Chip and Spin

Ross Anderson, Mike Bond, and Steven J. Murdoch

Computer Laboratory, University of Cambridge,
JJ Thompson Avenue, CB3 0FD, UK
{Ross.Anderson, Mike.Bond, Steven.Murdoch}@cl.cam.ac.uk

1 Introduction

The new UK "Chip and PIN" card payments scheme has recently gone live. It has been spun in the media so far as "a safer way to pay" and as "the biggest change to payment since decimalisation". However, the latest fraud figures show that fraud is up, not down – and the Chip and PIN scheme is being blamed [1]. So how secure is it really? And who will benefit most from its introduction? This note briefly considers liability issues, technical shortcomings and management failures.

2 Problems for the customer

Sadly it appears to be the customer who has most to lose in the transition to Chip and PIN; this section describes the issues briefly.

2.1 Liability and dispute resolution

Until the Chip and PIN initiative, manuscript signature was the only authentication mechanism used in the UK at Point-of-Sale (POS) terminals. Sales vouchers were governed by laws that evolved for cheques, of which the most important is that a forged signature is completely null and void. This gives the customer strong protection against abuse of a stolen card. Although some argument can be made about card terms and conditions, and about possible negligence by a customer, in practice the banking industry paid the costs of fraud.

ATM transactions, on the other hand, are authenticated by PINs. British banks for many years took the view that their systems were secure, and so customers who complained about 'phantom withdrawals' must be mistaken or lying. During 1992–4, there was a wave of phantom withdrawals, eventually traced to a criminal gang whose principals were convicted in 1994. During this period, several thousand people sued thirteen banks in a class action over phantom withdrawals. (The banks won.)

The late 1990s saw the growth of electronic banking services. Most banks introduced terms and conditions that, in effect, passed the burden of proof in disputes to the customer [2]. At the same time, fraud controls were relaxed. Early electronic banking systems allowed customers to pay money only to previously

nominated accounts, so that a customer could arrange to pay his regular bills online but would have to visit a branch or send a letter to pay someone new. This requirement has mostly been relaxed. There has since been a wave of 'phishing' attacks in which fraudsters trick customers into giving up electronic banking passwords using replicas of bank websites.

There has also been a second wave of phantom withdrawals, linked with false fronts attached to ATMs that harvest card and PIN data on the way through, and credit-card counterfeiting associated with retail outlets that are controlled by crooks, or whose staff have been subverted. The overall picture is one of rising technical fraud coupled with serious attempts by the banking industry to dump the liability on cardholders or merchants. It is well-known to students of security economics that when one party is responsible for protecting a system, while another party suffers when it fails, then security failure can be expected [5].

2.2 Regulatory issues

One might think that the shift of liability was intrinsic to the technology. If a cheque, or a signed paper, voucher, is forged, the signature is often nothing like the customer's. But using a PIN instead of a signature takes us into more difficult territory. If a card forger might have observed a PIN as it was used, or a bank insider might have stolen it from a computer system, then there is no obvious way to find out which [3]. However, in the USA, electronic banking is governed by 'Regulation E' which places the default liability squarely with the bank. Faced with a customer repudiation of a transaction, a bank must either pay up, or prove fraud by direct means; it is not sufficient to rely on general claims about system security. Banks can then take a business decision whether to install CCTV or other measures. Under this regulation, US banks have thrived – because the incentives are in the right place [6].

UK banks do have a voluntary code of practice [7] that is supposed to indicate liability during such circumstances, but it relies on ill-defined notions of "reasonable care", and does not specify the standard of proof a bank investigator must produce before denying a refund decision. So from the British customer's point of view, a signature is far preferable to a PIN. From the bank's point of view, however, a PIN is preferable. Demanding PIN entry converts more straightforward fraudulent signature denial attacks into the equivalent of ATM withdrawal denial attacks (where the banks often do not give the customer any benefit of the doubt). It may also be preferred by some merchants, as it absolves the shop-assistant of blame for transaction rejection (the computer is blamed). In the short term, though, merchants are unhappy; the new terminals cost money, and integrating them into a store chain's systems is a complex and expensive task. So the banks have changed the terms and conditions they offer merchants so that disputed signature-based transactions will in future be charged back to the merchant. In effect, if a fraudulent transaction was signed, the merchant pays; if a PIN was used, then we can expect that more and more it will be the customer who is asked to pay.

2.3 PIN use at the point of sale

The use of PINs in shops as well as at ATMs greatly increases the likelihood that PINs can be compromised by shoulder-surfing. This is not just a function of the frequency of PIN entry, but also of the environment and the associated social norms. People at ATMs are expected to be antisocial - to turn their back and shield their keyboard use – and other people in the vicinity are expected not to get too close. In a supermarket queue, on the other hand, shoppers may feel decidedly uneasy at shielding PIN entry, especially if there is an acquaintance in the same checkout queue or a nearby one. Also, many shop terminals do not appear to have been designed to facilitate customer security and ease of mind.

Social norms will evolve in time, and no doubt the security usability of terminals will improve. But it remains to be seen whether the cost will be an increase in crimes against the person. As PINs become both more important and more available, and as people are encouraged to use the same PIN for multiple cards, the shopper's purse will become a more attractive target for both pickpockets and muggers.

3 Problems with the system

Chip and PIN may well shift liability to the customer, but might this not be fixed by regulation? Does the system generate enough benefits that these might be shared out among the banks, the customers and the merchants, so that they all end up better off than before?

3.1 Fallback

At present, the system has a glaring weakness. If a card is presented at an ATM or POS terminal whose chip has been damaged, or which never had a chip, then the device falls back to magnetic strip operation. This means that a gang operating a dodgy business that skims their customers' credit cards can continue in business: now they record the PINs as well. They then take cards with broken chips and encode the magnetic strip with the skimmed data and the observed PIN. While previously they could only use a skimmed card to buy goods for resale (or in cahoots with a crooked merchant), they can now take money out of cash machines. The 'money laundering' overhead disappears and the criminal business becomes more profitable.

This is a known problem in countries that have used PINs at the retail point of sale along with magnetic strip technology. The banks are aware of the risk, and talk about withdrawing magstripe fallback at ATMs if the fraud gets out of control. But this will bring other costs with it; the several percent of cardholders whose chips fail because of static electricity each year will have to get new cards, and people whose card contacts get dirty will also have problems.

3.2 Cross-Border fraud

A further problem arises with the abolition of magstripe fallback – international travel. If UK ATMs will accept only chip cards, then US travellers will be unable to get cash from British ATMs; and if the magnetic stripe is removed from UK cards (as one banker suggested to us), then Brits in America will suffer the same inconvenience. But if respectable customers can transact abroad, so can the crooks.

Cross-border fraud has in the past held only limited advantages over local fraud. There is one big exception: France. France was the first country to introduce chip and PIN, about a decade ago. Fraud initially fell, but then climbed again. The crooks learned to use French stolen cards in outlets in Germany and the UK, and vice versa. In each case, the system falls back to the common available mode – magnetic strip – and the chip mechanisms are bypassed. Furthermore, the fallback mechanisms tend to be less robust than the magnetic strip mechanisms are in countries where this is the main technology. So although the move to chip and PIN cut fraud in France, it cut it much less than had been hoped. In fact, banks in Spain cut fraud even more, and without introducing chip cards: they simply set all floor limits to zero, forcing all transactions to be verified online [4].

So here too, a surge in cross-border fraud is quite predictable.

3.3 Counterfeiting

Counterfeit cards made using "skimming devices" have been the single largest problem associated with magstripe. The designers of the EMV specification have made a particular effort to include anti-counterfeiting technology in the design. They proposed counterfeit identification techniques based around both symmetric and public-key cryptography, known as *Static Data Authentication (SDA)* and *Dynamic Data Authentication (DDA)* respectively.

With DDA, each card contains a private key, which is used to sign a certificate attesting to its authenticity that incorporates fresh information from the POS terminal. Each terminal bears the VISA or Mastercard public key, and can walk along a short certificate chain to verify that the card's certificate is valid. Once the card is proven genuine, its local validation of the PIN can be trusted.

The cheaper alternative, SDA, uses only a symmetric key on the card, shared with the issuing bank. Once the card has verified the PIN, the terminal sends a summary of transaction data, and the card produces a MAC calculated over this, called a "transaction certificate". The terminal cannot know this key, as it would be too great a risk to give all terminals a key that would permit wholesale forgery if discovered.

The card can prove that it's genuine, but only if the terminal is *online*, that is, connected to the bank via a network. *Offline* machines, however, have no way of telling if the card is genuine on the spot. They can record the response the card gives during authorisation, and pass it on to the bank when they next connect (if they connect at all), but there is nothing more they can do. In fact,

SDA chipcards are even less secure with offline terminals than magnetic strip cards are! This is because the fraudster does not even need to know the PIN. It's the card's job to verify the PIN and respond with a yes/no answer to the terminal, so if the card is a counterfeit, the fraudster can program it to say "yes" no matter what PIN is entered.

Neither APACS nor any UK bank has publicly announced whether they are using Static Data Authentication (SDA) or Dynamic Data Authentication (DDA). However, we have analysed a number of UK Chip and PIN cards, and these results suggest they they have only static capability. One source at a software developer has predicted that UK banks will move to DDA over time. We expect this too, as the cost of fraud starts to climb.

3.4 EMV weaknesses

The technical specification behind Chip and PIN, EMV, is a public standard, and will come under attack. In practice it is very hard to build a complex system without introducing design errors. How and when will these be discovered during the lifespan of EMV, how serious will they be, and what are the likely consequences? Here is a brief description of one potential weakness that we have noticed. We are reluctant to test whether a vulnerability actually exists, because of the risk of legal action if we tamper with live banking systems; so we are not motivated to do a thorough search for vulnerabilities. (The crooks, of course, will be motivated.) We merely highlight some potential issues that are evident from the published specifications.

The *Cardholder Verification Method (CVM)* system determines how you will be required to prove your ownership of a card when asked to make a transaction. Your chip card and the Point-of-Sale (POS) terminal both maintain a list of methods that they find acceptable, and the order of preference. A POS terminal in a shop might typically ask for a PIN to be verified by the card, and if this fails, then go for a signature. Meanwhile, if your card is older, its CVM list might only contain 'Signature' and 'Nothing' (some terminals, such as rail ticket vending machines, can't input either a signature or a PIN, so their verification method is 'nothing', i.e., no authentication at all). The POS terminal and card negotiate together to find a mutually agreeable method.

However, it seems that this CVM list is not certified in such as way as to prevent modification. A fraudster might thus be able to take a stolen chipcard, change the CVM list using some relay device that intercepts the communications, and thus downgrade a card that requires a PIN into one that will accept signature. Alternatively, the relay device might pretend to the card to be a terminal that's happy with a signature, and to the terminal to be a card that requires a PIN and has happily accepted the bogus PIN that the criminal entered.

3.5 Middleperson attacks

One fundamental limitation of electronic payment systems is the user interface. If you go to a shop and pay by card, regardless of the technology, how do you

know how much your account is actually going to be charged? You see an amount appear on the till, but what if the display is lying to you? This problem has been understood for some time. It wasn't usually considered significant, because you can always complain later if you are charged the wrong amount.

Chip and PIN changes that. The interaction with your credit or debit card is simply done using electronic signals over the contacts with the card. These signals can be routed almost instantly anywhere in the world, just like a phone call. So when you put your card in the machine slot at a shop or restaurant, how can you be sure that it is genuinely talking to that machine?

Consider the following scenario: You go for lunch in a small restaurant in London, and pay using your chipcard at the end of the meal. What you don't know is that the waiter at the restaurant is corrupt. You ask for the bill, and the waiter goes off to fetch a handheld Chip and PIN machine that he brings over to you. Meanwhile, on the other side of town, his accomplice is loitering in a jeweller's store. The waiter sends an SMS message to his accomplice, who goes up to make a purchase. Just as you insert your card into the waiter's terminal, the accomplice puts a fake card into the jeweller's terminal. The waiter's sabotaged reader simply forwards all the traffic from your card wirelessly to the card in the reader at the jewellers, and pretends to ask you to pay for lunch. You enter the PIN, thinking you're paying for lunch, but in fact you're buying the crooks a diamond!

It is extremely hard to prevent this sort of attack, which is sometimes called a "middleperson" or "relay" attack. The attack might even be industrialised – for example, by subverting the computer that controls a network of vending machines, or by infecting point-of-sale terminals with a virus. The attacker would then tap into a continuous stream of transactions, and subvert one whenever needed. The problem is that there is no 'trusted path' – no way that a cardholder can communicate securely with the chip on their card to see exactly how much they are about to pay, and to whom.

3.6 Smartcard device security

The physical security of EMV-approved smartcards remains an open question. A destructive invasive attack to extract the secret key from a chipcard is possible, but unlikely, as the equipment required to probe out smartcard memory contents directly is expensive and difficult to use.

A more serious threat is a non-invasive attack using power analysis. Here, the crooks build a terminal that performs a series of transactions quickly on a card while measuring the current it consumes. Since chips use different amounts of power depending on the data being processed, it is in principle possible to extract the cryptographic key being used. Whether this is actually a threat depends on the number of transactions the terminal needs to do; if it a few dozen, the attack is probably feasible, while if it is a few tens of thousands, then it is probably not. The problem is that estimating the attack work factor depends on the state of the art in signal processing, which is improving all the time. This is an area of research in which we maintain a close interest.

3.7 Card not present

There is one final reason why the introduction of chipcards is unlikely to give even the modest fraud reductions achieved in France a decade ago. That is the huge volume of card-not-present transactions generated by the boom in online shopping. These have grown five-fold since 1999, and now account for 10% of all credit-card turnover. Card-not-present fraud rose 24% last year to £150.8m, almost 30% of the total.

Chip cards at present do nothing to reduce this. Worse, electronic commerce may be undermined by a climate in which banks increasingly try to push the costs and risks of fraud on to merchants and customers. (One of us has experience of a bank help-desk being uninterested in a disputed credit-card debit, telling us to pursue the dispute with the merchant.) Given the huge amounts that banks earn as their cut from online transactions, this may be shortsighted. Or perhaps British banks reckon that the better-regulated US financial markets will continue to facilitate the growth of global e-commerce, on which they can free ride.

4 Conclusions

Chip and PIN is being heavily spun in the UK media. Whilst our comments might help redress the imbalance of information, they cannot reduce the spin. The customer thus has a hard time determining the facts. Sadly, what we expect is that as banks manage to externalise their risks, their incentives to protect the system will be eroded and the system will become less secure. In the medium term, we might hope for better regulation – which should place the liability for security failure squarely with those principals who are in a position to manage the security risks. (The banking industry might also contemplate the potential political costs of risk-dumping in an age when corporate social responsibility is all the rage.)

Meanwhile the UK banking industry ought to look carefully at what they hope to achieve from chipcards, and make sure their estimates of attackers' capabilities are realistic. As late adopters of smartcard technology, they may not have as long a grace period from attack as the French banks did. It is particularly ominous that there appear to be card fraud gangs in the UK with eight-figure turnover, and the technical sophistication to engineer ATM false fronts that are so good as to be almost undetectable. At the very least, banks should have a well-thought-out plan to move from SDA to DDA technology.

In the longer term, some fresh thinking may be needed about payment system architectures. The lack of trusted path creates a fundamental limitation to the credit card form-factor for payments, and maybe some new ideas are needed.

References

1. M Brignall, "Chip and PIN Helps Push Bank and Credit Card Fraud to £505m", in The Guardian, March 8, 2005, at `http://www.guardian.co.uk/business/story/0,,1432601,00.html`

2. N Bohm, I Brown, B Gladman, "Electronic Commerce: Who Carries the Risk of Fraud?", Journal of Information Law and Technology, October 2000, at http://elj.warwick.ac.uk/jilt/00-3/bohm.html
3. Phantom Withdrawals Website, http://www.phantomwithdrawals.com
4. "Card Fraud: Banking's Boom Sector", Banking Automation Bulletin for Europe, March 1992, pp 1–5
5. "Why Information Security is Hard – An Economic Perspective", in *Proceedings of the Seventeenth Computer Security Applications Conference* IEEE Computer Society Press (2001), ISBN 0-7695-1405-7, pp 358–365
6. H Varian, "Managing Online Security Risks", New York Times, Jun 1, 2000; at http://www.sims.berkeley.edu/~hal/people/hal/NYTimes/2000-06-01.html
7. Banking Code of Practice 2005, http://www.bankingcode.org.uk/pdfdocs/BANKING%20CODE.pdf
8. Diner's Club Pty. SA vs. A. and V. Singh , High Court of Durban, South Africa, 2001–2003