Problems with the NHS Cryptography Strategy

Ross Anderson

Cambridge University Computer Laboratory

1 Executive Summary

Clinical data networking has the potential to improve patient care in various ways. Electronic referrals could cut hospital administration times; electronic discharge letters could help GPs provide better follow-up treatment; electronic pathology and radiology reports could cut the delays, errors and paperwork associated with paper systems; and telemedicine could give GPs and patients access to a wider range of specialists while cutting travelling cost and inconvenience.

One of the main obstacles to achieving these benefits is concern among both clinical professionals and patients about both the safety and privacy of electronic medical information. Errors in laboratory reports or referral letters could lead to incorrect treatment and cause harm or even death; putting clinical databases on-line could lead to breaches of privacy; and the move from paper to electronic records will introduce new medico-legal complexities when these records have to be relied on in evidence.

Some networks are made physically secure, using techniques such as armoured pressurised cables. However this is very expensive, and gives a relatively low level of assurance. It is usually more economic to protect electronic information on networks using cryptography. This includes two basic techniques:

digital signatures can help assure the safety aspects of electronic messages. They bind a message to its originator and can detect any alteration to the message after it has been signed;

encryption can help assure the privacy of a message by scrambling it under a key or keys whose corresponding decryption keys are only available to the message's authorised recipients.

The NHS Executive's Information Management group (IMG) understood the promise of cryptography and published a strategy on it in April 1996 [85]. Following questions raised by the BMA at a meeting in May [12], some clarifications were issued in [86], and in meetings with the BMA in June and July. There are now three encryption pilots underway, of which two are administratively directed and under IMG control while the third is more attuned to clinical needs. The IMG documents, together with the two pilots under IMG control, provide a coherent picture of a strategy for cryptography in the NHS.

This paper examines the documents and the strategy they purvey in order to assess its acceptability to the clinical professions. A number of problems have become apparent.

1. The IMG strategy pays insufficient attention to the safety aspects of clinical messaging. Digital signatures are at least as important as encryption and are the subject of European standards activity; but the strategy concentrates on encryption first, with signatures to be added later.

2. From the privacy viewpoint, the strategy uses an incorrect threat model: that most attacks will come from outsiders rather than insiders. This is based on IMG security policy documents that have now been superseded in the IMG's own thinking.

3. Whether the protection priority is encryption or signature, a means of managing encryption keys must be provided and its structure determines the trust relationships in the resulting system. It is prudent practice to make electronic trust relationships mirror those in the underlying world of professional practice, and this principle is agreed between the BMA and the Department of Health [60]. Yet the strategy seeks to replace the current collegiate trust in medical practice with a centralised structure: a small number of 'trusted third parties' are supposed to manage all the keys in the system. This will impose significant overheads and obstruct normal working practices, as well as being seen as a serious assault on professional independence.

4. The estimated costs are not credible. In addition to assuming that key management can be centralised cheaply, the strategy ignores the costs of standards development, evaluation, system migration and training. On the other hand, the cost of cryptographic software is set at four times the market price. The proffered explanation, that development and training costs are bundled into the licence fee, appears to assume a monopoly that would undermine the other costing assumptions.

5. The strategy is equivocal on the subject of key escrow. It initially appeared to be an attempt to have the NHS adopt a GCHQ escrowed encryption protocol [24] (or something similar); when asked at the June meeting for references to suitable protocols, the GCHQ protocol's precursor [45] was cited, and the writer interpreted the strategy document to mean that escrow will be a requirement ([85] p 58). This interpretation was expressly denied [86]. Yet at least one of the pilots under IMG control has the demonstration of escrow as one of its principal goals, and progress on the pilot under clinical control has been held up by an IMG demand that keys be generated centrally.

6. The strategy of setting up escrowed encryption keys first and then using them to distribute digital signature keys means that a doctor's signature key could be recovered without his knowledge and used to forge apparently valid signatures. This is unacceptable on both safety and medico-legal grounds.

7. The encryption mechanisms proposed are unnecessarily weak by the standards of current commercial cryptography and are unlikely to inspire public or professional confidence.

8. The protocol and certification mechanisms that the strategy appears to be recommending to the NHS also suffer from various problems, ranging from year 2000 compliance through difficulties with replacing compromised keys, protecting security labels on data, and conformance with European standards. They also appear to lack the durability required for documents that might be used in evidence many years after being signed.

9. Finally, the IMG's main cryptography strategy document contains a significant number of fundamental technical errors.

The heart of the matter is that the IMG cryptography strategy appears to encourage the NHS to adopt protection mechanisms very similar to those designed by CESG (a department of GCHQ) to protect government electronic mail. This is admitted by Andrew Saunders, the director of CESG and a main board director of GCHQ, in [75].

However, the GCHQ protocol mechanisms have different goals from those of the clinical professions. They attempt to keep a message between two officials secret from third parties, but available to both their superiors (and to the police and intelligence services) by ensuring that each official's departmental security officer has a spare copy of the key used to encrypt it. Furthermore, the key used to 'sign' the message is also available to authority. Thus if an embarrassing message is leaked, it is always possible to claim that it was forged — perhaps by the very security officer whose negligence permitted the leak. We can summarise this functionality as 'secrecy with plausible deniability'.

Clinical professionals, on the other hand, require safety and privacy. The origin and content of messages should be indisputable, whether for the purposes of immediate clinical decision making or for litigation many years later. Patient privacy must also be respected; GMC guidelines mean that the patient (or the clinical professional acting as his advocate) must have control over who can read his records [38], and this in turn means supporting access control mechanisms that respect the organisational and professional realities of healthcare. These safety and privacy goals are both incompatible with the GCHQ approach to securing electronic mail.

The GCHQ approved protocol suffers from further problems [14], which it shares with the NHS Executive's strategy insofar as this has been spelt out in detail. For example, both assume that control will be centralised, which is not only in conflict with professional autonomy but highly likely to be impractical in the NHS. Officials have been informed repeatedly by the BMA that the GCHQ approach is unacceptable. Yet despite repeated official denials that escrow and centralisation are the objectives of the NHSE strategy, we are concerned that these very aspects of the GCHQ approach to protection are being implemented in the Teesside pilot.

In our view, there is no realistic prospect that the current strategy could win the trust of patients and professionals, and thus enable the many potential benefits of clinical messaging to be realised.

We will now discuss the above points in detail. In sections 2–9 below, we will discuss the points made in paragraphs 2–9 above. The point in paragraph 1, of the relative importance of encryption and signature, will be discussed in section 6 below as it turns out to be closely related with point 6.

2 Threat Model

When considering the implications of networking clinical systems together, IMG originally assumed that the main additional threat would be external 'hackers' [53, 54, 55, 56, 57, 58, 59].

The BMA first pointed out in [6] that this was mistaken, and that the great majority of incidents would probably be due to abuse of authorised access by insiders. Further BMA documents discuss the issue in more detail [8, 9, 10]. The view that most attacks will be internal rather than external was confirmed in a study commissioned by IMG from the CCTA which reckoned that only 6% of attacks would be due to outsiders [58]. It has recently been further confirmed by MI5, whose UNIRAS unit now handles the reporting of computer security incidents in civil government; the head of UNIRAS stated that only 2% of attacks on UK government systems in 1994/5 were due to outsiders [33].

The head of IMG has since conceded that insiders pose the main threat, and that the former threat model no longer reflects government thinking [73]. This change of view needs to be incorporated into the cryptography strategy, which currently maintains that '*current concerns about NWN security are centred on two main threats: the possibility of unauthorised individuals logging on to the network (and) the possibility of eavesdropping on network traffic*' (p 10).

There is a serious engineering issue here. One of the most important tasks in systems engineering is to maintain synchrony between the many documents involved in a typical information systems project — such as the user requirements document, the hazard analysis, the performance benchmarks, the technical specification, the source code, the manuals and the operators' training material. When this vital synchrony is lost — as, for example, when modifications are made to a system without this being reflected in the specification and manuals — then problems can be expected to occur.

After IMG accepted that their initial threat model was wrong, the BMA asked them to issue an amended set of policy documents, starting with the 'NWN Threats and Vulnerabilities' [54], then leading on to a top level information systems security policy to replace [53], a data security policy to replace [55], and so on through [56, 57, 58] to the reference manual [59].

However, IMG refused to update their documents. This is unacceptably poor systems engineering practice; it has led to the adoption of a cryptography policy founded on assumptions that officials claim to have abandoned. Under the circumstances, the writers of this policy cannot be held wholly to blame for the erroneous direction of their document.

Methodological confusion can also be seen in the NWCS pilot, which appears to have been substantially completed before a risk analysis was attempted, and it was discovered that the CRAMM methodology used for this could not cope with encryption [62]. A risk analysis should precede the detailed technical specification of a system, let alone its implementation.

The real hazards to the safety and privacy of clinical information are discussed in [9]. They are the risk that clinical messages may become corrupted, whether accidentally or maliciously, and the lack of adequate access control mechanisms in heterogeneous distributed systems. This analysis leads naturally to the BMA's security policy [8] which describes the requirements for cryptography: to protect the safety of clinical messages (using digital signatures), and to protect their privacy (using encryption to support access control). A secondary objective is to assure the medical-legal reliability of electronic health records.

The prevention of eavesdropping by third parties is of minor importance. If it were the primary objective, then the expenditure of over ten million pounds could not be justified, as there is so far no known NHS case of wiretapping leading either to a financial loss or a privacy compromise. (There are however a number of cases where information has been corrupted or misrouted.)

These considerations have been at the forefront of debate since the BMA policy was published in January 1996. Yet the IMG strategy ignores them, and repeats the view that concern focuses on outsiders (as in the statement on unauthorised logon and eavesdropping quoted above).

After this inconsistency was pointed out to officials in May, a second document claimed that their *'work has not assumed any particular threat model'* ([86], p 2). This is hard to reconcile with the statement quoted above; with the statement on p 24 of the strategy that encryption guidance *'would have to be written in the context of the current Data Networking Security Policy, Guidelines and Codes of Connection'*; and the statement on p 42 that *'care must be taken to ensure that (cryptography) does not interfere with or run across the desired access control systems'*. It is also contrary to good cryptographic engineering practice: the design of cipher systems should take into consideration the environment in which they operate and the kinds of attack that may be expected ([18], chapter 8).

The requirement that encryption should support access control was conceded by the IMG at the June meeting. We will return later to the technical aspects of linking encryption with access control.

There appears to be further confusion about whether cryptography should protect traffic on local as well as wide area networks. On p 8 it is said that eavesdropping and tampering on LANs should be dealt with using different controls, but p 42 claims that encryption *'can be used to address security needs which exist outside the NWN, including the encryption of data transmitted over other networks such as local LANs'*.

In addition, the discussion of the threat model fluctuates somewhat over pages 10–12. On page 10 we learn that after 'strong authentication' is in place, *'Addressing the remaining eavesdropping threat is the subject of the present study'*. On the next page the threat model includes *'repudiation of a message sent earlier by the sender (for example, the potential disowning of a negligent pathology result)'*. Yet by p 12 the conclusion has been reached that *'confidentiality is seen to be the widest requirement. This justifies the requested approach which was of focusing first on the encryption solution and then examining ways by which the encryption solution could be extended'*. Whether the request for this approach was made by IMG or by GCHQ is not stated. We shall return to the importance of the order of encryption and signature, and its implications for policy, in section 6.

Lack of clarity about threats and protection goals extends to the NWCS pilot. The ethical objection to NWCS is that it creates a large centralised database, outside clinical control, that contains highly sensitive information (such as treatments for HIV and terminations of pregnancy), that identifies patients fully, and that contains information on most of the population. The US experience suggests that once such a database has been created, there will be inexorable pressures for legitimised access by all kinds of interests, starting with researchers and proceeding via policemen and social security fraud investigators to insurance companies and credit reference agencies. The encryption of data enroute to such a facility is irrelevant to these concerns.

Furthermore, where a system has a star topology (as with the NWCS where many healthcare providers feed data to one contractor), there is no need for elaborate public key management schemes; and where the Secretary of State owns all the data anyway, there is no visible purpose served by key escrow (which we will discuss more fully in the next section).

3 Trust Structure

Probably the most important architectural issue is the structure of trust. By this we mean how we organise the mechanisms whereby a principal finds out that a particular key may be used for some given purpose. It is well known that trust structures in the electronic world should mirror those in the underlying business or professional practice [71], and as we noted above, this has been agreed in principle between the BMA and the NHS. In this section we would like to consider in more detail precisely what this means, and what it implies for the NHSE's proposed cryptography strategy.

To take a simple example, private investigators obtain personal health information unlawfully by telephoning either the GP or the health authority of the target of their investigation, and claiming to be a hospital doctor involved in the target's emergency care. Such enquiries pose a dilemma: should the GP reveal sensitive information to an unidentified caller and risk a breach of privacy, or withhold it and risk that the patient will come to harm?

This is a live problem; one health authority that has taken the trouble to train staff to use careful telephone procedures [7] is now detecting about thirty false pretext telephone calls per week [41]. This indicates that over a hundred thousand such incidents take place nationwide per annum; most of them go undetected because of the cost and inconvenience of maintaining a high level of staff awareness and telephone discipline.

A cheaper and more convenient solution is possible with secure electronic mail. A hospital doctor who needs information can authenticate himself by digitally signing the request; the GP can check the signature and then encrypt the response using an encryption key that was sent with the signed request, and whose corresponding decryption key is known to the hospital doctor. The problem is now how the GP should recognise the signature as belonging to a bona fide hospital doctor.

The use of 'certification authorities' (CAs) is one established way to manage cryptographic keys. An example is given by the new SET protocols for credit card transactions over the Internet; each bank (or group of banks) will establish a service to which customers can send their encryption keys (and the verification keys corresponding to their digital signature keys) and have them certified. This certification has the effect of binding a customer's keys to his or her account number.

But the solution proposed by IMG is to use a 'trusted third party' (TTP), or a small number of them, rather than CAs. TTPs differ from CAs in that they retain copies of the private decryption keys in order to provide government access to encrypted traffic if required. They are an initiative of the US government whose objective is to bring the civil use of cryptography under the control of the US government and its allies.

The rationale offered for TTP services is usually as follows. Cryptography is about secrecy; secret communications could help criminals escape detection; so the government must have some means of decrypting everybody's traffic; and, as a quid pro quo, the government can provide a means for users of the Internet to identify each other. So the proposed solution is that TTPs certify users' public encryption keys, while retaining copies of their private decryption keys for use by police and other government agencies.

In the UK, the TTP initiative was launched by the DTI last year, and at a meeting called by the Ministry of Defence on the 27th June to publicise it, a DTI spokesman stated that a condition of TTP licensing would be that all keys used for confidentiality services should be escrowed, i.e. copies would be retained by the TTP and made available to the authorities on demand.

We will discuss the privacy, safety and medico-legal aspects of escrow in more detail in section 5 below. The point we wish to make here is that regardless of whether we are dealing with a TTP or a CA, the number of such services has a significant impact on the usability and thus the cost of the system.

The NHSE's main strategy document strongly favour a single TTP, or a small number of them, for the whole of the NHS. Although it does allow for the possibility that there might be more than one such service, the thrust of the argument assumes that there will be one, and the detailed costings supplied on pp 29 ff imply that there will be at most two (four full time staff equivalent per TTP, and a total budget of eight).

This represents a massive centralisation of the current structures of trust in medicine, which as pointed out in [8] are collegiate rather than bureaucratic.

Signing a doctor's key is the electronic equivalent of recognising that a doctors is in fact a registered medical practitioner. This function is, of course, carried out by the General Medical Council which holds the register of all persons qualified and licensed to practice medicine. Consequently, any government move to set up a national medical TTP could be seen as subsuming one of the essential independent roles of the General Medical Council.

After this point was raised in discussions with officials in May, the supplementary strategy document claimed that the GMC had been suggested in internal discussions as the natural location of the TTP ([86] p 9). We understand that it was argued to the GMC that so long as they had 'control' of the TTP it did not really matter who actually operated it. These arguments are of great concern; if they prevail, the effect would be to put key management under nominal clinical control but with the day to day administration in the hands of the NHS Executive, or possibly even a privatised service provider.

The desire of the UK and other governments to centralise trust appears tightly bound up with the key escrow agenda. The GCHQ secure email document declares and intention to centralise all state sector TTPs at the department level, except where geographic dispersal prevents this, and have GCHQ personnel operate them, at least initially [24]. The US National Security Agency, which is driving the whole TTP programme worldwide, attempted to specify ISAKMP — the future key management protocol for the Internet — in such a way that there could be at most 65536 certification authorities or trusted third parties in the world. This attempt has now failed. The reason for the attempted centralisation is clear — governments feel that a small number of key management centres would be easier for them to control.

However, even if it had been the intention of government to have professional registration bodies act as TTPs without external assistance, then surely provision would have been made in the strategy for more than two of them. There would have to be at least one (and for reliability reasons preferably two) for each registrable professional group, i.e. two for doctors, two for nurses, two for dentists, and so on.

But in the Teesside encryption pilot, there will be a single TTP operated initially by BT [81]; their key management scheme is to be adapted from that used in the Clearing pilot. Centralisation appears to be assumed.

Experience from the banking sector has shown that in organisations with decentralised personnel management, centralising trust is difficult, expensive and fragile. The writer advised a bank with 25,000 employees, managed through seven regional personnel offices, trying to administer mainframe passwords at a central site. With thirty staff and much message passing to and from the regions, the task was just about feasible, but was an inefficient way of running things and coped poorly with emergencies. Imposing such a solution on about a million staff in the thousands of NHS and supporting organisations would result in significant risks, costs and service degradation.

As a general principle, keys should be managed where the personnel management is done, and especially so where keys are used to support access control. In New Zealand, for example, a proposal to have doctors' keys managed by officials in the local district hospitals turned out to be impractical. The overhead of calling the hospital for keys to be issued or revoked with every temporary staff change was not supportable. It is now proposed that keys there should be managed at the practice level [40]. In the UK, with some 12,000 general practices, hospitals and community care facilities, centralised key management is even less likely to be workable.

So for practical reasons, cryptographic keys will have to be managed at the provider level. This means developing CA software that can be used by general practices to certify keys for their staff. A consequence of this will be that access by the police and by other government bodies such as social services and benefit agencies to clinical records will involve the knowledge and cooperation of at least one member of clinical staff at the provider — as is the case at present with manual records and standalone computer systems. This will accord with the agreement that future trust structures should reflect existing ones.

It is unthinkable that the Government would wish to change this arrangement and permit access to practice databases without the partners being aware of the intrusion, as this would represent a very significant change in trust relationships and would create extensive public alarm. The EC Data Protection Directive would have to be considered, as would the Data Protection Registrar's recent comments on key escrow to the effect that under section 8.2 of the European Convention on Human Rights, the invasion of privacy is permissible only where there is a 'pressing social need' [35].

No reasoned argument has been made for changing the current ethical provisions for access to medical information which currently give patients control over who has access to their records. The BMA's security principles are designed to give patients authority over the data flows originating from their records. The BMA will resist any moves that undermine this approach.

4 Cost Estimates

The NHSE's cryptography strategy includes a number of fairly detailed costings. However, these lack credibility for a number of reasons. We have already mentioned the lack of provision for software integration of hardware crypto functions under 'Global Cost Model 1' (p 38); there are many more omissions and dubious assumptions.

For example, the strategy assumes that there will be one, or at most two, ‘trusted third party’ key management centres, with all the key management tasks performed by eight full time staff equivalent. Assuming two million personnel changes in the NHS each year (hirings, firings, and moves), this would entail each member of staff administering about a hundred key changes per hour, which is perhaps an order of magnitude more than what could be achieved with even fairly cursory checking of credentials and entitlements. So even under IMG assumptions, the number of staff needed would be more like a hundred.

But these assumptions are suspect on a number of grounds. According to the strategy, the TTP centre or centres would not be staffed at evenings and weekends. So if a key were compromised on Friday evening, it could not be revoked until Monday — leaving the attacker a whole weekend to attack systems, steal personal health information, forge prescriptions, and so on. This is not acceptable. Even in the banking sector, where lives are not usually at risk, a 24-hour revocation facility is considered essential; a bank in Germany that did not provide one was held to be liable for card fraud [84].

Furthermore, as discussed above, there would be an enormous overhead in administering a centralised key management system. A GP who hired a locum for the day would presumably have to call the key management centre to get key material issued so that the locum could be added to the relevant access control lists. Assuming that the phone were answered, and assuming that there were no hitches, and assuming that ten requests could be handled per hour (which would be brisk), then the GP would still have spent perhaps ten minutes on the operation. This has also not been budgeted for.

In reality, as discussed above, there will be not one certification authority but many. The strategy takes no account of this in its costings. Nor does it attempt to make explore the cheaper functional alternatives.

A second cost issue is standards development. The report does not budget for this either. The issue having been raised, the supplementary strategy document claims that standards costs had been included in the £100 per user estimated licence fee for encryption software. This price is indeed high, as the supplier of the encryption software for GP-provider links is charging only £25 and PGP can be licensed for SFr30 (under £20); so the explanation appears plausible. However it was not made explicit in the strategy document, and in any case raises the question of whether a monopoly provision of cryptographic services is likely or indeed necessary for the standards work to be done. The long term cost implications for the NHS of such a monopoly also need to be considered.

A third missing cost is training. On p 30, the strategy document asserts that cryptographic security can be implemented without significant expenditure on user training, and on p 53 that the entire operation of cryptographic security can be made invisible to the user. This contradicts both commonsense and a talk given by IMG’s senior adviser at ESORICS 94 in Brighton.

A fourth missing cost is that of system migration. Introducing new ways of doing things is never free; yet the strategy’s line is that ‘much of the above project manpower could be provided by the NHS from its own staff resources’. In other words, the stated costs apply to items such as equipment purchase and security consultancy. The strategy appears to ignore the fact that a large part of NHS computing is contracted out, and that 'own staff resources' are meagre. Even where they exist, they still have to be paid for.

A fifth missing cost is evaluation. The report contains no budgetary provision for the assessment and testing of cryptographic products or implementations; it simply contains items such as *'HISS — 6 systems at £75K'* (p 39). But evaluation of these changes is necessary and expensive. The report suggests the use of ITSEC [43]; yet an ITSEC evaluation costs in the range of £100,000 to £1,000,000 depending on the product's complexity. At only £100,000 per system, evaluation will more than double the figures in 'Global Cost Model 2'.

When tackled on this point in the June meeting, IMG agreed that evaluation costs had been omitted, stating that ITSEC evaluation was in their view an expensive and unnecessary piece of bureaucracy. This view would not be supported by the director of CESG, according to whom almost all products submitted for ITSEC evaluation have had serious vulnerabilities found and corrected, with some products having as many as twenty [74]. In our experience too, the large majority of system security failures result from errors in implementation or operation that are discovered by chance and exploited in an opportunist way [3, 13]. Reducing the incidence and severity of these blunders is the purpose and function of evaluation. Ignoring evaluation will not only lead to incorrect and inadequate budgeting, but is also at odds with both government policy and best practice.

For all of the above reasons, we believe that the IMG's cost estimates are not robust and will not stand up to the Public Accounts Committee's scrutiny. A realistic strategy must recognise the need to defend cost estimates from this standpoint.

5 Key Escrow

As mentioned above, the question of 'trusted third parties' is the main policy issue before the cryptologic community at present, and this is part of the wider issue of key escrow. Should cryptographic keys be held in 'escrow', which means using some mechanism (that might involve third parties) so that the cryptography becomes transparent to law enforcement and other government agencies?

The US government has made four significant attempts so far to get users of cryptography to adopt key escrow of one kind or another. The TTP initiative is merely the latest of these; the history of this project in the USA, and the national policy concerns behind it, are described in a recent report from the US National Academy of Sciences [27].

As the NHS's strategy document offered advice on the ramifications of adopting cryptography, its views on escrow would have been of great interest. But 'escrow' is not discussed; instead, a euphemism ('key recovery') is used, and on p 58 the strategy document states that *'NHS should consider ... whether it wishes to implement the key recovery capability within it or not'*. The implication that I, as a security professional, drew from this wording is that escrow will be implemented and that the only question is whether the keys will be held by the NHS or by another government department such as GCHQ.

In response to this, the statement [86] was made that *'Nowhere do we state or imply that either doctors' encryption keys or signing keys should be escrowed'*. In support of this we are pointed to another excerpt from the original strategy which mentions the possibility of omitting key escrow but designing the system so that it can be added later. This gives rise to almost as much concern. The writer of the strategy has since stated categorically at a public meeting [51] that medical keys would not be escrowed, and a senior official stated at the June meeting that there would be no key escrow in the pilot [52].

However the encryption pilot in Teesside appears to focus on demonstrating the capability to do key recovery despite officially having a broader scope than this [36]. The key management scheme involves RSA keypairs being generated by a single TTP run by BT Syntegra; both the signed public key (in X.509 certificate format), and the private key, are then sent to the user [15]. Thus private keys, for both signature and encryption, are available to the TTP for escrow. This key management system is derived from the NHS Clearing service pilot [81], so we can assume that the Clearing pilot also uses escrow.

Finally, the GPPL pilot was held up for a while by an IMG demand that doctors' signing keys be generated centrally, and is still handicapped by a lack of clarity about what key management mechanisms will be acceptable to the NHS Executive.

We note that the NHSE document says (p 45) that the strategy for cryptographic algorithms and key management will have to be confirmed before the pilots commence. But even though the pilots are well underway, the strategy for cryptographic algorithms and key management would not appear to have been confirmed.

The following selected series of events and developments may be indicative of changing attitudes to cryptography in the UK government:

- in 1994, John Major stated in a parliamentary written reply to David Shaw MP that no further restrictions on encryption were envisaged in the UK;

- by mid 1995, a proposed key escrow architecture developed by Royal Holloway appeared at an Australian conference [45], and a senior GCHQ man was on hand to support a proposed Australian escrow policy [65]. After a flaw in the protocol was pointed out at this conference, an amended version of the scheme was published [46];

- following press reports of imminent European introduction of key escrow [66] [83], the EU's crypto policy body, SOGIS (the senior officials' group, information security) asked in November 1995 for a Council decision to spend 25 million ECUs to establish a European network of TTPs [79];

- by February 1996 the trade press was reporting that the US — together with France and Britain — was pushing the other countries in the OECD to adopt a common line on the use of TTPs for key escrow [26];
- in April 1996, the NHS encryption strategy was published and referred to '*a number of necessary National Policy decisions were ... by CESG in the last months of 1995'* (p 57);
- on the 10th June the government announced that it would set up a licensing scheme for trusted third party services [69] [70];
- when this policy was presented at a public meeting on the 27th June, a government spokesman said that escrowing of confidentiality keys would be mandatory [42]. He also made clear that the controls would focus on small-to medium enterprises and individuals rather than on large companies. Large companies would be exempt; the rationale given for this was that organisations with enough assets would be responsive to warrants.

We sincerely hope that it is not the position of government that hospitals and general practices would not be responsive to warrants. Yet despite clear denials that key escrow was intended or would be introduced, persistent attempts are still being made to introduce it via the crypto pilots.

This is unlikely to win the confidence of either public or professions. It also breaches the agreement that the electronic trust structure should reflect existing practice — under which to medical records by policemen, social workers, benefit agency employees and other government servants involves the knowledge and agreement of the clinician having responsibility for the patient, and therefore with the knowledge and consent of the patient unless the law specifically provides otherwise.

6 Order of introducing encryption and signature

The IMG strategy claims on p 13 that authentication and signature are not only a secondary requirement, but one which should be tackled after encryption is in place. On p 18, it considers signature to be only a 'possible requirement'. If signature is indeed required, then the system 'could be extended cost-effectively to support these additional security services' (p 13). The decision to introduce encryption first and signature second is repeated elsewhere (e.g., p 59). This closely follows GCHQ's strategy for secure email [24].

The IMG/GCHQ approach is wrong and dangerous. If escrowed encryption is used to distribute signature keys, then these signature keys become known to (or at least discoverable by) the authorities. So doctors' signatures could be forged by authority, or by third parties to whom the signature or escrow keys had been leaked. Safety would be at risk and the evidential force of digital signatures would be greatly reduced.

What should happen is that a clinician should generate a pair of keys — a private signature key, and a public signature verification key — and send the public key to a certification authority to be signed. But the GCHQ protocol does not support a mechanism for transporting keys from the user to authority. There is merely a 'token' that is used to convey both private and public keys from authority to the user. Despite IMG's many assurances that neither encryption nor signature keys will be escrowed, and an assurance in [86] that *'current discussion about escrowing keys relates only to encryption keys and not to signature keys'*, the Teesside pilot takes the GCHQ approach. Signature keys are generated by the TTP (in that case, BT) and then copies of them are given to doctors. BT thus acquires the capability to forge doctors' signatures, thus redoubling the threat of insider access that encryption mechanisms should be helping to combat.

Even if privacy keys are eventually required by law to be escrowed, the keys used for digital signature and for other forms of authentication must be treated differently. The government clearly has difficulty understanding this point, despite its being raised in numerous fora by the BMA and others. So it is worth explaining explicitly.

The stated purpose of key escrow is to enable government employees to monitor the contents of encrypted traffic (and, in some escrow schemes, to facilitate data recovery if users lose or forget their keys). Its stated purpose does not include allowing government employees to create forged legal documents. It would be highly undesirable if they were able to use this access to forge contracts or purchase orders: the scope for insider fraud and conspiracy to pervert the course of justice would be immense.

Any police officer will appreciate that if he can get copies of a suspect's bank statements, then perhaps he can use them in evidence; but if he can tracelessly forge cheques, then the suspect will argue that all the evidence was simply forged by the police. So if there is any possibility that a digital signature might be needed as evidence, then the key used to create it must not be escrowed.

In fact, we would go further than this: keys which are used only for authentication (rather than non-repudiation) should not be escrowed either.

For example, suppose that some piece of equipment (e.g. a telephone exchange, or a ventilator in an ICU) is controlled remotely, and message authentication codes are used to protect the integrity of control messages. Even if these messages are not retained for the purposes of evidence, it is clearly important to distinguish between authorising a law enforcement officer to monitor what is going on and authorising him to operate the equipment. If authentication keys are escrowed, then the ability to monitor and the ability to create seemingly authentic control messages become inseparable: this is almost certainly a bad thing. Returning to the medical context, it is unlikely that either doctors or patients would be happy with a system that allowed the police to forge prescriptions, or Pensions Agency officials to assume control of life support equipment.

We doubt that any government minister who understands this danger would wish to expose himself and his officials in such a way.

In such applications, we need an infrastructure of signature keys that is as trustworthy as we can make it. Bootstrapping the trust structure from a system of escrowed privacy keys is completely unacceptable on both safety and medico-legal grounds.

7 Strength of Encryption Mechanisms

According to present government guidelines, information whose compromise could threaten human life directly must be classified 'Secret'; there are certainly medical records that fall in this category. Yet the Red Pike algorithm which the NHS strategy recommends (and which is being used in the Teesside pilot) is only evaluated to 'Restricted' — two whole grades below. As its keys are only 64 bits long, it is already vulnerable to keysearch by large organisations, and will be open to attack by individuals within 10–15 years at most. The NHS Executive's own documents estimate the life of Red Pike at only five years. In addition, it does not command the confidence of the cryptographic community. It is worth while discussing these points in more detail.

7.1 Capability of attackers

As already mentioned, the choice of protection mechanisms must be informed by an analysis of the attacks that are likely to be made, and the length of time for which the information must be protected (the 'cover time'). This is a basic security engineering principle, and one the IMG's senior adviser discusses in his book [18]. We will discuss in turn the capability of opponents, and the cover time appropriate for medical data.

The IMG strategy states that encryption using single-round DES, with 56 bit keys, is not at present enough. The key length of DES is a well discussed problem that was first raised by Diffie and Hellman shortly after that algorithm was proposed as a standard [31]. If we attack DES by trying all the possible keys, then we will on average have to test 2^{55} of them before we find the right one. Now $2^{55} = 2^{20}$ x 2^{20} x 2^{15}, and 2^{20} is about a million; so a custom machine with a million processors, each capable of testing a million keys a second, could break a DES key in less than a day.

A 1993 analysis, assuming the faster chips available by then, argues that a machine built for $1m could break a DES key in about three hours; and it can be expected that within the next few months, the first public announcement of a successful DES keysearch will be made. A US company, in order to promote the use of a block cipher called RC5, has offered a number of public rewards for successful keysearch against ciphers of varying keylengths. A 40 bit key was found in 3.5 hours, and a 48 bit key in 13 days — in both cases using software on a number of machines in parallel. An attempt on their 56 bit DES challenge key is now getting underway, and a recent technical innovation allows DES keys to be searched 3–5 times more quickly than was previously the case [16].

7.2 Nature of attackers

The IMG strategy is not completely clear about the opponents from whom the confidentiality of personal health information is to be protected. According to the IBM taxonomy, these opponents could be naive outsiders (such as undergraduate hackers at American universities), sophisticated insiders (such as the IT staff of an NHS supplier attacking a system to obtain competitive information) and funded organisations (such as a foreign national intelligence organisation seeking to obtain personal health information on MPs and senior officials for the purposes of blackmail) [1]. If the aim to to prevent 'anybody' getting hold of information (as stated on page 12), then one must assume that the opponents include funded organisations.

The best assessment of national intelligence agency capabilities that we can make is as follows:

- 56 bit keys are routinely found and have been for years;
- 64 bit keys can be found with some effort;
- 80 bit keys could be found using unreasonable effort (i.e., comparable to the Apollo project)

So an assumption that opponents could include national intelligence agencies is not compatible with a 64 bit key length. Assuming a well designed algorithm, a 64 bit key could probably not at present be found by an average individual attacker using keysearch. Such an attacker might just be able to find a 56 bit key if he were to sell his house to pay for the equipment, so the strategy's claim that 56 bits are too few and 64 bits are enough, is consistent with a threat model consisting of average individuals only.

The NHS Executive appears to be already aware of this problem. According to the NWCS pilot board minutes, it was estimated that five years was the operational life of Red Pike [61]. It does not seem prudent to deploy a system that will on its own advocates' admission be obsolete by the time it is fully fielded.

7.3 Cover time

However, even if we assume that Red Pike will be replaced by something better in 2002 (and that this can be done without undue expense or disruption to fielded systems), it still does not follow that 64 bit keys are adequate to protect medical traffic today. The reason for this is that some of today's messages must be protected for a long time.

According to Moore's law, computing capability doubles every eighteen months. Thus the extra eight bits of protection afforded by using 64 bit rather than 56 bit keys translates into an extra twelve years of protection. Otherwise put, the capabilities of individuals lag slightly more than a decade behind those of governments; and if governments can already find 64 bit keys today, then individual attackers will be in the same position in 2010 at the latest.

But how long does medical data need to be protected?

Consider the effects of a revelation being made today that a senior Cabinet Minister had been treated for venereal disease while a teenager. This could clearly do harm, and clinicians seek to protect patients from such harm where practical. The timespan of protection required could, in such cases, be about a human lifespan — say 70 years. (It could be even longer in the case of genetic information.)

This translates into key lengths in the range of 112 bits (as offered by the standard variant of triple DES) through to 128 bits (as offered by algorithms such as WAKE, SAFER SK-128 and RC4).

A similar argument to the above can be found in a standard textbook on cryptography [77], which recommends a key length of 128 bits for personal information.

7.4 Cost of strong mechanisms

The cost of implementing this level of protection against keysearch is not significantly different from that of providing much weaker protection. Indeed, the '40-bit' version of RC4 alluded to in the NHS strategy and rightly described as being weak, is actually a perfectly acceptable 128-bit algorithm — but 88 of its key bits are sent in the clear in export versions of many US programs as a condition of export licencing.

There is no technical reason why strong protection should cost any more, and thus the IMG strategy's recommendation that a 64 bit algorithm should be used when 128 bit algorithms are available in the public domain makes no sense from the engineering point of view. It is also unlikely to inspire confidence in the public and the clinical community, once security professionals have the opportunity to spell this out in plain English.

7.5 Confidence in mechanisms

In addition to the 64 bit key length of the Red Pike algorithm, there is the problem that it is unpublished. There is strong and historically justified scepticism in the cryptographic community of secret mechanisms; it was enunciated as long ago as 1883 that the security of a cryptographic system should lie in the choice of the key rather than the obscurity of the algorithm [49], and this principle has stood the test of time [48]. The use of an unpublished algorithm runs directly counter to this.

At the June meeting, the IMG put its confidence in the crypto community's accepting Red Pike because of its provenance. We were highly sceptical of this claim, given the cryptographic community's historical preference for mechanisms that have withstood extensive peer review, and also because of the errors in GCHQ's email security protocol [14]. We therefore tested the IMG's claim by asking the following five questions of the sixty or so people who attended the cryptology and computer security sections of a 1996 Isaac Newton Institute research programme. These delegates had been selected by a programme committee as the most eminent researchers in the field.

1. Which course of action is in your view more likely to win the confidence of the public and the crypto community in the privacy of medical data communications: the UK government proposal to use Red Pike, or the BMA suggestion to choose from among the existing stock of public domain algorithms?
2. Do you consider that the Red Pike algorithm has the confidence of the cryptographic community?
3. Do you consider it likely that the Red Pike algorithm will win this confidence?
4. Do you consider it likely that the Red Pike algorithm will ever win public confidence?
5. Do you consider it prudent for the BMA to accept, sight unseen, an unpublished encryption algorithm which will subsequently be made available in software?

The BMA was unanimously supported on questions (1), (2) and (5). There was strong majority support on the other two; the minority said that once the algorithm was published (deliberately or otherwise) it might gain acceptance if it attracted and resisted serious attack efforts.

The IMG strategy states that the encryption algorithm chosen must be acceptable to the cryptographic community (p 56), and the people canvassed in the above test are the opinion leaders in that community. So it would seem that the use of Red Pike is incompatible with this eminently sensible goal.

8 Protocol Problems

Recently the NHS Executive has been urging providers to ensure that they can deal with the year 2000 problem. Yet the encryption pilots use a mechanism — X.509 certificates — which has only a two digit date field.

X.509 has other problems. It was originally designed an an authentication protocol, with non-repudiation added as an afterthought. As a result, the support that it offers for digital signatures is less than ideal. Its trust model may be thought of as similar to a credit card — certificates have an expiry date (like credit cards) and there is a 'certificate revocation list' of revoked certificates (like the hot card list of a credit card company).

This trust model has a number of limitations, including:

1. X.509 certificates vouch for identities rather than roles. This may be enough for retail commerce, but in healthcare it is common for an individual to have a number of distinct roles: the same person may be simultaneously a patient, a GP, a coroner and a board member of a local NHS trust. Identity certificates are not adequate to support this..

   Roles are also of great value in reducing the cost of administration. If a bank has 50,000 staff and 100 separate applications on its system, and one tried to administer each user's access rights separately, then this would involve managing 5,000,000 bits of security state, which would be expensive and error prone. However, all but a handful of staff can have their access rights completely specified by allocating them to one of a small number of roles, such as 'teller' and 'assistant branch accountant'.

2. X.509 does not support dual control. It will often be desirable for doctors' keys to be signed by more than one certifier; we envisage that a typical doctor's working key would be signed by a long-term personal key that was in turn signed by the GMC, and also by the hospital or general practice where he or she was employed [12]. In this way we can ensure that the medico-legal aspects of electronic trust are made explicit; both the doctor and the provider are liable for actions taken.

   Dual control is also important to prevent the compromise of a single certifier causing the entire system to fail: so long as a certificate still has one trustworthy signature, it can still be used. Such robustness is a vital attribute of medical systems.

3. In X.509, certificate revocation is the sole responsibility of whoever created the certificate in the first place. This is less than optimal in healthcare, where there are good reasons to want some certificates to be issued by one principal and revoked if necessary by others. For example, a hospital doctor's certificate should be able to be revoked by the hospital (in case he leaves or is dismissed), by himself (in case he resigns), and by the GMC (in the event that he is struck off).

4. X.509 offers poor support for signatures that must persist for a long time, such as the signatures on archived medical records and other legal documents that might be the subject of dispute many years after all the relevant certificates have expired.

At the very least, if X.509 is to be used, then health service specific extensions will have to be developed, tested and agreed. But it is prudent to note that although X.509 certificates are being introduced in a number of electronic commerce applications, there is a growing realisation that the standard is inadequate for more general use, and alternatives are being developed. (The front runner at present appears to be Microsoft's SDSI.)

It is also prudent to note that although the computer and communications industries are merging, their standards processes have not. Where there is a conflict, the computer industry standard usually wins. There are several reasons for this. To succeed, a communications standard must be supported by a wide range of computer systems; and the product cycle of computer companies is about fifteen months compared with fifteen years for phone companies.

The NHS already has expensive experience of this conflict. A policy to have X.400 (the phone company standard) adopted for electronic messaging is encountering resistance, as computer companies — and independent users such as GPs — at present see much more relevance to their business needs with SMTP instead. This mistake should not be repeated with certificates.

The IMG strategy inherits from X.509 the lack of support for necessary features such as year 2000 compliance, roles, dual control, and long-lived documents. However the GCHQ protocol, if adopted in its entirety, would introduce other problems too.

Firstly, users' keys are derived from their names; so a user whose key is compromised cannot be issued a new key, but only a copy of the same old compromised key. In addition, security labels (which in the medical context would mean access control lists) are not protected and could thus be manipulated by attackers. This is directly contrary to DTI guidance on protective security marking [29].

Secondly, this fragility extends to the escrow aspects. In the GCHQ proposal (and also in the Teesside and NWCS pilots as far as we can see), a single official has the power to remove all the protection offered by the cryptographic mechanisms. This is not a necessary attribute of escrow systems; the US government's Clipper initiative, for example, uses two escrow agents so that a single rogue official cannot completely break the system [27]. So even if key escrow eventually becomes a legal requirement, it is possible to implement it much more safely than the IMG proposes to. Further problems with the GCHQ protocol are described in [14].

Finally, IMG agreed at the July meeting that the protocols at least would be public. It is of considerable concern that we have been supplied only with project board minutes, not with technical documentation of the protocols that have been implemented in the Teesside and NWCS pilots.

9 Other Credibility Problems

The IMG encryption strategy has been commented on in [12] and at the June and July meetings mentioned above. Two particular areas of criticism were the reasons for the choice of algorithm and a number of specific points which we believe are in error.

9.1 Reasons for algorithm choice

At the June meeting, the BMA asked the IMG whether they had considered encryption algorithms such as SAFER, WAKE and Blowfish. They admitted looking at SAFER but had not heard of WAKE or Blowfish. This causes concern for a number of reasons:

- all three of these algorithms were unveiled at the same conference [2];
- at the conference in question, about a dozen new cryptographic algorithms were presented, only one of which has since been seriously broken [4]. The annual successors to this conference have steadily been adding encryption algorithms to the publicly available stock;
- SAFER, WAKE and Blowfish are all described in the standard undergraduate textbook used at Cambridge and elsewhere [77];
- all three are public domain and have had official source code published;
- at the time the IMG strategy was written, WAKE held the record as the fastest software encryption algorithm on which no shortcut attack was known, and was thus a clear contender for use in medical imaging applications. There is now a faster algorithm [22] but it is a derivative of WAKE. As it was developed by PictureTel, it is likely to be widely used in video applications;
- Blowfish was the subject of extensive publicity, with a major computer journal running a reward for attacks on it [76].

In view of this, we were unable to accept the IMG claim that strong encryption algorithms were not available, which was their declared reason for recommending Red Pike. It was clear to us that the full range of candidate encryption algorithms had not been properly considered. We pointed this out, and also pointed out that triple DES has been in the public domain for years.

When this point was made, it was explained that the choice of algorithm had been constrained by UK export regulations. However, this explanation is not consistent with p 26 of the strategy which talks about the export prospects of products developed according to the strategy with no mention of the export controls issue; it also appears to clash with p 52 where export controls are claimed to be a disadvantage of hardware as opposed to software cryptography.

When we pointed out that UK medical software is not exported, due to Read coding and other features, it was explained that the choice had been constrained by US rather than UK export regulations; that US companies operating in the UK as NHS suppliers, such as AT&T, would not be allowed by the US state department to use available algorithms such as triple DES even in products created by its staff in the UK and sold only in the UK.

As the writer believed that this conflicted with the statement on p 44 where the availability of Red Pike to non-UK systems developers is stated to be require 'more detailed discussions with CESG', he wrote to the minister responsible, Ian Taylor, suggesting that UK crypto policy is being dictated by US requirements. This was strongly denied [82]:

> 'for the record, I would like to stress that HMG Policy is not determined by US Government requirements. The desire to balance the needs of business for strong cryptography, with the requirements of law enforcement, has, and always will be, determined by the United Kingdom national interest.'

We therefore seek clarification of the following two points to determine whether what we believe is implied can be explained, or preferably refuted, with evidence and authority.

Firstly, as the IMG strategy states that 56 bit keys are inadequate for the NHS (p 55), and as 56 bit crypto is the maximum that US companies can export (even with escrow) we need to know how a supplier subject to US export controls could possibly comply with the strategy.

The implication could be drawn that some deal might be done between the British and American governments that would enable AT&T to sell products in the UK using a 64 bit algorithm provided that the algorithm was Red Pike. (This is not entirely clear; on p 57, the strategy states that *'within the last 12 months, the position has changed. CESG has responded to the situation ... by developing an algorithm known as Red Pike.'* One might infer that the action of CESG that remedied the alleged algorithm shortage was not the negotiation of a deal with the US government, but the creation of Red Pike.)

Secondly, we need to know why the UK government can prevail on the US government to give AT&T a dispensation for Red Pike, but could not also obtain a dispensation for the use of a public domain algorithm of US design such as triple-DES.

There are a number of other points of apparent confusion in the arguments that the strategy presents for the adoption of Red Pike:

- it is stated on p 61 that DES is only available to protect financial traffic. This is incorrect; the DTI has stated categorically that DES (and for that matter triple-DES) are unrestricted in domestic use [42]. Indeed, DES is used in 600,000 gas meters in the UK, as well as garage door openers, road toll tags and many other systems. It comes as a default in most smartcards;
- when we pointed this out at the June meeting, IMG claimed that DES was protected by IBM patents. The relevant patent [32] expired some time ago, and in any case IBM granted a royalty free licence for other people to make and sell equipment using it [77];
- Iit is also stated on p 56, that PGP would be undesirable since (inter alia) its bulk encryption algorithm, IDEA, is subject to patent held by a

Swiss company. On p 22 it criticises algorithms which are *'not sufficiently widely available (being controlled by a single vendor, or under other restrictions).'* Yet the strategy then goes on to propose an algorithm from a single supplier, namely GCHQ, and one that is apparently subject to (unstated) restrictions (p 63).

It is also of interest that the US National Institute of Standards and Technology has initiated a process to find a replacement for DES, which will be called AES (the Advanced Encryption Standard) [63]. This will presumably become an international standard and have an adverse impact on the marketability and credibility of systems using Red Pike.

One further comment in the IMG strategy is revealing. On page 44, it states that further discussions with GCHQ would be needed about *'the possibility of the NHS* **being allowed** *to use alternative algorithms'* (our emphasis). We shall return later to this point.

9.2 Technical and other errors

In addition to its lack of awareness of the cryptographic algorithms available in the open literature, the IMG strategy contains what we believe are further errors, of which the more obvious include the following.

1. The claim that digital signatures can only be generated using asymmetric techniques (p 55 and p 61) is false. Lamport invented signatures based on hash functions before any public key algorithms were published [30, 50], and the military used authenticators based on concatenated encryption by the 1970's [78]. The first of these references is cited in a book written by the IMG's senior adviser ([18] p 393). In recent years, there have been many research results on this topic (see, e.g., [19]).

2. Another claim that by the strategy document that is not supported by recent research is the assertion on p 57 that by *'reason of very large scale, it will be essential for the NHS to have a Key Management infrastructure that includes the use of asymmetric key management methods'*. This is a non sequitur. The scalability of trust structures and mechanisms has recently been widely discussed at research conferences, and it is now well understood that both symmetric and asymmetric key management scale in about the same way. Asymmetric systems may appear at first sight to be cheaper but this is only the case so long as the work undertaken by users and the cost of a certificate revocation service are excluded from the calculation. Once they are included, both approaches cost each of n principals about $k \log n$ effort; no convincing evidence has been adduced to show that the value of the constant k differs significantly from one technology to the other.

3. The claim that there are only two options for key management, namely RSA and Diffie Hellman (p 21) is wrong. Even without reference to recent

research, there is an established product called Kerberos that would be a much cheaper option in the event that the IMG persists in its wish to centralise key management in a small number of online servers. In fact, the security architecture proposed by GCHQ, and which IMG appears to be trying to adopt or adapt, gives no more functionality than Kerberos, but at much greater cost.

4. The claim that TTPs will *'add only a negligible amount of additional network traffic'* is unconvincing. If, for example, the GCHQ protocol were used, then the quantity of extra network traffic could be considerable: whenever two principals in different domains communicate, a reference must be made to both of their TTPs to get a key of the day. IMG may have intended that the whole of the NHS would be in a small number of security domains; they have claimed that professional bodies such as the GMC and the UKCC would be the TTPs [86]. In that case, the use of the GCHQ protocol could mean that a doctor who exchanged messages regularly with a nurse (perhaps in his own practice) would have to send messages to both the GMC and the UKCC to obtain fresh key material at the start of each day.

5. Various safety aspects are neglected. For example, the use of digital signatures to protect the integrity of clinical data will be best supported by integration with clinical EDIFACT messages rather than by enclosing these in an envelope. In the latter case, the signature may be discarded before the message reaches the destination clinical system, and various interactions with X.400 might also have to be considered. Both safety and medico-legal considerations suggest that the digital signature facilities be well integrated with clinical applications rather than hidden invisibly behind an application programming interface, as suggested by the IMG on p 16 and at the June meeting.

6. The interaction with access control will probably require that both encryption and key management are also integrated with EDIFACT messages for similar reasons.

7. The claim by the IMG at the July meeting that messages should be first encrypted and then signed is mistaken, as with many systems the signed ciphertext can be decrypted to a different plaintext using a different key [5]. Messages should first be signed and then encrypted, just as letters are first signed and then sealed in an envelope. Quite apart from the technical security reasons for this, signatures in healthcare often have to be verified by third parties, who will not normally have access to the decryption keys of the original recipient.

8. The strategy document makes a number of claims about PGP on p 56 which are incorrect, such as that it does not integrate well with standard email packages. In fact, it is integrated with far more mail systems than any other encryption program. It is also claimed to be unsuitable for

a large organisation; yet the UK academic network, UKERNA, aims to use it for its security infrastructure. The claim that PGP would need modification by suppliers to support TTPs is also untrue [67].

9. The NHS strategy claims security advantages for unpublished algorithms, but then suggests an algorithm that will be made available in software. As shown by RC2 and RC4, it is only a matter of time before such algorithms are reverse engineered and published. If IMG believes in the alleged advantages of unpublished algorithms, then logically their strategy should have recommended an algorithm available only in tamper resistant hardware such as Skipjack.

10. On p 23, the IMG strategy mentions that independent advice should be taken on the strength of Red Pike. Yet on p 26 it states that the people who evaluate the algorithm have to be acceptable to its owners. Independence can be hard to achieve under these circumstances.

11. The strategy's overall view that security is largely a matter of algorithm choice is mistaken. The engineering aspects are far more important [3] and these are largely neglected in the strategy.

12. The claim on p 50 that cryptography must be done at either the link or the application layer is wrong; session level encryption is common, and network layer encryption is about to be introduced to the Internet through IPSEC. In addition, the claim on p 51 that end-to-end security services can be located between the link and network layer is incorrect. There is further lack of clarity over the layer at which encryption will take place, and compatibility between services at different layers, in the following few pages and between there and p 53.

13. the claim on p 28 that 'increasingly NHS users will be using smart cards for system access' is a surprise, given the IMG decision to endorse non-smartcard authentication devices from SecurID and Digital Pathways for remote login, and pilots of medical records using optical rather than smart cards.

14. The implication that there will be a National Public Key Infrastructure that will manage all cryptographic keys in Britain may reflect the desires of GCHQ but is impractical [12].

15. Further unsupported minor assumptions in the strategy include:

    - that health care managers, rather than clinical professionals, are responsible for the security of personal health information (p 41);
    - that contracting data is not identifiable (p 43);
    - that only two principals will ever be party to a secure session (p 57), which would exclude secure electronic case conferences.

10 What is the IMG strategy for?

At the July meeting, the BMA was offered two separate explanations of the IMG strategy document. The first was that it covered the ramifications of using cryptography. It fails to do that because it does not tackle the critical issue of escrow.

The second was that an architecture had been proposed for cryptography in the NHS. This is true insofar as the report steers the NHS towards adopting the GCHQ strategy for secure email (or some close variant).

The stated terms of reference in the strategy document were on IMG's admission varied to downgrade the importance of digital signature. What other variations were there?

The original terms of reference also asked:

- what crypto should be used for. This is answered only in general terms;
- the costs and benefits. The costs are wrong, and placating medical concerns about protecting patients' privacy appears to be seen as the main benefit;
- the available standards and products. The only available product discussed appears to be PGP, and the assessment given of it is inaccurate;
- the additional costs of central provision. These are not discussed, and the more realistic option of distributed provision of digital signature, encryption and certification services is not even mentioned;
- for a study team including people with a sound understanding of the NHS and the technology; especially the complexity of NHS management arrangements, variety of systems, available products, and the complexity of key management in very large decentralised organisations. This is also lacking.

However, on page 6 the strategy provides a completely separate set of terms of reference: to examine contexts for encryption and other security services; to describe suitable techniques; to identify impacts on various parties; and to discuss an implementation approach. It is difficult to understand why one single IMG strategy document contains two so different terms of reference, one in the appendix and one in the body of the text. However, assuming that the version in the body of the text is the real terms of reference within which the experts were instructed to work, then the strategy can still be criticised. In our view, it does not examine the contexts in which encryption would be required with any completeness or accuracy; it does not describe suitable techniques; it does not identify and describe the costs and other impacts with any precision; and it contains no practical guidance on implementation. The proposals actually made are in our view wrong in almost every detail.

The influence that GCHQ had on the evolution of the IMG strategy ought to be stated clearly together with supporting documentation. We are told that the *'national interest in protection of person identifiable data'* led to consultation with GCHQ's protective organisation, CESG. Yet the national authority for personal information is not CESG but the Data Protection Registrar — who has waged a long battle against the intelligence services over whether these services should register under the Data Protection Act [34].

It is further stated that *'CESG's advice has been most helpful and, while not leading to the introduction of any specific functionality into the recommended NWN key management infrastructure, has influenced the recommendations made in this report so that they allow for this possibility'* (of interworking with a future national key management infrastructure).

There is also the remark (mentioned above) that the NHS would have to talk with CESG about "the possibility of ... **being allowed** to use alternative algorithms". Who is doing the allowing here, and under what law or regulation? According to the DTI, the use of cryptography within the UK is completely unrestricted [42], so we must insist that the authority for the statement quoted above is clearly identified. Which minister will be concerned with 'allowing' the use of alternative algorithms? Under what law or regulation will the minister be exercising that power?

When the director of CESG, Andrew Saunders, appeared at an MoD sponsored meeting to sell the TTP programme on the 26th June, he was asked whether his department had advised the Department of Health that the IMG strategy was sound, or whether they had at least had sight of it before its release. The question was not answered on that occasion, but we must insist that it is answered now.

Common sense suggests the following explanation. GCHQ wished to promote its 'secure' email protocol (as admitted in [24]). It involved the IMG and the NHS Executive as government agencies (strategy p 16: *'NHS would represent a large user community, and this size of market would encourage the development of an active and competitive market in Red Pike products'*). A policy decision was taken to use the GCHQ approved offering in the National Health Service, and the production of the IMG strategy document was an exercise in collecting arguments to provide ex post facto justification for this. The fact that the NHS Executive, a central government department, commissioned the report quite clearly constrained its authors, despite the lack of clarity over the report's remit.

The IMG cryptography strategy is consistent with this explanation, in which central government interests rather than individual patient privacy are paramount. Such an explanation also sheds light on why the IMG repeatedly assured the BMA that the pilots would not support key escrow, and yet went ahead with it none the less.

However, the whole GCHQ approach to securing email is inappropriate in the health care context, as are the arguments advanced in the IMG strategy to support its adoption.

11 Conclusion

The debate over security in NHS networking originally arose because of doctors' growing concern that many systems being built or proposed by IMG on behalf of the NHS Executive made large quantities of personal health information available centrally to administrators and other persons who are not clinicians, are not involved directly in the care of individual patients, and do not have the consent of patients to share this information. These systems include but are not limited to Clearing, HES, administrative registers, prescription pricing, and a growing number of disease specific databases.

Part of the Department of Health's response to this concern was the IMG strategy for cryptography. Unfortunately, as set out above, this strategy will not overcome the dangers of centralising personal health information, and it will not protect it from clandestine access by the police, benefit agencies, social work departments and other government bodies. On these grounds alone it is unlikely to win the trust of either professions or the public, both of which are essential if the NHS is to enjoy the benefits that networked clinical systems could bring. It is also unacceptable for a large number of detailed technical and other reasons.

What is needed is an architecture, based on open systems and standards where possible, that reflects the structure of trust in existing practice; which supports digital signatures for safety and encryption for privacy; which supports access controls of the type described in the BMA security policy; which has not been weakened to allow undetected intrusion by police and other government agencies; and takes into account the information technologies actually used or about to be introduced by British healthcare providers.

References

[1] "Transaction Security System", DG Abraham, GM Dolan, GP Double, JV Stevens, in *IBM Systems Journal* v 30 no 2 (1991) pp 206–229

[2] *'Fast Software Encryption — Fist International Workshop'*, R Anderson (editor), Springer Lecture Notes in Computer Science no 809

[3] "Why Cryptosystems Fail", RJ Anderson, in *Communications of the ACM* vol 37 no 11 (November 1994) pp 32–40

[4] "On Fibonacci Keystream Generators", R Anderson, in *'Fast Software Encryption — Second International Workshop'*, B Preneel (editor), Springer Lecture Notes in Computer Science no 1008 pp 346–352

[5] "Robustness principles for public key protocols", RJ Anderson, RM Needham, in *Advances in Cryptology — Crypto 95* Springer LNCS vol 963 pp 236–247

[6] "NHS Network Security", RJ Anderson, 30 May 1995

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

[8] *'Security in Clinical Information Systems'*, RJ Anderson, published by the British Medical Association, January 1996

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

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

[11] "A Security Policy Model for Clinical Information Systems", in *Proceedings of the 1996 IEEE Symposium on Security and Privacy* pp 30–43

[12] "The Zergo Report — Initial Comments", RJ Anderson (May 1996)

[13] "On the Reliability of Electronic Payment Systems", RJ Anderson, SJ Bezuidenhoudt, in *IEEE Transactions on Software Engineering* v 22 no 5 (May 1996) pp 294–301

[14] "The GCHQ Protocol and its Problems", RJ Anderson, MR Roe, *to appear at Eurocrypt 97*; available from `www.cl.cam.ac.uk/users/rja14`

[15] Discussion between B Beeby, R Anderson, G Kelly and J Williams, 20th November 1996

[16] "A Fast New DES Implementation in Software", E Biham, in *Preproceedings of the Fourth Fast Software Encryption Workshop* (Technion, Haifa, 20–22 January 1997) pp 241–252

[17] *'Secure Computer Systems: Mathematical Foundations'*, DE Bell, LJ LaPadula, Mitre Corporation report ESD-TR-73-278

[18] *'Cipher Systems'*, H Beker, F Piper, Northwood 1982

[19] "Directed Acyclic Graphs, One-way Functions and Digital Signatures", D Bleichenbacher, UM Maurer, *Advances in Cryptology — Crypto 94* (Springer LNCS v 839) pp 75–82

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

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

[22] "Optimizing a Fast Stream Cipher for VLIW, SIMD and Superscalar Processors", CSK Clapp, in *Preproceedings of the Fourth Fast Software Encryption Workshop* (Technion, Haifa, 20–22 January 1997) pp 254–268

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

[24] *'Securing Electronic Mail within HMG'*, CESG document T/3113TL/2776/11, 21st March 1996

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

[26] "United States pushes key escrow, security in OECD", in *Computer Fraud and Security Bulletin* (Mar 96) pp 7–9

[27] *'Cryptography's Role in Securing the Information Society'*, K Dam, H Lin, National Academy Press, Washington, D.C. 1996

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

[29] *'Keeping it Confidential — Protecting business information'*, Department of Trade and Industry, 1996

[30] "New Directions in cryptography", W Diffie, ME Hellman, in *IEEE Transactions on Information Theory* v IT-22 no 6 (Nov 76) pp 644–654

[31] "Exhaustive Cryptanalysis of the NBS Data Encryption Standard", W Diffie, ME Hellman, in *Computer* v 10 no 6 (June 77) pp 74–84

[32] *'Product Block Cipher for Data Security'*, WF Ehrsam, CHW Meyer, RL Powers, JL Smith, WL Tuchman, US Patent 3,962,539 (8th June 1976)

[33] "HMG's Unified Incident Reporting and Alert Scheme", P Fleury, talk given at *'Information Security — is IT Safe?'*, IEE, 27th June 1996

[34] "The law and its impact on the way data security issues are addressed", E France, at *Information Security — is IT Safe?"*, IEE, London, 27 June 1996

[35] "Future frameworks for data and privacy protection", E France, at *Liberty on the Line: Opportunities and Dangers of the Superhighway*, 14th November 1996

[36] "Encryption Trials — Aims, Objectives & Issues", D Garwood, 1996

[37] *'Good Medical Practice'*, General Medical Council

[38] *'Confidentiality'*, General Medical Council

[39] *'A Strategy for Security of the Electronic Patient Record'*, A Griew, R Currell, IHI, University of Wales, Aberystwyth, 14/3/95

[40] P Gutman, *personal communication*, July 96

[41] A Hassey, *Talk at Personal Information workshop, Cambridge, June 96*

[42] N Hickson, Department of Trade and Industry, speaking at *'Information Security — Is IT Safe?'*, IEE, Savoy Place, London, 27th June 1996

[43] *'Information Technology Security Evaluation Criteria'*, EU document COM(90) 314 (6/91)

[44] "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

[45] "A Proposed Architecture for Trusted Third Party Services", N Jefferies, C Mitchell, M Walker, in proceedings of *Cryptography Policy and Algorithms Conference*, 3–5 July 1995, pp 67–81; published by Queensland University of Technology

[46] "A Proposed Architecture for Trusted Third Party Services", N Jefferies, C Mitchell, M Walker, in *Cryptography: Policy and Algorithms*, Springer LNCS v 1029 pp 98–104; also appeared at the Public Key Infrastructure Invitational Workshop at MITRE, Virginia, USA, in September 1995 and PKS '96 in Zürich on 1st October 1996

[47] "Information and the NHS (for me or for them?), S Jenkins, in *Personal Information — Security, Engineering and Ethics* (21–22 June 1996), preproceedings published by the Isaac Newton Institute, Cambridge, pp 1–21

[48] *'The Codebreakers'*, D Kahn, Macmillam 1967

[49] *'La Cryptographie Militaire'*, A Kerkhoffs, in *Journal des Sciences Militaires*, 9th series, IX (Jan 1883) pp 5–38; (Feb 1883) pp 161–191

[50] "Constructing digital signatures from a one-way function", L Lamport, *SRI TR CSL 98* (1979)

[51] Talk given by J Leach at meeting organised by 'Scientists for Labour', 14th November 1996

[52] B Molteno, comments at BMA House, 13th June 1996

[53] *'Information Systems Security: Top level policy for the NHS'*, IMG document 2009 (b)

[54] *'NWN Threats and Vulnerabilities'*, 5 April 1995, IMG document NWNS/T1.22

[55] *'NHS-wide networking: data security policy'*, IMG document NWNS/T3.3

[56] *'NHS wide networking security architecture'*, 3 April 1995, IMG document NWNS/T1.21

[57] *Security Guide for IM&T Specialists'*, 3 April 1995, IMG document NWNS/T5.11

[58] *'NHS/CCTA Internet Security Report'* version 1.3

[59] *'NHS IS Reference Manual'*, December 1995

[60] Press release, NHS Executive, 17th October 1996

[61] NHS Wide Clearing System, Encryption Pilot Project, Project Board Minutes for 18/11/96

[62] NHS Wide Clearing System, Encryption Pilot Project, Project Board Minutes for 29/1/97

[63] "Announcing Development of a Federal Information Processing Standard for Advanced Encryption Standard", National Institute of Standards and Technology, in *Federal Register* v 62 no 1 (Jan 2 1997) pp 93–94

[64] "Managing Health Data Privacy and Security: a case study from New Zealand", R Neame, in *Personal Information — Security, Engineering and Ethics* (21–22 June 1996), preproceedings published by the Isaac Newton Institute, Cambridge, pp 207–215

[65] "Encryption and the Global Information Infrastructure: An Australian Perspective", S Orlowski, in proceedings of *Cryptography Policy and Algorithms Conference*, op. cit., pp 31–39

[66] "Euro-Clipper chip scheme proposed", D Peachey, in *Communications Week* no 151 (18 September 1995) pp 1, 81

[67] "Providing Secure, Recoverable Email", P Peterson, in *Network Security* (Aug 96) pp 15–19

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

[69] "Move to Strengthen Information Security", Press Association, 06/10 1808

[70] "UK to license information encryption services", Reuter RTf 06/10 1355, London, June 10th

[71] "Institutionell-organisatorische Gestaltung informationstechnischer Sicherungsinfrostrukturen", A Roßnagel, in *Datenschutz und Datensicherheit* (5/95) pp 259–269

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

[73] R Rogers, in discussion at BMA house, 2nd May 1996

[74] "Keynote Address", AW Saunders, in *Information Security — is IT Safe?*, IEE, London, 27 June 1996

[75] "CESG Recommendations for Secure Electronic Mail — A Statement by Andrew Saunders, Director CESG", available at `http://www.cesg.gov.uk/cesghtml/news/25_2_97.htm`

[76] "The Blowfish Encryption Algorithm", B Schneier, *Dr. Dobbs Journal* v 20 no 4 (Apr 94) pp 38 - 40

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

[78] "The History of Subliminal Channels", GJ Simmons, in *Information Hiding* (proceedings of first international workshop), Springer Lecture Notes in Computer Science v 1174 pp 237–256

[79] "Proposal for a Council Decision — in the field of security of information systems, concerning the establishment of a Europe-wide network of Trusted Third Party Services (ETS)", SOGIS (Senior Officials' Group, Information Security) 23/11/95

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

[81] *'Project Initiation Document — Tees Security Pilot'*, Issue 1 draft 3, Syntegra, 23rd October 1996

[82] Letter from Ian Taylor MP MBE to the writer, 16 December 1996

[83] "EC plans encryption rules in bid to police information superhighway", J Thorel, in *Nature* v 377 no 6547 (28/9/95) p 275

[84] Urteil im Zivilprozess Winter/Dresdner Bank AG, Amtsgericht Darmstadt, 24. Februar 1989, Geschäftsnummer 36 C 4386/87, Richter Jerschenkowski

[85] "The use of encryption and related services with the NHSnet", Zergo Ltd., published as NHSE IMG document number E5254, April 1996

[86] "Zergo's response to 'The Zergo report — initial comments' by Dr Ross Anderson", Zergo Ltd., 17th May 1996