CHAPTER

# 20

# Telecom System Security

*I rarely had to resort to a technical attack. Companies can spend millions of dollars toward technological protections and that's wasted if somebody can basically call someone on the telephone and either convince them to do something on the computer that lowers the computer's defenses or reveals the information they were seeking.*

**— Kevin Mitnick**

*There are two kinds of fools. One says, "This is old, therefore it is good." The other one says, "This is new, therefore it is better".*

**— Dean William Inge**

## 20.1 Introduction

The protection of telecommunications systems is an important case study for a number of reasons. First, many distributed systems rely on the fixed or mobile phone network in ways that are often not obvious, and the dependability of these networks is declining. For example, POTS — the 'plain old telephone system' — typically required exchanges to have backup generators with enough diesel to survive a six-week outage in the electricity supply, while cellular systems typically use batteries that will last at most 48 hours. What's worse, the electicity companies rely on mobile phones to direct their engineers when repairing faults. When people realised that this could cause serious problems where outages lasted more than two days, the electricity companies started buying satellite phones as a backup.

Second, the history of telecomms security failures is very instructive. Early attacks were carried out on phone companies by enthusiasts ('phone phreaks') to get free calls; then the phone system's vulnerabilities started to be

exploited by crooks to evade police wiretapping; then premium rate calls were introduced, which created the motive for large-scale fraud; then when telecomms markets were liberalized, some phone companies started conducting attacks on each other's customers; and some phone companies have even attacked each other. At each stage the defensive measures undertaken were not only very expensive but also tended to be inadequate for various reasons. The same pattern is repeating with the Internet — only with history much speeded up. A number of the policy issues that arose with wireline phones, such as wiretapping, have played themselves out again on the Internet.

Finally, the latest developments in telecomms, from VOIP at the consumer level to the adoption of IP networks as the underlying technology by telecomms providers, create further interactions. Skype's two-day failure in August 2007 following Microsoft's Patch Tuesday is just one case in point. Systems are becoming much more complex and interdependent, and are generally not being engineered to the old standards.

## 20.2 Phone Phreaking

The abuse of communication services goes back centuries. Before Sir Rowland Hill invented the postage stamp, postage was paid by the recipient. Unsolicited mail became a huge problem — especially for famous people — so recipients were allowed to inspect a letter and reject it rather than paying for it. People soon worked out schemes to send short messages on the covers of letters which their correspondents rejected. Regulations were brought in to stop this, but were never really effective [979].

A second set of abuses developed with the telegraph. The early optical telegraphs, which worked using semaphores or heliographs, were abused by people to place foreknowledge bets on races; if you could learn which horse had won before the bookmaker did, you were well away. People would bribe operators, or 'hack the local loop' by observing the last heliograph station through a telescope. Here too, attempts to legislate the problem away were a failure [1215]. The problems got even more widespread when the electric telegraph brought costs down; the greater volumes of communication, and the greater flexibility that got built into and on top of the service, led to greater complexity and volume of abuse.

The telephone was to be no different.

### 20.2.1 Attacks on Metering

Early metering systems were wide open to abuse.

- In the 1950's, the operator in some systems had to listen for the sound of coins dropping on a metal plate to tell that a callbox customer had paid,

so some people acquired the knack of hitting the coinbox with a piece of metal that struck the right note.

- Initially, the operator had no way of knowing which phone a call had come from, so she had to ask the caller his number. He could give the number of someone else — who would be charged. This was risky to do from your own phone, so people did it from call boxes. Operators started calling back to verify the number for international calls, so people worked out social engineering attacks ('This is IBM here, we'd like to book a call to San Francisco and because of the time difference can our Managing Director take it at home tonight? His number's xxx-yyyy'). So call box lines had a feature added to alert the operator. But in the UK implementation, there was a bug: a customer who had called the operator from a callbox could depress the rest for a quarter second or so, whereupon he'd be disconnected and reconnected (often to a different operator), with no signal this time that the call was from a callbox. He could then place a call to anywhere and bill it to any local number.
- Early systems also signalled the entry of a coin by one or more pulses, each of which consisted of the insertion of a resistance in the line followed by a brief open circuit. At a number of colleges, enterprising students installed 'magic buttons' which could simulate this in a callbox in the student union so people could phone for free. (The bill in this case went to the student union, for which the magic button was not quite so amusing.)

Attacks on metering mechanisms continue. Many countries have moved their payphones to chip cards in order to cut the costs of coin collection and vandalism. Some of the implementations have been poor (as I remarked in the chapter on tamper resistance) and villains have manufactured large quantities of bogus phone cards. Other attacks involve what's called *clip-on*: physically attaching a phone to someone else's line to steal their service.

In the 1970's, when international phone calls were very expensive, foreign students would clip their own phone on to a residential line in order to call home; an unsuspecting home owner could get a huge bill. Despite the fact that in most countries the cable was the phone company's legal responsibility up to the service socket in the house, phone companies were mostly adamant that householders should pay and could threaten to blacklist them if they didn't. Now that long distance calls are cheap, the financial incentive for clip-on fraud has largely disappeared. But it's still enough of a problem that the Norwegian phone company designed a system whereby a challenge and response are exchanged between a wall-socket mounted authentication device and the exchange software before a dial tone is given [673].

Clip-on fraud had a catastrophic effect on a family in Cramlington, a town in the North East of England. The first sign they had of trouble was hearing

a conversation on their line. The next was a visit from the police who said there'd been complaints of nuisance phone calls. The complainants were three ladies, all of whom had a number one digit different from a number to which this family had supposedly made a huge number of calls. When the family's bill was examined, there were also calls to clusters of numbers that turned out to be payphones; these had started quite suddenly at the same time as the nuisance calls. When the family had complained later to the phone company about a fault, their connection was rerouted and this had solved the problem.

But the phone company denied the possibility of a tap, despite the report from their maintenance person which noted that the family's line had been tampered with at the distribution cabinet. (The phone company later claimed this report was in error.) It turned out that a drug dealer had lived close by, and it seemed a reasonable inference that he'd tapped their line in order to call his couriers at the payphones. By using an innocent family's phone line instead of his own, he not only saved on the phone bill, but also had a better chance of evading police surveillance. But both the police and the local phone company refused to go into the house where the dealer had lived, claiming it was too dangerous — even though the dealer had by now got six years in jail. The Norwegian phone company declined an invitation to testify about clip-on for the defence. The upshot was that the subscriber was convicted of making harassing phone calls, in a case widely believed to have been a miscarriage of justice. There was discussion at the time about whether the closing of ranks between the phone company and the police was a bureaucratic reflex — or something more sinister. Since 9/11, it's emerged that many phone companies have been giving the police easy access to systems for years, often without warrants, in return for favours. The logical consequence was a policy of covering up anything that could stray into this territory — even if the coverup caused collateral damage. I'll discuss all this later in the third part of this book.

Stealing dial tone from cordless phones is another variant on the theme. In the 1990s, this became so widespread in Paris that France Telecom broke with phone company tradition and announced that it was happening, claiming that the victims were using illegally imported cordless phones which were easy to spoof [745]. Yet to this day I am unaware of any cordless phones — authorised or not — with decent air link authentication. The new digital cordless phones use the DECT standard which allows for challenge-response mechanisms [1283] but the equipment sold so far seems to simply send a handset serial number to the base station.

Social engineering is also widespread. A crook calls you pretending to be from AT&T security and asks whether you made a large number of calls to Peru on your calling card. When you deny this, he says that they were obviously fake and, in order to reverse out the charges, can he confirm that your card number is 123-456-7890-6543? No, you say (if you're not really alert),

it's 123-456-7890-5678. Now 123-456-7890 is your phone number and 5678 your password, so you've just given that caller the ability to bill calls to you.

The growth of premium rate phone services during the 1990s also led to scamsters developing all sorts of tricks to get people to call them: pager messages, job ads, fake emergency messages about relatives, 'low cost' calling cards with 0900 access numbers, you name it. (In fact, the whole business of tricking people into calling expensive premium numbers enabled crooks to develop a lot of the techniques we now see used in email as part of phishing attacks.) The 809 area code for the Caribbean used to be a favourite cover for crooks targeting U.S. subscribers; many people weren't aware that 'domestic' numbers (numbers within the USA's +1 international direct dialling code) extend outside the relatively cheap USA (and Canada). Even though many people have now learned that +1 809 is 'foreign' and more expensive, the introduction of still more Caribbean area codes, such as +1 345 for the Cayman Islands, has made it even harder to spot premium rate numbers.

Phone companies advised their customers 'Do not return calls to unfamiliar telephone numbers' and 'Beware of faxes, e-mail, voice mail and pages requesting a return call to an unfamiliar number' [22] — but how practical is that? Just as banks now train their customers to click on links in marketing emails and thus make them vulnerable to phishing attacks, so I've had junk marketing calls from my phone company — even though I'm on the do-not-call list. And as for governments, they have tended to set up weak regulators to oversee phone system abuses at home, and avoid anything that might get them involved in trying to regulate premium rate scams overseas. For example, they let phone companies harass their customers into paying bills for overseas services even when they knew that the overseas traffic was fraudulent.

Indeed, by no means all premium-rate scams involved obviously dodgy companies running sex lines; as I write in 2007, the British press are full of stores about how TV companies rip off their customers by getting them to call premium lines in order to compete, and vote, in all sorts of shows. It's turned out that many of these are recorded, so the calls are totally futile; and even the live ones are managed so that people who live in the wrong part of the country or speak with the wrong accent have no chance. The authorities tried to leave this to 'self-regulation' and on-air apologies from TV bosses, until a public outcry (whipped up by their competitors) led to the file being sent to the police in October 2007. It's a recurring pattern that the biggest scams are often run by 'respectable' companies rather than by Russian gangsters.

### 20.2.2 Attacks on Signaling

The term 'phone phreaking' refers to attacks on signaling as well as pure toll fraud. Until the 1980s, phone companies used signalling systems that worked *in-band* by sending tone pulses in the same circuit that carried the speech. The

first attack I've heard of dates back to 1952, and by the mid-to-late 1960s many enthusiasts in both America and Britain had worked out ways of rerouting calls. One of the pioneers, Joe Engresia, had perfect pitch and discovered as a child that he could make free phone calls by whistling a tone he'd heard in the background of a long-distance call. His less gifted colleagues typically used home-made tone generators, of which the most common were called *blue boxes*. The trick was to call an 0800 number and then send a 2600Hz tone that would *clear down* the line at the far end — that is, disconnect the called party while leaving the caller with a trunk line connected to the exchange. The caller could now enter the number he really wanted and be connected without paying. Phone phreaking was one of the roots of the computer hacker culture that took root in the Bay Area and was formative in the development and evolution of personal computers [835]. For example, Steve Jobs and Steve Wozniak first built blue boxes before they diversified into computers [502].

Phone phreaking started out with a strong ideological element. In those days most phone companies had monopolies. They were large, faceless and unresponsive. In America, AT&T was such an abusive monopoly that the courts eventually broke it up; most phone companies in Europe were government departments. People whose domestic phone lines had been involved in a service theft found they were stuck with the charges. If the young man who had courted your daughter was (unknown to you) a phone phreak who hadn't paid for the calls he made to her, you would suddenly find the company trying to extort either the young man's name or a payment. Phone companies were also aligned with state security. Phone phreaks in many countries discovered signalling codes or switch features that would enable the police or the spooks to tap your phone from the comfort of their desks, without having to send out a lineman to install a wiretap. Back in the days of Vietnam and student protests, this was inflammatory stuff. Phone phreaks were counterculture heroes, while phone companies were hand-in-hand with the forces of darkness.

As there was no way to stop blue-box type attacks so long as telephone signalling was carried in-band, the phone companies spent years and many billions of dollars upgrading exchanges so that the signaling was moved out-of-band, in separate channels to which the subscribers had no easy access. Gradually, region by region, the world was closed off to blue box attacks. There are still a few places left. For example, the first time that USAF operations were disrupted by an 'information warfare' attack by noncombatants was in 1994 when two British hackers broke into the Rome Air Force Base via an analog link through an ancient phone system in Argentina which they used to hold up investigators [1202]. There's also an interesting legacy vulnerability in wiretapping systems: common phone-tapping equipment was designed to be backwards compatible with in-band signalling, with the result that you can evade surveillance by using a blue box to convince the police equipment that

you've hung up. The telephone exchange ignores this signal, so you remain on the phone but with the police recording stopped [1151].

But to defeat a modern telephone network — as opposed to its law-enforcement add-ons — different techniques are needed.

### 20.2.3 Attacks on Switching and Configuration

The second wave of attacks targeted the computers that did the switching. Typically these were Unix machines on a LAN in the exchange, which also had machines with administrative functions such as scheduling maintenance. By hacking one of these less well guarded machines, a phreak could go across the LAN and break into the switching equipment — or in to other secondary systems such as subscriber databases. For a survey of PacBell's experience of this, see [271]; for Bellcore's, see [722].

Using these techniques, unlisted phone numbers could be found, calls could be forwarded without a subscriber's knowledge, and all sorts of mischief became possible. A Californian phone phreak called Kevin Poulsen got root access to many of PacBel's switches and other systems in 1985–88: this apparently involved burglary as much as hacking (he was eventually convicted of conspiring to possess fifteen or more counterfeit, unauthorized and stolen access devices.) He did petty things like obtaining unlisted phone numbers for celebrities and winning a Porsche from Los Angeles radio station KIIS-FM. Each week KIIS would give a Porsche to the 102nd caller, so Poulsen and his accomplices blocked out all calls to the radio station's 25 phone lines save their own, made the 102nd call and collected the Porsche. He was also accused of unlawful wiretapping and espionage; these charges were dismissed. In fact, the FBI came down on him so heavily that there were allegations of an improper relationship between the agency and the phone companies, along the lines of 'you scratch our backs with wiretaps when needed, and we'll investigate your hacker problems' [472].

Although the unauthorized wiretapping charges against Poulsen were dismissed, the FBI's sensitivity does highlight the possibility that attacks on phone company computers can be used by foreign intelligence agencies to conduct remote wiretaps. Some of the attacks mentioned in [271] were from overseas, and the possibility that such tricks might be used to crash the whole phone system in the context of an information warfare attack has for some years worried the NSA [495, 754]. Countries that import their telephone exchanges rather than building their own are in an even worse position; a prudent nations will assume that its telephone switchgear has vulnerabilities known to the government of the country from which they bought it. (It was notable that during the invasion of Afghanistan in 2001, Kabul had two exchanges: an old electromechanical one and a new electronic one. The USAF bombed only the first of these.)

But although high-tech attacks do happen, and newspaper articles on phone phreaking tend to play up the 'evil hacker' aspects, most real attacks are much simpler. Many involve insiders, who deliberately misconfigure systems to provide free calls from (or through) favored numbers. This didn't matter all that much when the phone company's marginal cost of servicing an extra phone call was near zero, but with the modern proliferation of value-added services, people with access to the systems can be tempted to place (or forge) large numbers of calls to accomplices' sex lines. Deregulation, and the advent of mobile phones, have also made fraud serious as they give rise to cash payments between phone companies [317]. Insiders also get up to mischief with services that depend on the security of the phone network. In a hack reminiscent of Poulsen, two staff at British Telecom were dismissed after they each won ten tickets for Concorde from a phone-in offer at which only one randomly selected call in a thousand was supposed to get through [1266].

As for outsiders, the other 'arch-hacker' apart from Poulsen was Kevin Mitnick, who got arrested and convicted following a series of break-ins, many of which involved phone systems and which made him the target of an FBI manhunt. They initially thought that he was a foreign agent who was abusing the U.S. phone system in order to wiretap sensitive U.S. targets. As I mentioned in Chapter 2, he testified after his release from prison that almost all of his exploits had involved social engineering. He came out with the quote at the head of this chapter: 'Companies can spend millions of dollars toward technological protections and that's wasted if somebody can basically call someone on the telephone and either convince them to do something on the computer that lowers the computer's defenses or reveals the information they were seeking' [895]. So phone company systems are vulnerable to careless insiders as well as malicious insiders — just like hospital systems and many others we've discussed.

A worrying recent development is the emergence of switching exploits by organisations. The protocols used between phone companies to switch calls — notably 5ESS — aren't particularly secure, as the move from in-band to out-of-band signaling was supposed to restrict access to trusted parties. But once again, changing environments undermine security assumptions. Now that there are many entrepreneurial phone companies rather than a handful of large ones, all sorts of people have access to the switching. An example is location service. This is provided for a fee by mobile networks; you can register your child's mobile, or your employees' mobiles, and trace them through a website. One entrepreneur undercut this service in the UK by using the switching interface exported by a local telco. While such issues can generally be resolved by contracts, litigation and regulation, there remains a lingering worry that attackers might bring down a telco by exploiting access to its switching and network management. This worry increases as telcos migrate

their networks to IP, and they start to converge with VOIP services that give users access to the IP layer. I'll return to VOIP later.

### 20.2.4 Insecure End Systems

After direct attacks on the systems kept on phone company premises, the next major vulnerabilities of modern phone systems are insecure terminal equipment and feature interaction.

There have been a number of cases where villains exploited people's answering machines. The same technique can be used for at least two different purposes: the relatively innocuous one of tricking someone into dialling a premium rate number, or the somewhat more sinister one of using their answering machine as a covert remailer for a voicemail message. The problem arises from phone company switches that give you dial tone twelve seconds after the other party hangs up. So a terrorist who wants to send an untraceable instruction to a colleague can record on your answering machine thirteen blank seconds, followed by the tones needed to dial his colleague's number and the secret message. He then calls again, gets the machine to play back its messages, and hangs up on it.

But the really big frauds using insecure end systems are directed against companies and government departments. Attacks on corporate *private branch exchange* systems (PBXes) had become big business by the mid-1990's and cost business billions of dollars a year [322]. PBXes are usually supplied with facilities for *refiling* calls, also known as *direct inward system access* (DISA). The typical application is that the company's sales force can call in to an 0800 number, enter a PIN or password, and then call out again taking advantage of the low rates a large company can get for long distance calls. As you'd expect, these PINs become known and get traded by villains [911]. The result is known as *dial-through* fraud.

In many cases, the PINs are set to a default by the manufacturer, and never changed by the customer. In other cases, PINs are captured by crooks who monitor telephone traffic in hotels anyway on order to steal credit card numbers; phone card numbers and PBX PINs are a useful sideline. Many PBX designs have fixed engineering passwords that allow remote maintenance access, and prudent people reckon that any PBX will have at least one back door installed by the manufacturer to give easy access to law enforcement and intelligence agencies (it's said, as a condition of export licensing). Of course such features get discovered and abused. In one case, the PBX at Scotland Yard was compromised and used by villains to refile calls, costing the Yard a million pounds, for which they sued their telephone installer. The crooks were never caught [1244]. This was particularly poignant, as one of the criminals' motivations in such cases is to get access to communications that will not be tapped. Businesses who're the victims of such crimes nevertheless find

the police reluctant to investigate, and one reason for this is that the phone companies aren't particularly helpful — presumably as they don't like having their bills disputed [1088].

In another case, Chinese gangsters involved in labor market racketeering — smuggling illegal immigrants from Fujian, China, into Britain where they were put to work in sweatshops, on farms and so on — hacked the PBX of an English district council and used it to refile over a million pounds' worth of calls to China. The gang was tackled by the police after a number of its labourers died; they were picking shellfish in Morecambe Bay when the tide came in and drowned them. The council had by now discovered the discrepancy in its phone bills and sued the phone company for its money back. The phone company argued that it wasn't to blame, even although it had supplied the insecure PBX. Here, too, the gangsters were interested not just in saving money but in evading surveillance. (Indeed, they routed their calls to China via a compromised PBX in Albania, so that the cross-border segment of the call, which is most likely to be monitored by the agencies, was between whitelisted numbers; the same trick seems to have been used in the Scotland Yard case, where the crooks made their calls via the USA.)

Such cases apart, dial-through fraud is mostly driven by premium rate services: the main culprits are crooks who are in cahoots with premium line owners. Most companies don't understand the need to guard their 'dial tone' and don't know how to even if they wanted to. PBXes are typically run by company telecomms managers who know little about security, while the security manager often knows little about phones. This is changing, albeit slowly, as VOIP technologies take over and the company phone network merges with the data network.

Exploits of insecure end-systems sometimes affect domestic subscribers too. A notorious case was the Moldova scam. In 1997, customers of a porn site were told to download a 'viewer' program that dropped their phone line and connected them to a phone number in Moldova (having turned off their modem speakers so they wouldn't notice). The new connection stayed up until they turned off their computers; thousands of subscribers incurred hundreds of thousands of dollars in international long distance charges at over $2 per minute. Their phone companies tried to collect this money but there was an outcry. Eventually the subscribers got their money back, and the Federal Trade Commission enjoined and prosecuted the perpetrators [456]. Since then there have been a number of copycat scams [870]; but as more and more people move to cable modems or broadband, and their PCs are no longer able to dial out on the plain old telephone system, this kind of abuse is getting less common. The latest twist is premium-rate mobile malware: in 2006, for example, the Red Browser worm cashed out by sending $5 SMSs to Russia [633].

Premium rate scams and anonymous calling are not the only motives. Now that phones are used more and more for tasks such as voting, securing entry into apartment buildings, checking that offenders are observing their parole terms, and authenticating financial transactions, more motives are created for ever more creative kinds of mischief, and especially for hacks that defeat caller line ID. For example, caller-line ID hacks make middleperson attacks on payment systems easier; SMS spoofing and attacks on the SS7 signaling in the underlying network can have similar effects [897].

And sometimes attacks are conducted by upstanding citizens for perfectly honorable motives. A neat example, due to Udi Manber, is as follows. Suppose you have bought something which breaks, and the manufacturer's helpline only has an answering machine. To get service, you have to take the answering machine out of service. This can often be done by recording its message, and playing it back so that it appears as the customer message. With luck the machine's owner will think it's broken and it'll be sent off for maintenance.

### 20.2.5 Feature Interaction

More and more cases of telephone manipulation involve feature interaction.

- Inmates at the Clallam Bay Correctional Center in Washington state, who were only allowed to make collect calls, found an interesting exploit of a system which the phone company ('Fone America') introduced to handle collect calls automatically. The system would call the dialled number and a synthesised voice would say: 'If you will accept a collect call from . . . (name of caller) . . . please press the number 3 on your telephone twice'. Prisoners were supposed to state their name for the machine to record and insert. The system had, as an additional feature, the ability to have the greeting delivered in Spanish. Inmates did so, and when asked to identify themselves, said 'If you want to hear this message in English, press 33'. This worked often enough that they could get through to corporate PBXes and talk the operator into giving them an outside line. The University of Washington was hit several times by this scam [476].
- A number of directory-enquiry services will connect you to the number they've just given you, as a service to motorists who can't dial while driving. But this can often be used to defeat mechanisms that depend on endpoint identification. Adulterers use it to prevent their spouses seeing lovers' numbers on the family phone bill, and naughty children use it to call sex lines despite call barring [977].

- Call forwarding is a source of many scams. In the old days, it was used for pranks, such as kids social-engineering a phone company operator to forward calls for someone they didn't like to a sex line. Nowadays, it's quite often both professional and nasty. For example, a fraudster may tell a victim to confirm her phone number with the bank by dialing a sequence of digits — which forwards her incoming calls to a number controlled by the attacker. So the bank's callback machanisms are defeated when the customer isn't aware that a certain sequence of dialed numbers can alter the behavior of her phone.
- British Telecom launched a feature called 'Ringback'. If you dial an engaged number, you can then enter a short code and as soon as the called number is free, both your phone and theirs will ring. The resulting call is billed to you. However, when you used ringback used from a pay phone, it was the phone's owner who ended up with the bill. People with private pay phones, such as pub landlords and shopkeepers, lost a lot of money, which the phone company was eventually obliged to refund [652].
- Conference calls also cause a lot of trouble. For example, football hooligans in some countries are placed under a curfew that requires them to be at home during a match, and to prove this by calling the probation service, which verifies their number using caller ID. The trick is to get one of your kids to set up a conference call with the probation service and the mobile you've taken to the match. If the probation officer asks about the crowd noise, you tell him it's the TV and you can't turn it down or your mates will kill you. (And if he wants to call you back, you get your kids to forward the call.)

This brings us to the many problems with mobile phones.

## 20.3 Mobile Phones

Since their beginnings as an expensive luxury in the early 1980s, mobile phones have become one of the big technological success stories. By 2007, we now have over a billion subscribers; it's said that over a billion phones will be sold this year and the total subscriber base may rise to two billion. In developed countries, most people have at least one mobile, and many new electronic services are being built on top of them. Scandinavia has led here: you get a ferry ticket in Helsinki by sending a text message to the vending machine, and you get a can of Coke the same way. You can also scan a bar code at a bus stop with your phone camera, and get sent a text message 90 seconds before the next bus arrives; that way you don't have to stand out in the snow.

Growth is rapid in developing countries too, where the wireline network is often dilapidated and people used to wait years for phone service to be installed. In some places it's the arrival of mobile phone service that's connected villages to the world. Criminals also make heavy use of mobiles, and not just for communications: and in large tracts of the third world, mobile phone units have become a de facto currency. If you get kidnapped in Katanga, the kidnappers will tell your relatives in Kinshasa to buy mobile phone units and text them the magic numbers. In developed countries, the criminal interest is largely in communications, and most police wiretaps are now on mobile numbers.

So mobile phones are very important to the security engineer, both as part of the underlying infrastructure and as a channel for service delivery. They can also teach us a lot about fraud techniques and countermeasures.

### 20.3.1 Mobile Phone Cloning

The first generation of mobile phones used analog signals with no real authentication. The handset simply sent its serial numbers in clear over the air link. (In the U.S. system, there were two of them: one for the equipment, and one for the subscriber.) So villains built devices to capture these numbers from calls in the neighborhood. (I've even seen a phone that a student had reprogrammed to do this by a simple software hack.) One of the main customers was the *call-sell operation* that would steal phone service and resell it cheaply, often to immigrants or students who wanted to call home. The call-sell operators would hang out at known pitches with cloned mobiles, and their customers would queue up to phone home for a few dollars.

So a black market developed in phone serial numbers. The call-sell market was complemented by the market for anonymous communications for criminals: enterprising engineers built mobile phones which used a different identity for each call. Known as *tumblers*, these were particularly hard for the police to track [636]. The demand for serial numbers grew rapidly and satisfying it was increasingly difficult, even by snooping at places like airports where lots of mobiles were turned on. So prices rose, and as well as passive listening, active methods started to get used.

Modern mobile phones are cellular, in that the operator divides the service area up into cells, each covered by a base station. The mobile uses whichever base station has the strongest signal, and there are protocols for handing off calls from one cell to another as the customer roams. (For a survey of mobile phone technology, see [1061].) The active attack consists of a fake base station, typically at a place with a lot of passing traffic such as a freeway bridge. As phones pass by, they hear a stronger base station signal and attempt to register by sending their serial numbers.

A number of mechanisms were tried to cut the volume of fraud. Most operators developed or bought intrusion detection systems, which watch out

for suspicious patterns of activity. A number of heuristics were developed. For example, genuine mobiles which roam and call home regularly, but then stop calling home, have usually been stolen; other indicators include too-rapid movement (such as calls being made from New York and LA within an hour of each other) and even just a rapid increase in call volume or duration.

In the chapter on electronic warfare, I mentioned RF fingerprinting — a formerly classified military technology in which signal characteristics that arise from manufacturing variability in the handset's radio transmitter are used to identify individual devices and tie them to the claimed serial numbers [534]. Although this technique works — it was used by Vodafone in the UK to almost eliminate cloning fraud from analogue mobiles — it is expensive as it involves modifying the base stations. (Vodafone also used an intrusion detection system that tracked customer call patterns and mobility, described in [1283]; their competitor Cellnet simply blocked international calls from analogue mobiles, which helped move its high value customers to its more modern digital network.) Another proposed solution was to adopt a cryptographic authentication protocol, but there are limits on how much can be done without changing the whole network. For example, one can use a challenge-response protocol to modify the serial number [485]. But many of the mechanisms people proposed to fortify the security of analog cellular phones have turned out to be weak [1305].

Eventually the industry decided to upgrade to a new digital system. Revenue protection was an issue, but far from the only one; digital systems offered more efficient use of bandwidth, and a whole host of new features — including easier international roaming (important in Europe with lots of small countries jammed close together), and the ability to send and receive short text messages. (Text messages were almost an afterthought; the designers didn't realise they'd be hugely popular.) From the operators' viewpoint, the move to standard digital equipment cut costs and enabled rapid, wide-scale deployment.

### 20.3.2 GSM Security Mechanisms

The second generation of mobile phones adopted digital technology. Most handsets worldwide use the *Global System for Mobile Communications*, or GSM, which was designed from the start to facilitate international roaming; it was founded when 15 companies signed up to the GSM Association in 1987, and service was launched in 1992. As of 2007, the GSM system extends to over two billion handsets in over 200 countries; a typical developed country has more handsets in service than it has people [133]. The USA, Japan, Korea and Israel had different second-generation digital standards (although the USA has GSM service too). Since about 2001, most countries also have a third-generation service, which I'll describe in the next section.

The designers of GSM set out to secure the system against cloning and other attacks: their goal was that GSM should be at least as secure as the wireline system. What they did, how they succeeded and where they failed, make an interesting case history.

The authentication protocols are described in a number of places, such as [232] (which also describes the mechanisms in an incompatible U.S. system). The industry initially tried to keep secret the cryptographic and other protection mechanisms which form the core of the GSM protocols. This didn't work: some eventually leaked and the rest were discovered by reverse engineering. I'll describe them briefly here.

Each network has two databases, a *home location register* (HLR) that contains the location of its own mobiles, and a *visitor location register* (VLR) for the location of mobiles which have roamed in from other networks. These databases enable incoming calls to be forwarded to the correct cell.

The handsets are commodity items. They are personalised using a *subscriber identity module* (SIM) — a smartcard you get when you sign up for a network service, and which you load into your handset. The SIM can be thought of as containing three numbers:

1. there may be a personal identification number that you use to unlock the card. In theory, this stops stolen mobiles being used. In practice, many networks set an initial PIN of 0000, and most users never change it or even use it;
2. there's an *international mobile subscriber identification* (IMSI), a unique number that maps on to your mobile phone number;
3. finally there is a *subscriber authentication key* $K_i$, a 128-bit number that serves to authenticate that IMSI and is known to your home network.

Unlike the banks, which used master keys to generate PINs, the phone companies decided that master keys were too dangerous. So instead of diversifying a master key $KM$ to manufacture the authentication keys as $K_i = \{\text{IMSI}\}_{KM}$, the keys are generated randomly and kept in an authentication database attached to the HLR.

The protocol used to authenticate the handset to the network runs as follows (see Figure 20.1). On power-up, the SIM may request the customer's PIN; if this isn't configured, or once it's entered correctly, the SIM emits the IMSI, which the handset sends to the nearest base station. It's relayed to the subscriber's HLR, which generates five *triplets*. Each triplet consists of:

- RAND, a random challenge;
- SRES, a response; and
- $K_c$, a ciphering key.

The relationship between these values is that RAND, encrypted under the SIM's authentication key $K_i$, gives an output which is SRES concatenated with $K_c$:

$$\{RAND\}_{K_i} = (SRES|K_c)$$

The standard way to do this encryption is using a one-way function called Comp128, or A3/A8. (A3 refers to the SRES output and A8 to the $K_c$ output). Comp128 is a hash function with 40 rounds, described in detail in [226], and like most proprietary algorithms that were designed in the shadows and fielded quietly in the 1980s and 90s, it turns out to be vulnerable to cryptanalysis. The basic design of the function is much like in Figure 5.9 — each round consists of table lookups followed by mixing. There are five tables, with 512, 256, 128, 64 and 32 byte entries each, and the hash function uses them successively in each block of five rounds; there are eight of these blocks. This may seem very complex, but once its design became public, a vulnerability was soon noticed. Four of the bytes at the output of the second round depend only on the value of the same bytes of the input. This four-byte to four-byte channel is called a *narrow pipe* and it's possible to probe it by tweaking input bytes until you detect a collision. Once all the details have been worked out, it turns out that you need about 150,000 suitably chosen challenges to extract the key [1306, 1307]. The effect is that given access to a SIM issued by a network that uses Comp128, the authentication key can be extracted in several hours using software that is now freely available.

This attack is yet another example of the dangers of using a secret crypto primitive that has been evaluated by only a few friends; the cryptanalytic techniques necessary to find the flaw were well known [1287] and if Comp128 had been open to hostile public scrutiny, the flaw would most probably have been found. Thankfully, a phone company can replace Comp128 with a proper hash function such as SHA-256 without affecting anyone else; the hash function is present only in the SIM cards it issues to its customers and the software at its HLR. In any case, there don't seem to be any industrial-scale attacks based on a vulnerable hash function; the normal user doesn't have any incentive to crack his $K_i$ out from his SIM as it doesn't let him bill calls to (or steal calls from) anyone else.

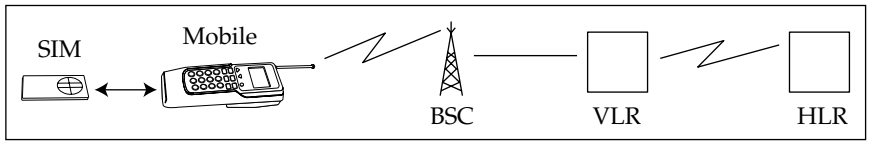


**Figure 20.1:** GSM authentication system components

Anyway, the triplets are sent to the base station, which now presents the first RAND to the mobile. It passes this to the SIM, which computes SRES. The mobile returns this to the base station and if it's correct the mobile and the base station can now communicate using the ciphering key $K_c$. So the whole authentication protocol runs as in Figure 20.2.

| | |
|---|---|
| SIM → HLR | IMSI |
| HLR → BSC | (RAND, SRES, $K_c$), . . . |
| BSC → SIM | RAND |
| SIM → BSC | SRES |
| BSC → mobile | $\{\text{traffic}\}_{K_c}$ |

**Figure 20.2:** GSM authentication protocol

There are several vulnerabilities in this protocol. In most countries the communications between base stations and the VLR pass unencrypted on microwave links[1]. So an attacker could send out an IMSI of his choice and then intercept the triplet on the microwave link back to the local base station. A German mobile operator, which offered a reward of 100,000 Deutschmarks to anyone who could bill a call to a mobile number whose SIM card was held in their lawyer's office, backed down when asked for the IMSI [44].

Second, triples can be replayed. An unscrupulous foreign network can get five triples while you are roaming on it and then keep on reusing them to allow you to phone as much as you want. (Home networks could stop this by contract but they don't.) So the visited network doesn't have to refer back to your home network for further authorisation — and even if they do, it doesn't protect you as the visited network might not bill you for a week or more. So your home network can't reliably shut you down while you roam and it may still be liable to pay the roamed network the money. Dishonest networks also defraud roaming customers by *cramming* — by creating false billing records, a practice I'll describe in more detail later. So even if you thought you'd limited your liability by using a pre-paid SIM, you might still end up with your network trying to collect money from you. This is why, to enable roaming with a pre-paid SIM, you're normally asked for a credit card number. You can end up being billed for more than you expected.

[1]The equipment can encrypt traffic, but the average phone company has no incentive to switch the cryptography on. Indeed, as intelligence agencies often monitor the backhaul near major switching nodes as an efficient means of getting warrantless access to traffic, a phone company that did switch on the crypto might find that persons unknown started jamming the link to make them stop.

The introduction of GSM caused significant shifts in patterns of crime generally. The authentication mechanisms made phone cloning difficult, so the villains switched their modus operandi to buying phones using stolen credit cards, using stolen identities or bribing insiders [1352]. Robbery was another issue. We've had a spate of media stories in Britain about kids being mugged for their phones. Mobile phone crime did indeed increase 190% between 1995 and 2002, but to keep this in context, the number of subscribers went up 600% in the same period [583]. Some of the theft is bullying — kids taking smaller kids' phones; some is insurance fraud by subscribers who've dropped their phones in the toilet and report them as stolen as their insurance doesn't cover accidental damage; but there is a hard core of theft where muggers take phones and sell them to fences. Many of the fences either work at mobile phone shops that have authorised access to tools for reprogramming the International Mobile Equipment Identifier (IMEI), the serial number in the handset, or else have links to organised criminals who ship the handsets abroad. Things are worse in Brazil, where kidnapping is endemic: there are now fake kidnappings in which a child's phone is snatched by a criminal who phones its parents to demand a ransom[2].

From about 1997, prepaid mobile phones were introduced. This kicked off a period of rapid growth in the industry as the technology became available to people without credit ratings. For example, prepaids make up 90% of the market in Mexico but 15% in the USA. Worldwide, they're over half, and growing. They also made anonymous communication much more practical, and many criminals have started using them. The issues include not just evading police wiretapping but stalking, extortion, bullying and other kinds of harassment. Prepaids also facilitate simple frauds; if your identity isn't checked when you buy a phone, there's little risk to you if you recharge it with a stolen credit card [343].

It must be said though that most people who use prepaid phones for crime are only getting lightweight anonymity, and remain undetected only because the police don't put serious effort into catching petty crooks. If a really serious crime is committed, traffic analysis will be used, and most criminals don't have any clue of the level of operational discipline needed to stop this. As I already remarked, the alleged 9/11 mastermind Khalid Shaikh Mohammed was caught when he used a prepaid SIM from the same batch as one that had been used by another Al-Qaida member; and after the failed 21/7 London bombings, the would-be bomber Husein Osman fled to Rome, where he was promptly caught. He had changed the SIM in his mobile phone en route; but

[2]There are also completely fake kidnappings in which the bad guys just lie: they say they've snatched the child, and if they call its phone it will be killed. The perpetrators of these crimes are often in prison and have little to lose. 20% of parents still pay up rather than take the risk.

call records show not just the IMSI from the SIM, but also the IMEI from the handset. If you've got all the world's police after you, just changing the SIM isn't anything like enough. Operational security requires a detailed technical understanding of how networks operate, and levels of training and discipline that are most unusual outside national intelligence agencies.

Finally, prepaid mobiles were a gift to crooked call-sell operators. As the billing infrastructure is only invoked when a phone goes on-hook, it's possible for a bad guy to make a call last all day: the numbers called by each of his clients are simply added to a long conference call one after the other. At the end of the day, the phone goes on-hook, a bill for thousands of dollars is generated, the alarm goes off, and the crook tosses the phone in the river. The following day he buys another. That's why network operators typically tear down any call that lasts more than a few hours.

In addition to authentication, the GSM system is supposed to provide two further kinds of protection — location security and call content confidentiality.

The location security mechanism is that once a mobile is registered to a network, it is issued with a *temporary mobile subscriber identification* (TMSI), which acts as its address as it roams through the network. The attack on this mechanism uses a device called an *IMSI-catcher*, which is sold to police forces [488]. The IMSI-catcher, which is typically operated in a police car tailing a suspect, pretends to be a GSM base station. Being closer than the genuine article, its signal is stronger and the mobile tries to register with it. The IMSI catcher claims not to understand the TMSI, so the handset helpfully sends it the cleartext IMSI. This feature is needed if mobiles are to be able to roam from one network to another without the call being dropped, and to recover from failures at the VLR [1283]. The police can now get a warrant to intercept the traffic to that mobile or — if they're in a hurry — just do a middleperson attack in which they pretend to be the network to the mobile and the mobile to the network.

The GSM system is supposed to provide call content confidentiality by encrypting the traffic between the handset and the base station once the authentication and registration are completed. The speech is digitized, compressed and chopped into packets; each packet is encrypted by xor-ing it with a pseudorandom sequence generated from the ciphering key $K_c$ and the packet number. The algorithm commonly used in Europe is A5/1.

A5/1, like Comp128, was originally secret; like Comp128, it was leaked and attacks were quickly found on it. The algorithm is shown in Figure 20.3. There are three linear feedback shift registers of lengths 19, 22 and 23 and their outputs are combined using exclusive-or to form the output keystream. The nonlinearity in this generator comes from a majority-clocking arrangement whereby the middle bits $c_i$ of the three shift registers are compared and the two or three shift registers whose middle bits agree are clocked.

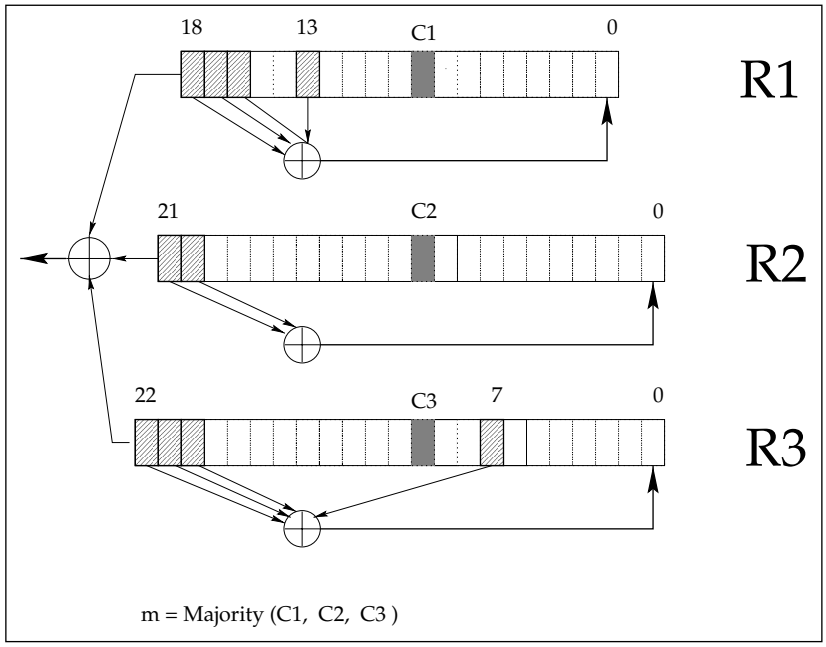


**Figure 20.3:** A5 (courtesy of Alex Biryukov and Adi Shamir)

The obvious attack on this arrangement is to guess the two shorter registers and then work out the value of the third. As there are 41 bits to guess, one might think that about $2^{40}$ computations would be needed on average. It's slightly more complex than this, as the generator loses state; many states have more than one possible precursor, so more guesses are needed. This led to an attack using a lot of FPGAs that is sold for about $1m[3]. Then Biryukov and Shamir found some optimizations and tradeoffs that let A5 be broken with much less effort. Their basic idea is to compute a large file of special points to which the state of the algorithm converges, and then look for a match with the observed traffic. Given this precomputed file, the attack can use several seconds of traffic and several minutes' work on a PC, or several minutes of traffic and several seconds' work [173]. Reverse engineering actual systems also showed that the keying of A5 was deliberately weakened. Although in theory A5 has a 64-bit key (the initial loads for the shift registers) the actual implementations set the ten least significant key bits to zero. Anyway, the response of the GSM vendors to these disclosures was to introduce a third cipher, A5/3, which is based on a strong block cipher known as Kasumi that's also used in third-generation mobile phones.

[3] A system of 15 boxes, each with 20 cards, each with 18 ICs, each with 32 cores, each running at 150MHz, checking one key per Hz, takes 7.5 sec per key and burns 15KW.

However, this attempt to fix the content-confidentiality problem has been undermined by an attack that exploits the underlying authentication protocol and that is now coming into widespread use by police and surveillance folks. It relies on the fact that phone companies in many countries use a weakened version of A5/1, called A5/2. Mobile phones are generally able to use either algorithm, so that they can roam freely. The attack was invented by Elad Barkan, Eli Biham and Nathan Keller [117], and runs as follows. If you're following a suspect who uses his mobile, you record the call, including the initial protocol exchange of challenge and response. Once he's finished, you switch on your IMSI-catcher and cause him to register with your bogus base station. The IMSI-catcher tells his phone to use A5/2 rather than A5/1, and a key is duly set up — with the IMSI-catcher sending the challenge that was used before. So the mobile phone generates the same key $K_c$ as before. As this is now being used in a weak cipher, it can be cracked quickly, giving access to the target conversation recorded previously. In fact it doesn't matter whether that was protected using the medium-security A5/1 or the high-security A5/3. The facts that there's a low-security algorithm, that keys are determined by one of the two principals, and that keys are shared between algorithms, together make the whole system weak.

The A5/1 vulnerability was introduced following pressure from Europe's intelligence agencies, and the ostensible justification was that they didn't want even halfways decent security exported to potentially hostile countries, as it would be harder to spy on them. The conspiracy theorists had a field day with all this — for example, there was a political row in Australia when it turned out that A5/2 was being used there, as it was seen as implying that European governments saw Australia as an intelligence target rather than an ally.

The truth is, as always, more subtle. Weak ciphers can definitely help in some tactical situations. Consider the case I mentioned in the chapter on electronic warfare, where the New Zealand navy sent a frigate to monitor a coup in Fiji. If the Fijian phone company had been allowed to use A5/1 rather than A5/2, this would not have frustrated the mission: the sigint officers could snatch the triplets off the microwave links, hack the location register, and if all else fails, brute force a key. But being able to break traffic quickly is very convenient.

On the other hand, imposing weak cipher security on less developed countries can also have operational costs. In 2007, the Taleban started tapping mobile phone calls made by British soldiers to friends and relatives back home, whom they then called. The wife of one RAF officer was told: 'You'll never see your husband alive — we have just killed him'. It was some hours before she could confirm that he was still safe and well [603]. British troops have now been banned from using mobile phones — which doesn't exactly help morale.

In domestic or friendly-country operations, it's even more complex. The agencies can get lawful access to cleartext from phone companies, but sometimes the warrantry procedures are (or are felt to be) too cumbersome. Hence

the sweetheart deals with some phone companies to get access without warrants; this has led to a number of scandals recently. For example, it emerged that AT&T had been giving the US authorities access to itemised billing data for years. Sweetheart deals were common with companies that used to be monopolies, but are becoming less common as most countries have now deregulated telecomms and have hundreds of phone companies, not just one. Some phone companies are distrusted by the authorities because they're owned by foreigners or even by gangsters. Others are tiny, have never been tapped before, the police just don't know who to talk to. Other companies are respectable and willing to comply with warrants, but unable to do so. One of Britain's most closely guarded law-enforcement and intelligence secrets for several years was that one of the mobile phone networks couldn't be conveniently tapped. Engineers tried to hook up their complex, messy systems to the intelligence community's complex, messy systems but the project went off the rails and they just could not get the thing to work. The police were really scared that crooks might find this out, and migrate en masse to that network.

How are we to make sense of this mess? It's as well to admit that there are many reasons, some honourable and some less so, why government employees may want to get access to traffic by technical means rather than by using the lawful interfaces provided for the purpose. But that doesn't mean that such access should be allowed. The huge improvements in liberty, prosperity and quality of life that the West has enjoyed since the Enlightenment and the Industrial Revolution are due in large part to our replacing the divine right of kings with freedom under the law in a democracy. We elect legislators to make the rules, and if we decide through them that the police can't torture suspects any more, then they'd better not try. Similarly, if our legislators decide that police and intelligence agencies should get warrants to wiretap, then the agencies must obey the law like anyone else. The GSM security story provides a good example of how the deliberate insertion of vulnerabilities can have wide-ranging and unforeseen engineering consequences. It must also be said that the facilities that phone companies and ISPs are now being compelled to provide for properly warranted law-enforcement access are often so poorly engineered that they can be abused [1151]. In 2004–5, persons unknown tapped the mobile phones of the Greek Prime Minister and about a hundred of that country's political, law enforcement and military elite, by subverting the wiretapping facilities built into Vodafone's Greek network. Both Vodafone, and their equipment supplier Ericsson, where heavily fined [1042]. Colleagues and I warned about this problem years ago [4] and I expect it to get worse. I'll discuss it at greater length in Part III.

There are further policy issues with location privacy. There was a storm in Switzerland in 1997 when the press found that the phone company was routinely giving location data to the police [1030], while in the USA, the FCC ordered mobile phone companies to be able to locate people 'so that 911 calls

could be dispatched to the right place'. This was imposed on every user of mobile phone service, rather than letting users decide whether to buy mobile location services or not. Privacy activists were not happy with this.

Anyway, the net effect is that the initial GSM security mechanisms provided slightly better protection than the wireline network in countries allowed to use A5/1, and slightly worse protection elsewhere, until the protocol attacks were discovered and exploited. Now privacy is slightly worse everywhere, as people with the right equipment can get fairly straightforward access to traffic. But it's probably not a big deal. Relatively few people ever get followed around by an investigator with professional snooping equipment — and if you're such a person, then ways to detect and prevent semi-active attacks on your mobile phone are just a small part of the tradecraft you need to know. If you're an average subscriber, the privacy threat comes from possible abuse of data collected by the phone company, such as itemized billing data. From that viewpoint, the vulnerabilities in the communications security mechanisms neither expose you to additional wiretapping, nor prevent the frauds that are likely to cause you the most grief.

### 20.3.3 Third Generation Mobiles – 3gpp

The third generation of digital mobile phones was initially known as the *Universal Mobile Telecommunications System* (UMTS) and now as the *Third Generation Partnership Project* (3gpp, or just 3g). These systems are now available almost everywhere, the exception being in China which is working on its own proprietary variant. Elsewhere, the security is much the same as GSM, but upgraded to deal with a number of GSM's known vulnerabilities. Third generation systems entered service in 2003–2004; the main advantage of 3g over GSM is higher data rates; instead of the 9.6kb/s of GSM and the tens of kilobits per second of GPRS, third-generation data rates are in the hundreds of thousands to millions of bits per second. The vision is that 3g will enable all sorts of mobile services, from mobile TV to laptops that just go online anywhere.

The overall security strategy is described in [1310], and the security architecture is at [1298]. The crypto algorithms A5/1, A5/2 and Comp128 are replaced by various modes of operation of a block cipher called Kasumi [696]. Kasumi is public and is based on a design by Mitsuru Matsui called Misty, which was properly peer-reviewed and has now withstood public scrutiny for a decade [844]. All keys are now 128 bits. Cryptography is used to protect the integrity and confidentiality of both message content and signalling data, rather than just content confidentiality, and the protection at least runs from the handset to a fairly central node, rather than simply to the local base station. This means the picking up the triples, or the plaintext, from the microwave backhaul is no longer an attack. The authentication is now two-way rather

than one-way, ending the vulnerability to rogue base stations; so IMSI-catchers don't work against third generation mobiles. Instead, there is a properly engineered interface for lawful interception [1299]. This can supply key material as well as plaintext, so that if the police follow a suspect, record a call and identify the mobile in the process, they can decrypt that call later, rather than being limited to the plaintext of calls recorded by the phone company after they get their warrant approved.

The protocol mechanics work as follows (see Figure 20.4). In the basic 3gpp protocol, the authentication is pushed back from the base station controller to the visitor location register. The home location register is now known as the *home environment* (HE) and the SIM as the *UMTS SIM* (USIM). The home environment chooses a random challenge RAND as before and enciphers it with the USIM authentication key $K$ to generate a response RES, a confidentiality key $CK$, and integrity key $IK$, and an anonymity key $AK$.

$$\{RAND\}_K = (RES|CK|IK|AK)$$

There is also a sequence number SEQ known to the HE and the USIM. A MAC is computed on RAND and SEQ, and then the sequence number is masked by exclusive-or'ing it with the anonymity key. The challenge, the expected response, the confidentiality key, the integrity key, and the masked sequence number made up into an *authentication vector* $AV$ which is sent from the HE to the VLR. The VLR then sends the USIM the challenge, the masked sequence number and the MAC; the USIM computes the response and the keys, unmasks the sequence number, verifies the MAC, and if it's correct returns the response to the VLR.

| | |
|---|---|
| USIM → HE | IMSI (this can optionally be encrypted) |
| HE → VLR | RAND, XRES, $CK$, $IK$, $SEQ \oplus AK$, $MAC$ |
| VLR → USIM | RAND, $SEQ \oplus AK$, $MAC$ |
| USIM → VLR | RES |

**Figure 20.4:** 3gpp authentication protocol

The UMTS standards set out are many other features, including details of sequence number generation, identity and location privacy mechanisms, backwards compatibility with GSM, mechanisms for public-key encryption of authentication vectors in transit from HEs to VLRs, and negotiation of various optional cryptographic mechanisms.

The net effect is that confidentiality will be improved over GSM: eavesdropping on the air link is prevented by higher-quality mechanisms, and the current attacks on the backbone network, or by bogus base stations, are excluded. Police wiretaps are done at the VLR. In a number of countries, third-generation mobiles were hard for the police to tap in the first few years,

as they had to learn to operate through formal channels and also to integrate their systems with those of the network operators. In the second phase, it's proposed to have end-to-end encryption, so that the call content and some of the associated signaling will be protected from one handset to another. This led to government demands for a *key escrow protocol* — a protocol to make keys available to police and intelligence services on demand. The catch is that if a mobile phone call takes place from a British phone company's subscriber using a U.S. handset, roaming in France, to a German company's subscriber roaming in Switzerland using a Finnish handset, and the call goes via a long distance service based in Canada and using Swedish exchange equipment, then which of these countries' intelligence agencies will have access to the keys? [1299] (Traditionally, most of them would have had access to the call content one way or another.)

One solution pushed by the agencies in Britain and France is the so-called Royal Holloway protocol [663], designed largely by Vodafone, which gives access to the countries where the subscribers are based (so in this case, Britain and Germany). This is achieved by using a variant of Diffie-Hellman key exchange in which the users' private keys are obtained by encrypting their names under a super-secret master key known to the local phone company and/or intelligence agency. Although this protocol has been adopted in the British civil service and the French health service, it is at odds with the phone company security philosophy that master keys are a bad thing. The protocol is also clunky and inefficient [76].

So 3gpp won't provide a revolution in confidentiality, merely a modest improvement. As with GSM, its design goal is that security should be comparable with that of the wired network [621] and this looks like being achieved.

### 20.3.4 Platform Security

The final point I need to make here is that as mobile phones become more widespread and more programmable, they may suffer from the malware problems that have plagued the PC. They have followed the pattern predicted by security economics. At first, the platform vendors — the firms selling operating systems, such as Symbian, Microsoft and Linux — didn't incorporate much security, as it would have got in application developers' way and appealing to complementers is vital when building share in a new market with network externalities. Then, as one platform pulled ahead of the others, the malware writers targeted it. In 2007, viruses and worms are being detected at the rate of about 300 per annum for Symbian phones, and one or two a year for the others. Symbian has started a program of hardening their platform, with progressively more sophisticated access controls, code signing, and so on.

Mobile phone platforms have also acquired DRM mechanisms. The Open Mobile Alliance DRM version 1 supports the download of ringtones, while version 2 supports music and video download with more general mechanisms. Version 1 is widely used, but version 2 has been held up by an interesting dispute. The main OMA promoter, Nokia, would like to become the distribution channel of choice for mobile music, just as Apple has become for PCs with iTunes. However the network operators are extremely reluctant to let Nokia have a business relationship with their customers directly, and this has held up deployment.

In general, security is made more difficult by the long and complex supply chain that delivers mobile phone service to customers. IP companies like ARM own the chip designs; foundries such as Infineon make the chips; handset designers like Samsung manufacture the actual mobiles; Symbian provides the operating system; Nokia may provide some more software for a download interface; a network operator such as Vodafone provides the national infrastructure; and a local operating company then bills the customer. There has been a tendency for everyone in this chain to see security as a problem to be tossed over the fence to the companies on either side. On top of this there might be apps provided by games vendors, and corporate apps such as fleet management that are provided by third-party software houses. Add the next generation of location-based services and offerings from the likes of Google and eBay, and the whole thing becomes complex beyond belief.

One final aspect of mobile phone platforms is locking. In many countries, handsets are subsidised out of future call revenue: you get a cheap phone in return for a year's contract with a network. The downside is that your phone is locked to the network. Even some prepaid phones are locked, and carry an implicit subsidy from expected future token sales. However, in some countries, this business model isn't permitted. There has thus arisen a brisk trade in unlocking tools and services, some of which are devised by finding exploits in the phone software, and others using the kind of hardware reverse-engineering techniques I described in Chapter 16. Legal skirmishing between the phone companies and the unlocking services came to a head after Apple launched the iPhone, which had a twist on this business model: the networks for which you could buy it were those that had paid Apple the most for the privilege. This annoyed iPhone purchasers as the U.S. network of choice, AT&T, was fairly slow. The iPhone was duly hacked, and AT&T sent its lawyers after the unlockers. It also shipped a software upgrade that disabled unlocked phones.

Laws on locking and unlocking vary widely. In the USA, the Digital Millennium Copyright Act (DMCA), which prohibits interference with any technical mechanism that enforces copyright, has a specific exemption for cellphone unlocking, in order that copyright law shouldn't be abused to stifle competition in the mobile phone market. It's now being argued that

this exemption covers only people who unlock their own phones, but doesn't extend to the sale of unlocking tools or services [821]. So in America, unlocking is legal if you're a customer (though the provider may brick your phone), and may be open to challenge if you do it commercially. At the other end of the scale, courts in France and Germany have held mobile phone locking to be illegal: Apple will have to offer unlocked models for sale there if it offers the product at all. Such variations in law and business practice have led to the development of a thriving grey market whereby phones are shipped from one country to another; this in turn has driven secondary markets for unlocking tools and even for assorted frauds.

### 20.3.5 So Was Mobile Security a Success or a Failure?

Whether mobile-phone security has been a success or a failure depends on whom you ask.

From the point of view of cryptography, it was a failure. Both the Comp128 hash function and the A5 encryption algorithm were broken once they became public. In fact, GSM is often cited as an object lesson in Kerckhoffs' Principle — that cryptographic security should reside in the choice of the key, rather than in the obscurity of the mechanism. The mechanism will leak sooner or later and it's better to subject it to public review before, rather than after, a hundred million units have been manufactured. (GSM security wasn't a disaster for most cryptographers, of course, as it provided plenty opportunities to write research papers.)

From the phone companies' point of view, GSM was a success. The shareholders of GSM operators such as Vodafone have made vast amounts of money, and a (small) part of this is due to the challenge-response mechanism in GSM stopping cloning. The crypto weaknesses were irrelevant as they were never exploited (at least not in ways that did significant harm to call revenue). There are one or two frauds that persist, such as the long conference call trick; but on balance the GSM design has been good to the phone companies.

From the criminals' point of view, GSM was also fine. It did not stop them stealing phone service: the modus operandi merely changed, with the cost falling on credit card companies or on individual victims of 'identity theft' or street robbery. It did not stop calls from anonymous phones; the rise of the prepaid phone industry made them even easier. (The phone companies were happy with both of these changes.) And of course GSM did nothing about dial-through fraud.

From the point of view of the large-country intelligence agencies, GSM was fine. They have access to local and international traffic in the clear anyway, and the weakened version of A5 facilitates tactical signint against developing countries. And the second wave of GSM equipment is bringing some juicy features, such as remote control of handsets by the operator [1061]. If you can

subvert (or masquerade as) the operator, then there seems to be nothing to stop you quietly turning on a target's mobile phone without his knowledge and listening to the conversation in the room.

From the point of view of the police and low-resource intelligence agencies, things are not quite so bright. The problem isn't the added technical complexity of GSM networks: court-ordered wiretaps can be left to the phone company (although finding the number to tap can be a hassle if the suspect is mobile). The problem is the introduction of prepaid mobile phones. This not only decreases the signal to noise ratio of traffic analysis algorithms and makes it harder to target wiretaps, but also encourages crimes such as extortion and stalking.

From the customer's point of view, GSM was originally sold as being completely secure. Was this accurate? The encryption of the air link certainly did stop casual eavesdropping, which was an occasional nuisance with analog phones. (There had been some high-profile cases of celebrities being embarrassed, including one case in Britain where Prince Charles was overheard talking to his mistress Camilla Parker-Bowles before his divorce from Princess Diana, and one in the USA involving Newt Gingrich.) But almost all the phone tapping in the world is done by large intelligence agencies, to whom the encryption doesn't make much difference.

Things are even less positive for the subscriber when we look at billing. Cryptographic authentication of handsets can't stop the many frauds perpetrated by premium rate operators and phone companies. If anything it makes it harder to wriggle out of bogus charges, as the phone company can say in court that your smartcard and your PIN must have been used in the handset that made the call. The same will apply to 3rd generation phones. The one minor compensation is that GSM facilitated the spread of prepaid phones, which can limit the exposure.

So the security features designed into GSM don't help the subscriber much. They were designed to provide 'security' from the phone company's point of view: they dump much of the toll fraud risk, while not interrupting the flow of premium rate business — whether genuine or fraudulent.

In the medium term, the one ray of comfort for the poor subscriber is that the increasing complexity of both handsets and services may create regulatory pressure for transparent mechanisms that enable the customer to control the amount she's billed. There are a number of factors pushing for this, such as the growing vulnerability of platforms to malware; and a number of factors pushing in the other direction, such as the phone companies' desire to keep pricing opaque so that they can rip customers off. I mean this not just in the strict sense that phone companies often defraud their customers, but also in the colloquial sense that confusion pricing is a mainstay of phone company economics. I'll go into this in more detail in the next section, where I'll look at the economics of telecomms and how they relate to fraud and abuse.

### 20.3.6 VOIP

The latest development in telephone is voice over IP (VOIP), in which voice traffic is digitised, compressed and routed over the Internet. This had experimental beginnings in the 1970s; products started appearing in the 1990s; but decent call quality requires bandwidth of about 80 kbit/sec, more than can be got from a dial-up modem. As a result, VOIP only really took off once a critical mass of households had broadband connections. Since about 2005, it has become big business, with eBay purchasing Skype in 2006 for $2.6bn. In fact, most normal phone calls are digitized and sent over IP networks belonging to the phone companies, so in a technical sense almost all phone calls are 'VOIP', but in line with common usage I'll use the term only for those calls made by customers over an IP service that they access directly, via their ISP.

The VOIP market is still fragmented, and suffers many of the problems you'd expect. Most products were shipped quickly in a race to market, with security an afterthought at best. There was the usual crop of stack overflows and other vulnerabilities at the implementation level. The most popular VOIP protocol, the Session Initiation Protocol (SIP), turns out to have vulnerabilities that enable third parties to wrongly bill calls to subscribers [1375]. These could give rise to difficult disputes between VOIP services and their customers.

There are many issues with VOIP as the market shakes down. The leading system, Skype, is a closed system using proprietary protocols while its many competitors are fragmented. One business opportunity is 'click-to-call' whereby an advertisement could contain a VOIP link enabling a prospective customer to click to speak to a representative; early efforts have run up against compatibility problems. Technical problems range from differing protocols through how to deal with the jitter (time delay variation) caused by the variable quality of service on the Internet, and how to deal with network address translation are firewalls.

The interaction with security is complex. Corporate security policies can result in firewalls refusing to pass VOIP traffic. Phone calls can be more secure if made over VOIP, as encryption is easy to add and with some services (such as Skype) it comes turned on by default. For some time at least, it may be more difficult for police and intelligence services to get access to call contents and to records of who called whom; many services have not yet got round to engineering a law-enforcement interface, and in the case of Skype, its peer-to-peer nature might make that difficult. The FBI is currently pushing for the CALEA regulations to be applied rigorously to VOIP providers, while the industry and civil liberties groups are resisting. (I'll have more to say about this in Part III.) Wiretaps are not the only issue; click-to-call clearly will have other security, privacy and safety issues in the context of social networking sites, especially those used by minors.

A more high-profile regulatory issue is that governments want emergency calls made through VOIP services to work reliably, and provide information about the location of the caller. This is hard; an IP packet stream can be coming from anywhere, and no-one owns enough of the Internet to guarantee quality of service. At a deeper level than that, the dispute over dependability is just the latest tussle in the forty years' war between computer companies and phone companies, which I've alluded to from time to time. Computer companies are innovative and entrepreneurial, racing to market with products that are just about good enough; phone companies are slow-moving and heavily regulated, with a fifteen-year product cycle and services engineered for 99.999% availability. Ever since people started using modems on a large scale in the 1960s, there has been one battle after another — which the computer companies pretty well always win. A glimpse into the future may have been provided by a two-day Skype outage in August 2007, caused by poorly-engineered protocols. After a large number of computers rebooted following Microsoft's 'patch Tuesday' security update, tens of millions of Skype clients simultaneously tried to contact the network. The congestion caused the network to fail, yet the clients kept on trying. We can probably expect a lot more failures. Although a VOIP handset looks like a phone and works like a phone, it's more like email in terms of reliability. If the power goes off, so does your service.

The main problems beyond that have to do with the incompatibility of the phone companies' business model and the nature of the Internet. Phone companies make their money by charging you vastly different rates for different types of content: their typical rates work out at a hundredth of a cent per megabyte for cable TV, eight cents per megabyte for wireline phone, three dollars a megabyte for mobile phone, and a whopping three thousand dollars a megabyte for text messages. The opportunity exploited by VOIP is to arbitrage these, and it's hardly surprising that the phone companies do what they can to get in the way. ISPs who are also phone companies deliberately cause problems for VOIP to stop it competing with their phone service; this is particularly pronounced with mobile IP service. For these reasons, we'd better look at the economics of phone companies and the interaction with security.

## 20.4 Security Economics of Telecomms

Phone companies are classic examples of a business with extremely high fixed costs and very low marginal costs. Building a nationwide network costs billions and yet the cost of handling an additional phone call is essentially zero. As I discussed in Chapter 7 on Economics, this has a couple of implications.

First, there's a tendency towards dominant-firm markets in which the winner takes all. Indeed for many years telephone service was considered in

most countries to be a 'natural monopoly' and operated by the government; the main exception was the USA where the old AT&T system was heavily regulated. After the breakup of AT&T following an antitrust case, and Margaret Thatcher's privatisation of BT, the world moved to a different model, of regulated competition. The details vary from one country to another but, in general, some sectors (such as mobile phones) had a fixed number of allowed competitors; others (such as long-distance provision) were free for companies to compete in; and others (such as local loop provision) remained de facto monopolies but were regulated.

Second, because the marginal cost of service provision is zero, the competitive sectors (such as long-distance calling) saw prices drop quickly to a very low level — in many cases below the level needed to recoup the investment. (The large investments made during the dotcom bubble end up being treated as sunk costs.)

In such a market, firms will try to maintain price discrimination by whatever means they can. In many telecomms markets, the outcome is *confusion pricing* — products are continually churned, with new offerings giving generous introductory discounts to compete with the low-cost providers, but with rates sneakily raised afterwards. The effect is to discriminate between 'people who will spend money to save time, and people who will spend time to save money'. If you can be bothered to continually check prices, you can get really good deals, but often at the cost of indifferent service. If you don't have the time to keep scrutinising your phone bills, and the latest emails you get from your ISP advising you of the latest service changes, you can find that the call home you made from your mobile while on business in France just ate up all that month's savings. In the end, you can end up paying about the same amount per month that you did in the past. Andrew Odlyzko, a scholar of phone-company economics, suggests the eventual way forward will be fixed-price contracts: for a certain amount per month you'll get a home phone, ISP service, a couple of mobiles and a decent allowance of air minutes and texts [982]. In the meantime, telecomms pricing remains murky, contentious and far from transparent. This leads directly to abuse.

### 20.4.1 Frauds by Phone Companies

One of the steadily growing scams is the unscrupulous phone company that bills lots of small sums to unwitting users. It collects phone numbers in various ways. (For example, if you call an 800 number, then your own number will be passed to the far end regardless of whether you tried to block caller line ID.) The wicked phone company then bills you a few dollars. Your own phone company passes on this charge and you find there's no effective way to dispute it. Sometimes the scam uses a legal loophole: if you call an 800 number in the USA, the company may say 'Can we call you right back?' and if you agree

then you're deemed to have accepted the charges, which are likely to be at a high premium rate. The same can happen if you respond to voice prompts as the call progresses. These practices are known as *cramming*.

I was myself the victim on an attempt at cramming. On holiday in Barcelona, my wife's bag was snatched, so we called up and cancelled the phone that she'd had in it. Several months later, we got a demand from our mobile provider to pay a few tens of dollars roaming charges recently incurred by that SIM card in Spain. In all probability, the Spanish phone company was simply putting through a few charges to a number that they'd seen previously, in the knowledge that they'd usually get away with it. My mobile service provider initially insisted that even though I'd cancelled the number, I was still liable for calls billed to it months afterwards and had to pay up. I got out of the charges only because I'd met the company's CEO at an academic seminar and was able to get his private office to fix the problem. Customers without such access usually get the short end of the stick. Indeed, UK phone companies' response to complaints has been to offer its customers 'insurance' against fraudulent charges. That they can get away with this is a clear regulatory (and indeed policing) failure.

Another problem is *slamming* — the unauthorized change of a subscriber's long distance telephone service provider without their consent. The slammers tell your local phone company that you have opted for their service; your phone company routes your long distance calls through them; they hope you don't notice the change and dispute the bill; and the telephone charges can then be jacked up. Some local phone companies, such as Bell Atlantic, allow their customers to freeze their chosen long distance carrier [22].

It would be a mistake to assume that cramming and slamming are just done by small fly-by-night operators. AT&T is one of the worst offenders, having been fined $300,000 not only for slamming, but for actually using forged signatures of subscribers to make it look as if they had agreed to switch to their service. They got caught when they forged a signature of the deceased spouse of a subscriber in Texas [390]. As for the UK, slamming wasn't even made illegal until 2005.

Another problem is the fly-by-night phone company. As anyone in the USA is legally entitled to set up a phone company, it is straightforward to set one up, collect some cash from subscribers, and then vanish once the invoices for interconnect fees come in. Companies also advertise sex lines with normal phone numbers to trap the unwary, then send huge bills to the subscriber at his residential addresses and try to intimidate him into paying. The regulation of premium-rate providers and their business practices is a widespread problem.

And it's not just the small operators that indulge in sharp practice. An example that affects even some large phone companies is the short termination of international calls.

Although premium-rate numbers are used for a number of more or less legitimate purposes such as software support, many of them exploit minors or people with compulsive behaviour disorders. So regulators have forced phone companies in many countries to offer premium-rate number blocking to subscribers. Phone companies get round this by disguising premium rate numbers as international ones. I mentioned scams with Caribbean numbers in section 20.2.1 above. Now many other phone companies from small countries with lax regulators have got into the act, and offer sex line operators a range of numbers on which they share the revenue.

Often a call made to a small-country phone company doesn't go anywhere near its ostensible destination. One of the hacks used to do this is called *short termination*, and here's an example that surfaced a few years ago. Normally calls for the small Pacific country of Tuvalu went via Telstra in Perth, Australia, where they were forwarded by satellite. However, the sex line numbers were marked as invalid in Telstra's system, so they were automatically sent via the second-choice operator — a company in New Zealand. (The girls — or to be more precise, the elderly retired hookers who pretend to be young girls — were actually in Manchester, England.) Technically, this is an interesting case of a fallback mechanism being used as an attack vehicle. Legally, it is hard to challenge as there is an international agreement (the Nairobi Convention) that stops phone companies selectively blocking international destinations. So if you want to stop your kids phoning the sex line in Tuvalu, you have to block all international calls, which makes it harder for you to phone that important client in Germany.

Problems like these are ultimately regulatory failures, and they are increasingly common. (For example, in the Moldova scam I mentioned earlier, the calls didn't go to Moldova but to Canada [251].) They are continuing to get worse as technology makes new, complex services possible and the regulators fail to keep up.

## 20.4.2 Billing Mechanisms

Billing mechanisms are a growing source of problems, as the economic forces discussed above lead to ever-more-complex rate cards. Even the phone companies themselves sometimes fall foul of the growing complexity. Sometimes their rates get so complex that people can arbitrage against them; there has been more than one case in which an entrepreneur found he could set up an international premium-rate service and be paid more per minute for calling it, than it cost him to call it using the best discount rate. Phone companies have tried to recover such trading profits through the courts, claiming fraud — with mixed success.

The security of the billing mechanisms covers a much wider range of issues. Present arrangements are inadequate for a number of reasons.

- A *call detail record* (CDR) is only generated once the calling phone goes on-hook. This was a long-established feature of wireline networks, but once the environment changed to mobile it became a serious problem. As I mentioned above, someone running a call-sell operation can set up a long conference call using a stolen or prepaid mobile, which clients join and leave one after the other. The phone stays off-hook continuously for hours. As soon as it goes on-hook, a CDR for several thousand dollars is generated, and the alarm goes off. The operator throws the phone in the river and starts using the next one. By 1996, this had become so serious that Vodafone introduced a six hour limit on all mobile calls. But it won't be acceptable to just drop all 3gpp calls after six hours. Many users are expected to have always-on internet connections (such as from their laptops) with relatively light packet traffic most of the time.
- More seriously, the back-end accounting system was designed in the days when phone companies were sleepy government departments or national monopolies, and there were no premium-rate services through which real money could be extracted from the system. So it has little in the way of audit and non-repudiation. In effect, phone company A tells phone company B, 'please debit your customer no. X the sum of $Y and send me the money' — and it does. Even when these debits are mistaken or fraudulent, phone company B has no incentive to quibble, as it gets a cut. The result, as we saw with the cramming cases above, is that fraud slowly rises as insiders abuse the system. This is no longer fit for purpose in a world with many phone companies, quite a few of which are unscrupulous. The regulators aren't effective, and the only real backward pressure comes from the growing number of prepay customers.
- The phone companies also want to be able to charge for relatively high-value product and service delivery, extending the current premium services through location-based services ('give me a map showing me how to drive to the nearest McDonalds') to music and video downloads and extra services such as the Finnish ferry tickets, cans of coke from vending machines, and (most recently) parking meters in London. The accounting system will have to become a lot more robust, and dispute resolution mechanisms fairer and more transparent, if this potential is to be realised.
- All this interacts with platform security. If malware becomes widespread on mobile phones, then the botnet herders who control subverted phones will be able to pay for all sorts of goods and services by getting infected machines to send text messages. Recent history suggests that any exploits that can be industrialised to make money on a large scale, will be.

So how can phone payment systems be improved?

One proposed way of implementing this is to incorporate a micropayment mechanism [92]. The idea is that the phone will send regular *tick payments* to each of the networks or service providers which are due to be paid for the call. The tick payments can be thought of as electronic coins and are cryptographically protected against forgery. At the start of the call, the handset will compute a number of phone ticks by repeated hashing: $t_1 = h(t_0)$, $t_2 = h(t_1)$, and so on, with $t_k$ (for some credit limit $k$, typically $2^{10}$ units) being signed by the phone company. The phone will then release ticks regularly in order to pay for the services as the call progresses. It starts by releasing $t_k$, then $t_{k-1}$, then $t_{k-2}$ and so on. If a charge is subsequently disputed — whether by a subscriber or a network operator — the party claiming an amount of (say) $j$ ticks must exhibit the ticks $t_{k-j}$ and $t_k$, the latter with a certificate. As the hash function $h$ is one-way, this should be hard to do unless the handset actually released that many ticks. The tick $t_{k-j}$ can how be checked by applying the hash function to it $j$ times and verifying that the result is $t_k$[4]. One advantage of tick payments is that as well as protecting the phone companies from conference call frauds, it could protect the subscriber from many more. It could enable phone application designers to empower users in new ways: for example, you might by default decide to accept calls, or play games, or do other functions, only so long as they cost less than a dollar, and require user intervention for anything more expensive. At present, that kind of functionality just isn't available, except via the rather clunky mechanism of only loading so much airtime into your phone.

The industry's proposed solution is to redesign the call data record to contain a lot more information. In addition to time, duration and called number, it will have fields for data quantity, location and quality-of-service. This is not just to support possible future differential charging for quality of service, but also to help with emergency call tracking, and to comply with a 2002 European directive on telecomms requires all mobile operators retain location information on mobile phones for at least a year, for law enforcement purposes. There was a proposal for an online cost-control mechanism to limit the charges incurred for each user [901], but this appears to have stalled. The cost-control mechanisms are not being standardized but can involve forwarding charging data from either the local network or the gateway to the home environment which will be able to have the call terminated if the available credit is exhausted (as with a prepaid SIM card) or if the use appears to be fraudulent. It's tempting to draw a parallel with NATO's failure to make IFF work properly across different countries' armed forces, which I discussed in Chapter 19. The public good (in this case, transparent and dependable cost

[4]This protocol is an example of multiple simultaneous discovery, having been invented by our group at Cambridge, by Pedersen, and by Rivest and Shamir, independently in 1995 [40, 1013, 1077].

control) isn't high on the agendas of the stakeholders who'd have to work together to make it happen. Indeed, it's even worse. It would be hard to find a general anywhere in North America or Europe who didn't agree that decent IFF would be a Good Thing; but many of the major stakeholders depend for their existence on confusion pricing and would see a decent charging system as a serious threat.

The extreme difficulty of engineering a global solution has left the market open to a variety of local ones. In the USA, there is a move to integrate RFID-based credit cards with NFC-compatible mobile phones. In the UK, a scheme called PayForIt was launched in 2007 by the main mobile operators that aims to replace premium SMS services with a WAP-based protocol that 'provides a uniform payment experience' and requires 'clear pricing, access to terms and conditions and merchant contact details before a purchase is confirmed'. O2, one of the big networks, required all its users to switch from June 2007. They said there would be 'reduced customer care issues'.

Personally, I am sceptical, because of the lack of bankable guarantees for the customer; the terms and conditions state that 'any queries or complaints regarding Goods and Services must be referred to the Supplier,' while firms advertising PayForIt transaction acquisition quote fixed prices and seem ready to accept all comers, respectable or otherwise. Perhaps the scheme will simply hold up the growth of phone-based payments by making them more fiddly and restricting them to more capable handsets, while not resolving the underlying trust problem. A law that enforced customer rights would be better for all. In its absence, the phone companies appear to be setting up a banking system but without the regulations and controls that long and bitter experience has shown necessary for bank stability and trustworthiness. In civilised countries, mafiosi are not allowed to run banks. But gangsters have a basic human right (in America at least) to own a phone company. So phone companies should not run banks.

## 20.5 Summary

Phone fraud is a fascinating case study. People have been cheating phone companies for decades, and recently the phone companies have been vigorously returning the compliment. To start off with, systems were not really protected at all, and it was easy to evade charges and redirect calls. The mechanism adopted to prevent this — out-of-band signalling — proved inadequate as the rapidly growing complexity of the system opened up many more vulnerabilities. These range from social engineering attacks on users through poor design and management of terminal equipment such as PBXes to the exploitation of various hard-to-predict feature interactions.

On the mobile front, the attempts to secure GSM and its third generation successor make an interesting case study. Their engineers concentrated on communications security threats rather than computer security threats, and they concentrated on the phone companies' interests at the expense of the customers'. Their efforts were not entirely in vain but did not give a definitive solution.

Overall, the security problems in telecomms are the result of environmental changes, including deregulation, which brought in many new phone companies. But the main change was the introduction of premium rate numbers. While previously phone companies sold a service with a negligible marginal cost of provision, suddenly real money was involved; and while previously about the only serious benefit to be had from manipulating the system was free calls, or calls that were hard for the police to tap, suddenly serious money could be earned. The existing protection mechanisms were unable to cope with this evolution. However, the major phone companies are so threatened by price competition that their business models are now predicated on confusion pricing. So the incentives for an overhaul of the billing mechanisms are just not there.

Ultimately, I suspect, the regulator will have to step in. The best solution in the USA could be the extension of Regulation E, which governs electronic banking, to phone companies — as they have become de facto banks. When you can use your mobile phone to buy ferry tickets and songs, and feed parking meters, it's performing all the functions of an electronic purse, which if issued by a traditional payment service provider such as VISA or Mastercard would fall squarely under Reg E. Also, I believe that either the FCC or the Federal Reserve should have the right to ban known criminals from owning or managing regulated phone companies. Europe is even further behind, but there is some action in the regulatory pipeline: a draft Directive that will impose a Europe-wide duty on companies to disclose security breaches in telecomms to affected customers, similar to the breach notification laws in force in over 30 U.S states in 2007. This at least will be a start.

## Research Problems

Relatively little research is done outside phone company and intelligence agency labs on issues related specifically to phone fraud and wiretapping. There is growing interest in traffic analysis, which I'll discuss later; and in the likely effects of next-generation value added services, which are bound to introduce new feature interactions and other vulnerabilities. The interaction between all communications (especially mobile), platform security, and the mechanisms used to protect distributed systems security, also looks like fertile ground for both interesting research and expensive engineering errors. Society

expects greater resilience and availability from the phone system than from the Internet — for example, to get through to emergency services — and as the two systems converge there will be some interesting assurance problems.

## Further Reading

There are a lot of scattered articles about phone fraud, but nothing I know of that brings everything together. The underlying technologies are described in a number of reference books, such as [1061] on GSM, and more can be found on websites such as [1190]. An overview of UMTS can be found in [642], and a number of relevant research papers at [92]. To keep up with phone fraud, a useful resource is the Discount Long Distance Digest [390]. NIST has a guide on how to evaluate your PBX for vulnerabilities [943]. Finally, there's a survey of threats to mobile payment systems by the Mobile Payment Forum that gives a summary of the state of play as of 2002 [897].