Maintaining consumer confidence in electronic payment mechanisms

Nicholas Bohm[1], **Ian Brown**[2] and **Brian Gladman**[3]
[1]Electronic Commerce Working Party, The Law Society, 113 Chancery Lane, London WC2A 1PL
[2]Hidden Footprints Ltd., Gower Street, London WC1E 6BT
[3]Information Security Consultant

Abstract
Credit card fraud is already a significant factor inhibiting consumer confidence in e-commerce. As more advanced payment systems become common, what legal and technological mechanisms are required to ensure that fraud does not do long-term damage to consumers' willingness to use electronic payment mechanisms?
Keywords: non-repudiation, security, liability

Introduction

Consumers regularly cite the possibility of fraud as a reason to avoid on-line purchases. While they understand the precautions required in real-world use of credit cards, and may understand the liability régime that applies, they are rightly cautious about the significant changes involved in the move to an electronic environment. Well-publicised cases of large-scale card fraud linked to the Internet have further damaged confidence.

Many of these fears are misplaced. The ease with which credit card numbers can be obtained as they travel over the Internet, even before the widespread deployment of secure links between customer and merchant, is greatly overestimated. Insecure merchant databases are in reality the source of most stolen details.

But insecurity in customer *and* merchant systems is likely to promote on-line fraud which, when combined with trends in risk allocation by retail banks, will seriously damage already fragile consumer perceptions of the safety of on-line payment schemes. This could be very damaging for electronic commerce.

This paper discusses some of these technical problems and the risks they introduce. We compare traditional transactions such as payments by cheque or credit card with the use of newer remote data systems. We then analyse who bears the risk of fraud, and explore measures used or needed to reduce it. We argue that the approach taken by banks is unfair to their customers in some cases and fails to encourage the development of adequate security measures. The possibility of large-scale fraud where risk falls on the consumer will damage the consumer confidence required in electronic payment mechanisms for e-commerce to succeed.

Our analysis is based on English law except where otherwise stated: the law of other jurisdictions may not be the same.

Forged cheques

If your bank debits your account with payment of a cheque that you did not sign, it has no authority for the debit it has applied and must credit your account with the amount charged. The quality of the forgery and care taken by the bank are irrelevant: a cheque is a bill of exchange, and under section 24 of the Bills of Exchange Act 1882:

> "... where a signature on a bill is forged ... , the forged ... signature is wholly inoperative, and no right to retain the bill or to give a discharge therefor or to enforce payment thereof against any party thereto can be acquired through or under that signature, unless the party against whom it is sought to retain or enforce payment of the bill is precluded from setting up the forgery ... . "

The Bills of Exchange Act 1882 did not introduce new law. It codified the contemporary common law, and reflected the more general rule which still prevails in English law. If someone wishes to enforce a document against you on the basis that you are bound by it because you signed it, and if you deny that you signed it, then it for them to prove that the signature it bears was made or authorised by you, and not for you to prove that it was not.

An obvious advantage of the existing rule is that the bank can decide for itself (at its own risk) what level of care to apply to signature verification. Items of small value will not usually be checked at all, but unusual or very large items may be checked not only by careful inspection of the signature and comparison with a specimen card, but also by alternative means such as a telephone call to the customer.

If the bank rejects a cheque presented for payment by the forger, nobody suffers an unfair loss. But if the forged cheque is presented by a merchant who has accepted it from the forger in exchange for goods or services, the merchant suffers the loss despite having had no means of verifying the genuineness of the signature. For many years, merchants either accepted that risk or declined to take cheques. Cheque fraud led more and more merchants to refuse cheques, which annoyed bank customers and cut into the banks' fee income. The banks therefore introduced the use of cheque guarantee cards covering cheques up to a modest limit (£50 when introduced in 1965 and now more usually £100 or £250). The effect was to transfer the risk of small forgeries from the merchant to the bank: the banks could be seen as delegating to the merchant the signature verification process (using the signature on the card for comparison) in relation to smaller amounts (where they might themselves already apply little or no care to verification checks).

Although the rule that the bank bears the risk of forgery is plain, it does not follow that customers can easily reject any debit to their account based on a cheque simply by claiming that it is a forgery. Although some forgeries are crude enough to be obvious to anyone, others are considerably more skilful. If the bank produces a cheque bearing a signature which even on close inspection is indistinguishable from the customer's signature, perhaps supported by the evidence of a professional document examiner, then the customer cannot expect to succeed by mere unsupported denial. The customer will in effect have to rebut the evidence produced by the bank, and may in some cases be unsuccessful in doing so even though the signature is indeed a forgery.

To make this point does not amount to exposing some fundamental flaw in procedures that rely on signatures: it merely shows, as is evident to common sense, that those procedures are not perfect. Controlled trials show that professional document examiners misattribute 6.5% of documents while untrained persons of comparable educational attainment perform much worse with a mismatch rate of 38.3% [Kam97, Kam98]. Indeed, examiners assert that forged signatures are almost always easy to distinguish from genuine ones on close examination, however convincing they may be on casual inspection [Harrison58].

But because banks cannot practicably examine all signatures closely enough to detect forgeries which may be evident on close examination, and because a minority of forgeries are so good that they cannot be detected at all, we conclude that they run

real risks (and indeed incur real costs) from forgery of cheques and other written instructions. Acceptance of these risks and costs has not proved a major impediment to UK current account banking as a business. We suggest that this conclusion should be used as a point of comparison for the acceptability of the corresponding risks in other forms of payment transactions discussed below.

Credit and debit card liability rules

Credit and debit cards came into common use in the United States in the late 1950s, and were introduced in the UK by Barclays Bank in 1966. Their use expanded very rapidly in the 1980s, perhaps stimulated by the growing disparity between the amounts for which cards could be used and the more modest amounts covered by cheque guarantee cards.

Card transactions do not involve cheques, with the result that section 24 of the Bills of Exchange Act 1882 does not apply. Card issuers (referred to here for convenience as banks) are therefore free to apply different rules from those governing the risk of forgery of cheques, and the rules embodied in the terms and conditions on which they issue credit and debit cards are indeed different.

Although banks' terms vary in their details, the general rule is that the customer is responsible for all transactions carried out by the use of the card with the customer's authority, and for all other (i.e. fraudulent) transactions carried out by the use of the card, up to a limit of £50. This limited liability for fraudulent transactions ceases when the customer informs the bank that the card has been lost or stolen. These rules reflect the provisions of sections 84 and 171 of the Consumer Credit Act 1974 and the regulations made under it relating to credit cards.

By comparison with the case of cheque forgery, this régime transfers to the customer a limited part of the risk of fraudulent use of the customer's card. Such use of the card depends on physical possession of the card, however, and the customer can reduce the risk by taking good care of the card and by promptly reporting its loss. Taking care of articles like cards or door keys is largely a matter of common sense (to be contrasted with the precautions required to protect electronic systems, as discussed below). The £50 exposure can be seen as providing an incentive to the customer to take care of the card and report its loss promptly.

The balance of the risk that is not carried by the customer is borne by the bank or the merchant. The terms governing the relationship between the merchant and the bank determine this allocation. Where the cardholder is present at the transaction, and where the merchant has not been plainly careless in accepting a non-conforming signature, and has complied with limits on the amount of an individual transaction and other applicable rules, the bank normally carries the risk. Merchants therefore have an incentive to take appropriate care in accepting card transactions, but are guaranteed payment by the bank if proper care has been taken, just as if they had accepted a cheque with a cheque guarantee card (and with the advantage that much higher amounts can be covered).

It is clear from this discussion that possession of the relevant card plays a substantial role in authenticating a card transaction, and that signature verification is much less significant than in the case of cheques. This conclusion is supported by the fact that in the UK signatures are usually made on multi-part forms, with the customer retaining the top copy. In any subsequent dispute, the copy or copies of any voucher available for expert examination will bear only "carbon" copies of the customer's (or alleged customer's) signature. Forgeries are an order of magnitude more difficult to detect from carbon copies.

Cards may also be used in transactions where the cardholder and the merchant do not meet, and no voucher may be signed. Examples are the use of a card for mail orders, by telephone, by electronic mail or through a web page. (Such transactions are classified as "cardholder not present" or sometimes as "MO/TO", meaning "mail order or telephone order". We refer to them for simplicity as remote card transactions.) The incidence of risk in remote card transactions is quite different from that where the card is presented by the customer to the merchant.

In a remote card transaction, the customer provides the merchant with information apparent from the face of the card: its type (typically Visa or Mastercard), its number, its expiry date and the name of the cardholder. (Except in the case of mail order, the customer provides no signature; and in mail order the merchant cannot compare the signature with that on the card.) The ability of the customer to provide the card information does not depend on possession of the card: it is available to anyone through whose hands the card has passed in the course of earlier transactions, perhaps to the cardholder's family and friends, and to anyone who may have received or intercepted the information as it was transmitted by telephone or through the Internet. Where the purpose of a remote card transaction is to order goods for delivery, the merchant may be able to check the address of the cardholder through the bank, and can decline to deliver the goods except to that address. Where available this procedure provides some protection from fraudulently placed orders. Where the order is for online services, however, such as the downloading of software or the provision of access to online databases, no such precaution can be taken. In these cases there is very little impediment to fraud (either by the customer falsely repudiating a genuine transaction, or by an imposter using the customer's card details without authority). (We have disregarded the process of "authorisation", where a merchant complies with a requirement to check with the bank whether a transaction may proceed. The reason is that while this process enables the bank to check that the customer has not reported the card stolen or is not exceeding a credit limit, for example, it does not enable the bank to check that the card information is being used by the customer rather than an imposter. It does not guarantee payment where the customer is not present at the transaction, and therefore does not alter the balance of the risks under discussion.)

The liability régime is simple, although its implications do not seem to be widely understood. If the cardholder denies having entered into a remote card transaction in which the relevant card information was provided, and there is no evidence of delivery of goods to the customer or voucher signed by the customer, the bank has no basis on which to debit the customer's account. The mere use of the card information is not enough to show that the customer authorised the transaction, because of the wide class of other persons to whom the information is available. Merchants are of course members of that wide class, possessing card information in abundance: for the merchant, "forgery" of a remote card transaction is a trivial task. Faced with apparently unmanageable risks of this kind, the banks have adopted the simple approach of requiring the merchants to carry the risk. If a cardholder repudiates a remote card transaction for which there is no evidence of delivery of goods to the customer, or voucher signed by the customer, the bank makes a "chargeback", i.e. obtains reimbursement from the merchant of anything paid to the merchant in respect of the transaction (and may also make an administrative charge). The merchant is in practice unable to transfer this risk to anyone else, since he is unlikely to be able to prove who initiated the relevant transaction.

The banks naturally appreciate the perilous position in which this régime places the merchants. They are sometimes unhelpful to cardholders who repudiate remote transactions, by refusing to reverse the repudiated debit and responding that the cardholder must resolve the dispute with the merchant; but they are aware that this stance is unsustainable where the cardholder denies having participated in any transaction with the merchant, and in the face of persistence by the cardholder they will accept that the transaction must be reversed. (This is not to say that customers face no problems: in some cases they have required considerable persistence in the face of evasions, and faced long delays; in others they have suffered foreign exchange losses where debit and credit of the same amount in foreign currency has left them with a shortfall.)

The greatest risk to the merchant obviously arises from the provision of online services. Although this is not a new risk, the range of services which can be provided online has expanded greatly with the commercialisation of the Internet. The problem of managing the resulting risks for merchants may well prove to be a growing impediment to the growth of electronic commerce in online services.

For transactions carried out by the cardholder using a web browser to connect to a supplier's web page, it is possible to establish a secure connection so that the card information is delivered in encrypted form (using protocols such as TLS or SSL [Dierks99]). This procedure is widely followed, and provides protection against interception of the card information in transit. It cannot affect the wide availability of card information from other sources, and since the procedure cannot provide evidence that the supplier of the card information is authorised by the cardholder to conclude the transaction, it does not materially reduce the merchant's risk.

Visa and Mastercard have promulgated a standard for Secure Electronic Transactions, referred to as "SET" [SET99]. It would enable the merchant to check that the bank will accept the cardholder's authority as genuine, and would thereby presumably remove the risk from the merchant, or at least reduce it. The SET standard has not gained acceptance, perhaps because it is over elaborate and its implementation would be burdensome and expensive. The SET specifications do not deal with the legal régime covering relations between the bank, the merchant and the customer, presumably because this is a matter for individual banks and because the existing régime is expected to continue to apply. If the merchant's risk of chargeback is to be removed or reduced by treating the customer as present in a SET transaction, it therefore seems probable that the customer will be precluded from repudiating a SET transaction which appears to have been authorised by that customer. But the risk to the customer of losing control of the means of authorising SET transactions (which consists of information stored in electronic form) is very different from the risk of losing a plastic card, as we explain below.

As electronic commerce grows, and merchants experience increasing levels of chargeback from the use of conventional card information in the new electronic medium, it is likely that there will be growing pressure from merchants for the adoption of procedures to lessen their exposure. But for the reasons explored below, any temptation for the banks to use the adoption of new technical security procedures to transfer those risks from the merchant to the customer should be sternly resisted, not only in the interest of the customers but in the wider interest of public confidence in electronic commerce.

Verification procedures

The procedure for verifying the authenticity of banking instructions by using a PIN, password or other security information falls into the class of procedures based on a shared secret. When the customer uses a PIN, the bank uses its knowledge of the PIN to check that it is genuine. Such procedures may be contrasted with signature verification, which relies instead on the physiological property that a person can easily make his own signature but cannot easily make another person's signature well enough for a forgery to pass careful examination.

People cannot give away their physiological properties, but either of the parties to a shared secret can reveal it to facilitate fraudulent use. Where two parties share a secret whose misuse can cause loss, it might seem remarkable that they should agree that one of them should assume sole responsibility for the loss of the secret. But that is in effect what the banks have expected of their customers: both the bank and customer must know the PIN, the customer to use it and the bank to check it. Both could reveal it to a third party who could misuse it.

Banks are of course regulated bodies required to be managed by persons fit and proper for the purpose, and (with occasional spectacular exceptions) do not pursue financial crime as a corporate purpose. But technical security measures are very difficult to design and implement successfully. One of the more notorious weaknesses of commerce and industry in Britain has been its inability to obtain the full potential benefit of information technology through failure to integrate it with other parts of a business. The computer department, even when called the information technology department, is rarely on the career path to the boardroom. Computer specialists, even when employed by banks, are rarely either managers or integrated into the ethos of management. They can easily become impatient of mainstream managers' failure to understand the potential of information technology. Although no less honest than other professional people, computer specialists in banks are particularly vulnerable to the stresses of cultural isolation, low esteem, and the temptation to prove they can outwit the system. These considerations militate against any claim that the culpable disclosure of security information must necessarily be more likely to originate with the customer than with the bank.

The use of biometric information to authenticate customer transactions is seen as one answer to the problem of reliance on shared secrets. Iris recognition is now being tested by the Nationwide Building Society as an alternative to PINs [Hawkes98]. A cash dispenser compares various properties of a card user's iris with a stored record, making it extremely difficult for anyone but the card owner to withdraw money using it. Their system is said to make only one error in 131,000 cases if the probabilities of falsely accepting or rejecting an individual are set equal. Even this might be an unacceptably high error rate if very large numbers of false attempts were feasible, which is of course not the case where the user must be personally present and the system is operated in the presence of attendants.

The risks are different in unattended or remote operation, where a photograph or video of the user's iris might be presented to the camera; and the risks are liable to increase significantly if use becomes widespread in a variety of applications and many businesses come to have databases of customer personal identification and linked iris codes. And it remains the case that anyone with sufficient access to a bank's financial systems may still be able to create false transactions linked to a customer. No doubt banks have procedural mechanisms to limit such risks, but there is no independent evidence by which customers can judge for themselves the

effectiveness of such procedures, despite the fact that customers may be expected to rely on them by carrying the risk of fraud.
Where the risk to the customer is effectively limited to £50, and where that limited risk can reliably be terminated by notification of the loss of a card, with the bank carrying the balance of the risk of fraudulent transactions, the outcome of the liability régime may seem reasonable enough. But where there is no financial limit, or where fraud can occur without the loss of the card to alert the customer, to require the customer to bear the whole of the risk seems to us to be decidedly unreasonable. Both of these situations can arise with current online banking products.

Digital signatures

It might reasonably be thought that authentication essentially depends on using either biometrics (like handwriting, voice recognition, fingerprints, retinal scans, etc) or shared secrets. This was indeed true until the publication in 1976 of New Directions in Cryptography by Martin Hellman & Whitfield Diffie [Diffie76]. That paper, which laid the foundations of public key cryptography, showed that it was possible to establish a procedure by which (transposing it into the context of this paper) customers can control unique, secret "signature" keys for which they can provide related non-secret information that can be used by a bank to verify that instructions issued by them have been signed using these keys. (We refer to this verification information below as a "verification key".) Provided that this scheme is soundly implemented, and that owners keep their signature keys secret and under their own control, transactions signed with them can be attributed to their owners with a high degree of confidence. In such schemes there is no shared secret since the bank does not know the value of the signature key and cannot discover it from the information it is given.
It might be thought that with the customer in sole control of a signature key, the problems of liability could acceptably be solved by requiring the customer to accept responsibility without limit for all use of the signature key (at least until the bank is notified of a compromise). But that conclusion would involve considerable dangers, which we explain below after reviewing some other security issues.

The security of banking computer systems

While it is possible to establish a confidential channel between a bank and a customer, this does not eliminate the possible impact of security vulnerabilities in the computer systems used by banks and customers for on-line transactions.
Although UK banks have denied that weaknesses in their computer systems are responsible for alleged fraudulent transactions, the evidence discussed above and in Ross Anderson's papers [Anderson93, Anderson94] highlights failings in such systems which can have a serious impact on customers. The banks have been unwilling to allow independent experts to examine their systems, justifying this stance by claiming that they need to keep the design and operation of their systems secret in order to protect them from attack.
This approach, known in security circles as "security through obscurity", is now widely discredited, because any advantages provided by secrecy are offset by the fact that this secrecy allows serious faults to exist in systems for long periods without being discovered. The consequences are well illustrated by the ATM phantom withdrawal problem, where the banks have been asserting for years that the design of their systems make such events impossible in the face of steadily growing evidence that they must be wrong. As cases have come to court, defence expert witnesses have

gained steadily more access to the details of banking computer systems, and have discovered that banking computer systems do not exhibit the invulnerability that the banks claim for them.
If banks carried the whole risk involved in on-line payment, the vulnerabilities of bank computer systems would be of lesser public concern, but the prosecution of a customer for demanding repayment of sums he claimed were wrongly debited to his account [Anderson94] shows the serious consequences of a bank's attempt to transfer the risk to a customer.
It follows that customers' interests are not adequately protected even by an acceptance in principle by the banks that they will themselves carry all the risks of fraud in online transactions. In practice the banks will employ mechanisms to prevent fraud, and where these mechanisms fail the banks will sometimes wrongly seek to transfer the consequences to their customers. While at first sight account security measures such as PINs, passwords and digital signatures may seem to protect customer interests, their weaknesses will sometimes be used by banks to explain failures that are in reality the result of internal problems with their own systems. In this sense, therefore, it can be argued that security based on the secrecy of the mechanisms employed by the bank operates more in the interests of the bank than of its customers.
The security of online payment systems from a customer perspective is therefore not very satisfactory. Although there is no doubt that the vast majority of customers will not experience problems, for the small number that find themselves victims of security failures in banking computer systems the consequences can easily be very serious. Customers should seek a bank that offers better security than that provided by PINs and passwords alone, and that has allowed independent experts to audit and publish the results of security reviews of the computer systems it uses to provide online services. They may be in for a long search; in the meanwhile, they might do well to place limits on the amounts which can be transferred from their accounts on the basis of electronic instructions.

Client PC security

But even if banking computer systems were perfect, the majority of the computers used by customers will be home PCs that are most unlikely to meet any serious security requirements.
People tend to be very trusting of others and can often be persuaded to reveal their PINs and passwords when they should not do so. Many people have difficulty installing software on their PCs and find an "expert" neighbour or friend to help. It is not unusual to find that the helper will be given the codes needed to operate the service being installed, in order to check that it is working and to demonstrate its use to the real customer. Undoubtedly most helpers are honest but inevitably a few will use such knowledge for fraudulent personal gain.
A further concern is that typical PCs do not provide much real protection for PINs and passwords unless careful control is maintained over access to the PC as well as control over the software that is installed. PCs will often be used by several family members for a wide range of different pursuits. It is not difficult for anyone who has ongoing access to install software that will capture sensitive account and password data entered by users for later collection. This could easily be achieved by another family member or by someone called in to maintain the machine.
Such attacks can be even easier to mount if the software used for online transactions is not very carefully designed. Most modern PC operating systems can appear to run

several applications at once. They do this by temporarily moving applications and the data they are using from memory on to files on disc called swap files. Such files will often hold sensitive data such as passwords or security keys themselves, and they can be read with utilities that are widely available. A knowledgeable programmer could easily write software that searches the swap file to find the information. Recent research has shown that some security information has characteristics that are easy to detect unless it has been deliberately disguised, and this makes such attacks all the easier to design. A computer maintainer armed with software of this kind could easily recover such information as a matter of routine.

Although these forms of attack are probably rare at present, this is not the result of any inherent technical difficulty but because the gains are limited while online transactions are not yet widespread.

In addition to attacks requiring physical access, PCs used on the Internet are vulnerable to attacks in which software is remotely installed to capture and transmit a user's keyboard data to a remote locations. Since users are routinely asked to install "add-ons" such as applets and active controls, most users accept this as routine and will not understand how easy it is for a fraudulent site to install an applet that appears to offer one service but in reality captures and transmits security data back to the site in question. It would also be perfectly feasible to modify and redistribute an honest applet from a reputable company to do this. A number of cases have been reported recently in which commercial software has been found to provide its supplier with information about its user's activities, without the user having been made aware of the fact.

An even more potent attack would be one based on a computer virus (software designed to transfer itself from one computer to another unknown to their users, either on diskettes or over the Internet, and capable of affecting the working of any computer it reaches). Current viruses exhibit a range of behaviours from benign (or even beneficial) effects through to those of a highly malicious character, designed to inflict substantial damage on a victim's PC or the data it contains. But it is straightforward to write a virus that, once installed, looks for and captures PINs, passwords, account details and other sensitive data for transmission back to the virus writer when the victim next goes on line. By making such a virus covert - that is, as silent as possible, so that a PC user is unaware of its presence - it could easily do its job over months or even years without being detected.

If a customer gains access to a bank account through a local network, such as one operated by their employer, additional interception risks may be involved. Many companies will operate a "firewall" to protect their internal computer systems from external attacks. These will often prevent security protocols from operating "end to end" between the bank's system and the PC on the customer's desk. This may prevent the customer from gaining on-line access to the bank account or may require that access is gained indirectly using other computer facilities. In that case the additional computer and network connections involved may introduce further interception risks. The result could be access by other employees to information passing between the customer and the bank, such as passwords and other security information.

The inherent difficulties involved in computer security are discussed in The Inevitability of Failure: The Flawed Assumption of Security in Modern Computing Environments, a paper by scientists from the US National Security Agency [Loscocco98]. That agency is responsible for the security of US Government communications and for monitoring and deciphering foreign communications for intelligence purposes. Any paper by the NSA on computer system security carries

very high weight indeed. The thrust of this paper's argument is that it is unrealistic to expect that security mechanisms can be implemented in software without computer operating systems that offer effective security features of a kind that do not exist in current products.
Although purely software based PC payment procedures seem acceptable now, for reasons such as those discussed it is hard to believe they will continue to be seen as a robust solution for the longer term. Although the use of signature keys based on public key cryptography can greatly reduce the risks presented by the use of PINs, passwords and other shared secrets, even then the customer is dependant on keeping the private key secret despite the need to use it in a PC. In such an environment the customer is exposed to risks of the private key being compromised without having any means of detecting the compromise until fraudulent use comes to light. A sophisticated attack might leave no evidence of how it occurred, and the customer is therefore weakly placed to resist an assertion by the bank that the transaction must have been authorised.

Hardware based solutions

The unavoidable security limitations of software have led many to look for hardware solutions, such as those based on smartcards. Although software is easy to modify and hence subvert, this is less true of hardware, which makes it attractive for implementing security critical features. While hardware solutions offer better security assurance than software, they are also more expensive.
If secret data can be held in hardware, it is much less vulnerable to being discovered by an attacker. Smartcards are vulnerable to a number of forms of attack, but much less so than software since the expertise required is more specialised and the tools needed are less commonly available. But expertise in microelectronics is not rare, and many laboratories will have the necessary equipment. Several techniques have been developed to discover the internal secrets of smartcards, and some of these have been shown to be very successful for particular cards [Kömmerling99].
Such attacks have already become a serious problem for the purveyors of pay-per-view TV. At one point, Sky TV reckoned that smartcard forgery was costing in excess of 5% of its turnover. Once they are widely introduced, payment smartcards will clearly present an even more attractive target. Some attacks on cash dispenser cards have involved sophisticated and expensive techniques to deceive customers into giving their PIN to a fake machine, and recent research has shown that many smartcards are vulnerable to a fake machine extracting their secrets by observing the power they consume while calculating a digital signature. Organised crime will certainly be able to obtain the means to attack smartcards when the rewards justify the effort.
Moreover, the undoubted advantages of smartcards when compared with security mechanisms based on software are not as easy to harness as they seem. First, where smart cards are used to hold secret information, it makes little sense to transfer this information into a PC for use, since this will remove the very protection that the smartcard is intended to provide. So in order to maintain the security of the information, it has to be used on the card itself, and this is likely to require the card to have reasonable processing capability of its own. There are obvious cost implications. Secondly, at a practical level, almost no mass market PCs come with smartcard readers, and this seems unlikely to change unless the need is widely recognised and the costs involved are small.

Thirdly, a PIN or a password will be used to activate the card in order to guard against the fraudulent use of lost or stolen cards. If this is entered through the PC's keyboard it will be vulnerable to all the attacks discussed earlier. In this case it can be argued that the loss of the PIN is not so serious since fraud will require both the card and the PIN. Although this is true, if an attack has been mounted on a PC through a virus as described earlier, it would not be hard to extend the virus to use captured password data with the smartcard the next time it is inserted by the user.

These points are not made in order to deny the value of smartcards, but simply to point out that while they will offer big improvements when compared purely with software, they are not a perfect solution.

In order to overcome one of these vulnerabilities, at least one smartcard manufacturer is now offering a secure smart card reader with a small keypad for the entry of the PIN. This avoids the use of the PC for PIN entry, but it remains vulnerable to an attack in which a fraudulent application running on the PC (or a point of sale terminal) displays one transaction to the user while asking the inserted smartcard to authorise a completely different one. For example, a personal signature card used to sign credit card transactions is vulnerable to an attacker who presents a point of sale terminal to the user which purports to perform a genuine transaction but simultaneously authorises another transaction that is seriously to the user's detriment - examples might range from another credit card transaction for a large amount up to a re-mortgage of the user's home.

Smartcards are often seen as the perfect answer for implementing digital signatures, because signature keys kept on such cards can in principle have values not even known by their owners. This can prevent an owner from repudiating a genuine signature by publishing the key and claiming it to have been compromised before the transaction.

But providing a useful identity based signature which cannot be repudiated by this means remains very difficult, because (1) in order to ensure that the signature key is secret it must be generated on the card; (2) for the same reason it must never leave the card, and this requires that the transaction or document to be signed must be imported on to the card, with the signature process be performed by the card; and (3) the card has to export a verification key that allows the signature to be verified and associated with a person authorised to perform the transaction.

As already indicated, meeting requirements (1) and (2) currently requires relatively expensive "state of the art" hardware solutions, while meeting requirement (3) turns out to be difficult because it raises social and legal issues about how a person can be identified in a unique way.

A person's name alone is clearly not sufficient since names are not unique; but neither are names with birthdays or names with addresses (which can in any case change frequently). The use of verifiable biometric data - for example, fingerprints, iris or retinal scans or DNA data - offers a more robust solution but will be expensive. Use of such data also raises a number of ethical and privacy concerns such as those that come to the fore when identity cards are mooted: there are many circumstances where an individual may legitimately wish to use a pseudonym which has no link to any other name the individual uses. Moreover, while the costs might be contemplated for use with cash dispensers or point of sale terminals, it is less obvious that the cost of secure biometric data collection devices will soon become low enough for them to become "commodity" peripherals for home PCs. For this reason their value in the foreseeable future by consumers is somewhat uncertain.

In many respects typical PCs offer far more performance than is necessary for controlling the transactions involved in electronic commerce. They are designed for high levels of functional performance, but their resulting complexity makes the achievement of security objectives much more difficult. In many respects the ideal vehicle for e-commerce is a small self-contained computer system such as a palmtop with a small keyboard, a screen and (possibly) an infrared port to enable it to communicate with a home PC, a bank's or a merchant's computer system, or a point of sale terminal. By keeping this device simple, and by having a keyboard, a screen, a processor and secure storage in one small self-contained unit, it would be possible to have a highly assured capability for signing transactions without being dependent on other devices such as PCs or point of sale terminals.

Still further security could be achieved by having the secure storage for such a device implemented on a plug-in smartcard. It is possible to envisage a secure device with an integral keypad, screen, smartcard reader, PC interface and biometric input such as a fingerprint reader. If such a device could be manufactured at reasonable cost it could serve both as a point of sale terminal in a merchant environment and as a PC peripheral at home. In practice the merchant terminal would have to be more robust physically, but the two devices could share much of their security design in common. We think that one essential element in achieving public confidence in such a design will be its openness to scrutiny by independent experts, and the abandonment of "security through obscurity".

Devices of this kind will nevertheless be expensive. We do not think they will come into widespread use without being subsidised by banks and others who benefit from the growth of electronic commerce and have the skills to collaborate in their design. The most certain way to ensure that the banks have the necessary incentive to pursue this programme is to ensure that they carry the risks of the fraud that the programme would help to prevent. Such a programme is not without precedent: the spread of mobile telephony depends on large subsidies by network service providers to reduce the cost to users of buying mobile telephones.

Conclusion

Online electronic transactions are generally not yet well secured against fraud, but the apparent security of encryption and digital signatures is tempting some banks into transferring the risk of fraud to the customer. Where such a risk is discovered to have fallen on an innocent customer, electronic commerce could suffer seriously from the effect on public confidence. Banks should be among the first to promote the deployment of good security devices for their customers' use: ensuring that in the meanwhile it is the banks that carry the risk of fraud gives them the incentive they need.

Acknowledgments

The authors gratefully acknowledge the help they have received from Ross Anderson, Richard Clayton, Gilead Cooper, Alex Hamilton and Peter Landrock in comments on earlier drafts.

References

[Anderson93] R. Anderson. Why Cryptosystems Fail. Proceedings of 1st Conference on Computer and Communications Security '93, Fairfax, Virginia, USA, November 1993.

[Anderson94] R. Anderson. Liability and Computer Security: Nine Principles. Proceedings of European Symposium on Research in Computer Security, Brighton, UK, November 1994.
[Dierks99] T. Dierks and C. Allen. The TLS Protocol Version 1.0. RFC 2246, January 1999.
[Diffie76] W. Diffie and M. Hellman. New directions in cryptography. IEEE Transactions on Information Theory, 22(6) pp.644-654, November 1976.
[Harrison58] W. Harrison. Suspect Documents pp.373-426. New York: Praeger Publishers, 1958.
[Hawkes98] Nigel Hawkes. Machines will pay up in the blink of an eye. The Times [http://www.cl.cam.ac.uk/users/jgd1000/atm.jpg], April 1998.
[Kam97] M. Kam, G. Fielding and R. Conn. Writer Identification by Professional Document Examiners. Journal of Forensic Sciences 42 pp.778-786, 1997.
[Kam98] M. Kam, G. Fielding and R. Conn. Effects of Monetary Incentives on Performance of Nonprofessionals in Document-Examination Proficiency Tests. Journal of Forensic Sciences 43 pp.1000-1004, 1998.
[Kömmerling99] O. Kömmerling and M. Kuhn. Design Principles for Tamper-Resistant Smartcard Processors. Proceedings of USENIX Workshop on Smartcard Technology, Chicago, USA, May 1999.
[Loscocco98] P. Loscocco, S. Smalley, P. Muckelbauer, R. Taylor, S. Turner and J. Farrell. The Inevitability of Failure: The Flawed Assumption of Security in Modern Computing Environment. Proceedings of the 21st National Information Systems Security Conference [http://csrc.nist.gov/nissc/1998/proceedings/paperF1.pdf], October 1998.
[SET99] SET Secure Electronic Transaction LLC. The SET Specification. [http://www.setco.org/set_specifications.html] 1999.