linked to key certification.

3.3 Propagation of access control

In any case, once clinicians have acquired suitably certified key material, the integrity of access control lists across a network can be enforced by means of a ruleset such as the following:

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 certification, or by some 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.

Abuse can also be made harder by a rule that records must be given rather than snatched; access requests should never be granted automatically but subject to patient consent — or, in the case of emergency, to a case by case clinical decision.

Accreditation can be enforced in the usual way by not supplying key material until the documentation is complete. This is one advantage of central or at least structured certification over the web-of-trust approach.

Encryption is by no means the only comsec option; anonymity may often be simpler. For example, a system for delivering laboratory reports to GPs might replace the patient's name with a one-time serial number, which could be bar-coded on the sample label. The test results might then be transmitted in clear (with suitable integrity checks).

3.4 The importance of effective audit

When records are moved from paper to electronic form, abuse can become orders of magnitude easier. Previously, an intruder might have had to walk into an office where he has no business and look in a filing cabinet at risk of being challenged; but for a hospital employee to look at a clinical record on screen is an intrinsically innocuous act as far as bystanders are concerned. In this way, computerisation eliminates one of the major controls on information leakage.

Compensating controls are needed, and access controls alone are not enough. A clinician can always falsely declare that a patient has been admitted unconscious and request a copy of the record; if there is no systematic effort to detect and punish such abuse, then it can be expected. So our compensating controls must include an audit system that presents the intruder with a credible chance of being caught. Otherwise, systems will fail to meet the agreed goal that electronic records must be at least as secure as the paper records that they replace.

Now one of the interesting facts about clinical systems is that authority is not trusted. When building a military system, we can assume that the President or Prime Minister is on our side; and banking systems are not usually designed to prevent frauds by senior executives.

Medicine is different. For generations and in many countries, the authorities have striven to increase their access to personal health information, while both patients and clinicians have resisted this. In the UK, for example, the argument over who owns the record has been going on since at least 1911.

This complicates the design of an audit system. Where shall the audit trail be kept, and who shall be trusted to act on it?

Under the current UK arrangements, the responsibility for detecting and reacting to security incidents is left to local line management. In the words of the responsible minister, "there is no central collection of statistics on recorded instances of unauthorised access to personal health information, whether via computer systems or paper records" [Hor96]. Similarly, patients are unlikely to be told. There may be external audits, but their ineffectiveness at detecting abuse is well known. After all, the auditors' main desire is to be reappointed. So what can be done?

Our approach has been to provide two auditors, both of whom have an interest in detecting abuse and acting on it. The first is the patient, who must be informed of all the people who get access to his record. This notification will also cover security breaches, as we treat them as additions to the access control list.

The second is the central body that records which clinician accessed which record outside her own care team. We suggest that this be the body responsible for clinical discipline, such as the GMC for UK doctors. Its function will be to look for potentially abusive access patterns.

The exact balance between distributed and centralised audit will be a function of how healthcare is organised in the country in question. For example, Simmons' idea of flagging for investigation all accesses that are not followed by an invoice may be very effective, but it might have to be implemented in a distributed way in the US and centrally in the UK in order to get access to payment information.

3.5 Statistical security

Our security policy relates to personal information, and records may be removed from its scope if they are de-identified and aggregated, as often happens for research or census purposes. The problem is that the process is often incompetently designed; for example, a recent survey of HIV and AIDS proposed that patients' names be replaced by Soundex codes of their surnames, and accompanied by their birth dates and postcodes [MS95].

This is clearly inadequate. Britain has established guidelines 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].

This topic has been researched extensively in the context of census data [Den82], but the problem is even harder in the medical case. If an attacker can submit 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. A Norwegian proposal is that researchers should only be granted access to linkable data on a regional rather than national basis, and even then within protected space; researchers would travel to the regional registry, present their authorisation, run their queries, and come away with only statistical results [Boe93].

However most research does not involve access to large volumes of data. A typical scientist might want to study the records of everyone diagnosed with Creutzfeld-Jakob disease in the last 20 years; she can request consent from the deceased persons' relatives. In fact, she needs to do this if she is to get vital background information on the victims' lifestyles.

3.6 Medical records or patient records?

So far, most electronic clinical record systems have mirrored the paper-based practice in that each clinical team has its own filing system and information flows between them in the form of 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 doctor-based rather than patient-based; information flows between them in the form of summaries; and the lifetime record that links them all together is the record kept by the patient's GP.

There has been interest recently in a different model of clinical information, namely that there should be a single unified patient record that is opened on confirmation of pregnancy, closed on autopsy, and accumulates all the clinical notes and data in between [MRI94]. Proponents of this model often claim that the records are patient based rather than doctor based, though in practice it may mean moving the primary record from the patient's GP to a hospital, health authority, HMO or even insurer.

Many people will consider this to be rather undesirable; it will also be in conflict with the inertia of tradition and of installed systems. There are also many data management problems that affect security. Records may be very large (such as CAT scans and the records of long chronic illnesses); some records contain other patients' personal information too (e.g, birth records contain data on the mother); and records of some treatments cannot be transferred because of statutory prohibitions (e.g. treatment in prisons and STD clinics).

Now suppose that I walk into a hospital and claim that my demons are bothering me. When asked my name I reply 'John Major'. May the psychiatrist get the prime minister's record and append a diagnosis of schizophrenia? In other words, does a patient-based record force us to authenticate patients 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?

The above is by no means an exhaustive list. For a discussion of some of the security policy complexities of unified electronic patient record systems, see Griew and Currell [GC95]. As their paper makes clear, unified electronic patient records would force us to make our policy model significantly more complex.

We suggest that the unified record would be a bundle of disparate objects whose access control lists might only intersect in the patient himself. It is far from clear what engineering gains may be had from forcing all these objects to reside in the same store. The onus is on proposers of such systems to provide a clear statement of the expected health benefits, and to analyse the threats, the cost of added countermeasures and the likely effects of the residual risk.

4 Standards

Encryption of medical records has been mandated by the data protection authorities in Sweden for several years, and is being introduced in Norway. As already mentioned, a number of countries are building trusted certification authorities which will sign doctors' keys [AD94]. A European standardisation group for Security and Privacy of Medical Informatics (CEN TC 251/WG6) is working on a draft standard which recommends the encryption of identifiable clinical data on large networks.

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 New Zealand Health Information Privacy Code [NZ94], and the Office of Technology Assessment report cited above may also be referred to. They each contribute in different ways to our understanding of threats, of the principle of consent, and of the technical options.

However there is as yet no access control model in the sense understood by the computer security community, and it is hoped that this model may help clarify what medical systems builders should be trying to achieve with all these mechanisms.

5 Relation with Other Models

Our model can express Bell-LaPadula and lattice models, where the partial order is inclusion of access control lists. However the converse does not hold, since we maintain state about how many objects a particular subject has accessed, and have the externality of a strong notification requirement.

It is unlikely that our model will replace Bell-LaPadula in a traditional military application such as managing stores, since such applications are essentially capability based (there are more soldiers than security labels) whereas medicine is access control list based (there are more patients than doctors). However, there may be applications, such as intelligence, where the large number of security labels makes an access control list approach more economic. Perhaps strong notification to case officers of all access to intelligence records would have led to the earlier capture of Aldrich Ames; we understand that his access to the records of the agents whom he betrayed was notified to senior officials, but they did nothing. Perhaps the case officers would have done more; we can only speculate. Another application might be to enable account executives in private banking and other high value service industries to control access to information about their clients.

The one existing policy model which can capture most of the principles set out here is Clark-Wilson [CW87]. Let a constrained data item be a record together with its ACL; let the initial validation procedures be firstly, record opening, secondly, the validation of laboratory and other data by a clinician's signature, and thirdly, the process of adding a new name to the ACL by consultation and notification; and let the transformation procedures be the acts of appending material to the record and of passing information to some subset of the ACL. The Clark-Wilson audit requirements are fulfilled as all records are append-only, and all additions to ACLs are notified.

Strong notification is still not completely captured (though it could be if each patient were also a system user). In theory, secure time is required to ensure that an attacker does not change the system clock and cause records to be deleted. However, most of our policy model can clearly be built on a Clark-Wilson base.

This is curious, as Clark-Wilson is commonly thought of as an integrity model, and yet here we are using it to instantiate a security policy whose primary goal is confidentiality and which is strictly more expressive than the lattice and Bell-LaPadula models. The research community might care to consider the implications.

6 Conclusion

We have discussed the threats to the confidentiality, integrity and availability of personal health information in the light of experience in the UK, the USA and elsewhere, and proposed a clinical information security policy that enables the principle of patient consent to be enforced in the many heterogeneous distributed systems that are currently under construction.

Its goal is to ensure that any lack of consent is propagated and enforced. This gives rise to a privacy property that is much stronger than the confidentiality enforced by multilevel models, but which may be similar in some respects to the compartmented mode policies used in the intelligence community.

One curious fact about our model is that it can be most closely expressed using the machinery of Clark-Wilson to protect its access control lists. This suggests that there may be other links between the various aspects of confidentiality, integrity and availability as they are expressed in security models and implemented in real systems.

Acknowledgements: The research described in this paper was funded by the British Medical Association, and valuable input was received from a number of clinicians, 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 Gus Simmons, Bob Morris, Stewart Lee, Roger Needham, Markus Dichtl, Bruce Christianson, Ian Jackson, Mike Roe, Mark Lomas, Jeremy Thorp, Roy Dainty and Ian Keith; and from philosophers including Beverly Woodward, Ann Somerville and Keith Tayler. I am also grateful to the Isaac Newton Institute for hospitality while this paper was being written.

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

[And95] "NHS wide networking and patient confidentiality", RJ Anderson, in *British Medical Journal* v 310 no 6996 (1 July 1996) pp 5–6

[And96a] *'Security in Clinical Information Systems'*, RJ Anderson, published by the British Medical Association, January 1996; also available from `http://www.cl.cam.ac.uk/users/rja14/#Med`

[And96b] "Clinical system security: interim guidelines", RJ Anderson, in *British Medical Journal* v 312 no 7023 (13 Jan 1996) pp 109–111

[And96c] "Patient Confidentiality — At Risk from NHS Wide Networking", RJ Anderson, *to appear in Proceedings of Healthcare 96, March 96*

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

[Ben95] *'Medical Records Confidentiality Act of 1995'*, B Bennett, US Senate S.1360, 24th October 1995

[Boe93] *'Pseudonymous Medical Registries'*, E Boe, Norwegian Official Report 1993:22

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

[Bro95] "Sanitas launches health credit card", S Brown, Cards International (6 October 1995) p 3

[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

[BL73] *'Secure Computer Systems: Mathematical Foundations'*, DE Bell, L LaPadula, Mitre Corporation Technical Report ESD-TR-73-278 (1973)

[BSI95] *'Chipkarten im Gesundheitswesen'*, Bundesamt für Sicherheit in der Informationstechnik, Bundesanzeiger 4 May 1995

[Cae95] Personal communication, WJ Caelli, July 1995

[CB95] "Confidentiality of medical records: the patient's perspective", D Carman, N Britten, British Journal of General Practice v 45 (September 95) pp 485–488

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

[CW87] "A Comparison of Commercial and Military Computer Security Policies", in Proceedings of the 1987 IEEE Symposium on Security and Privacy pp 184–194

[Den76] "The Lattice Model of Secure Information Flow", DER Denning, Communications of the ACM v 19 no 5 (May 1976) pp 236–242

[Den82] *'Cryptography and Data Security'*, DER Denning, Addison-Wesley 1982

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

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

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

[GMC1] *'Good Medical Practice'*, General Medical Council, 178–202 Great Portland Street, London

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

[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

[Haw94] "Confidentiality of personal information: a patient survey", A Hawker, Journal of Informatics in Primary Care (March 1995) pp 16–19

[Hor96] 'Personal information', Mr Hormam, written answers, Hansard (29 Jan 1996) p 528

[HRM93] "RMs need to safeguard computerised patient records to protect hospitals", Hospital Risk Management 1993 no 9 pp 129–140

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

[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 the RCGP, 1988

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

[Lan95] "Proposed Confidentiality Law Angers Canadians", The Lancet (16 December 1995) p 1618

[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 PHCSG Conference, BCS, pp 36–42

[MS95] " 'Soundex' codes of surnames provide confidentiality and accuracy in a national HIV database", JY Mortimer, JA Salathiel, Communicable Disease Report v 5 no 12 (10 Nov 1995) pp R183–R186

[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

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

[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

[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

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

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

[Sim96] GJ Simmons, *personal communication, 1996*

[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

[ISM95] "Telephone stalker has access to confidential records", Information Security Monitor (September 1995) p 2

[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

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

[WHFG92] "The Active Badge Location System", Roy Want, Andy Hopper, Veronica Falcao, Jonathon Gibbons, in ACM Transactions on Information Systems v 10 no 1 (January 1992) pp 91–102