Security Engineering

Lesson 13

Multilateral Security, Secure Systems and Applications

Dr. Marenglen Biba

Outline

- Introduction
- Compartmentation, the Chinese Wall, and the BMA Model

Introduction

- Often, our goal is not to prevent information flowing "down" a hierarchy but to prevent it flowing "across," between departments.
- Relevant applications range from healthcare to national intelligence, and include most applications where the privacy of individual customers', citizens' or patients' data is at stake.
- They account for a significant proportion of information processing systems, but their protection is often poorly designed and implemented.

Multilevel and Multilateral

Figure 8.1 Multilevel security.

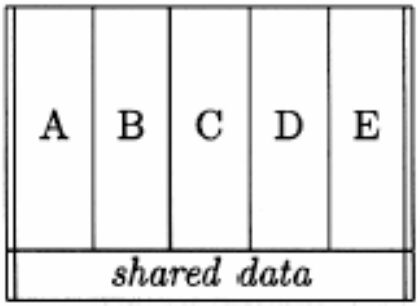

Figure 8.2 Multilateral security.

Terminology

- Information flow controls of the type we are interested in are known by a number of different names.
- In the U.S. intelligence community, for example, they are known as *compartmented security* or *compartmentation*.
- We will use the European term *multilateral security*.

Outline

- Introduction
- Compartmentation, the Chinese Wall, and the BMA Model

Multilateral Models

- There are (at least) three different models of how to implement access controls and information flow controls in a multilateral security model.
- *Compartmentation*, used by the intelligence community.
- *Chinese Wall* model, which describes the mechanisms used to prevent conflicts of interest in professional practice.
- *BMA model*, developed by the British Medical Association to describe the information flows permitted by medical ethics.
- Each of these has potential applications outside its field of origin.

Lattice Model

- Codewords are a way of expressing access control groups, and can be dealt with using a variant of Bell-LaPadula, called the *lattice model*.
- As an illustration, suppose we have a codeword, say, ‘Crypto’.
- Someone cleared to ‘Top Secret’ would be entitled to read files classified ‘Top Secret’ and ‘Secret’, but would have no access to files classified ‘Secret Crypto’ unless he or she also had a crypto clearance.

Lattice Model

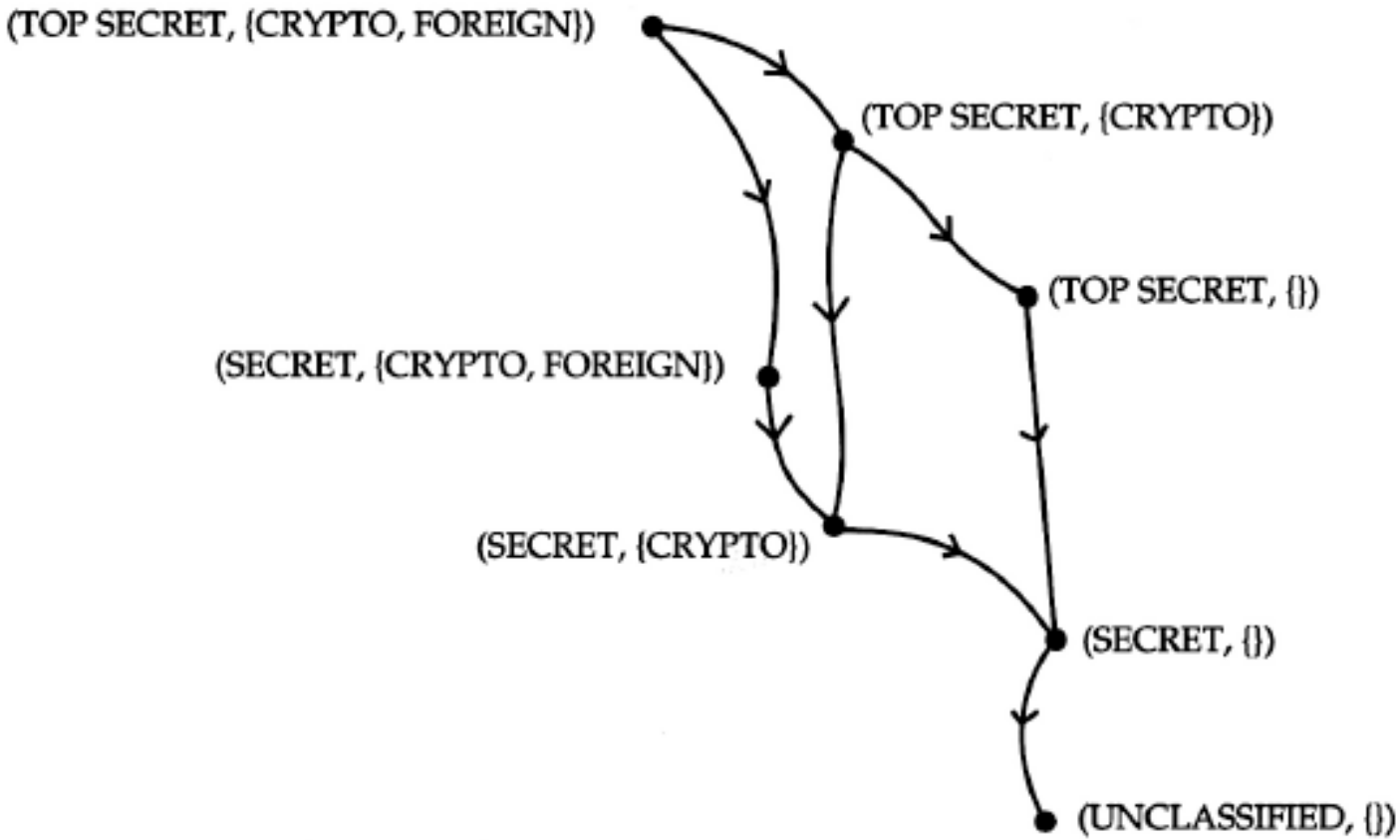

Figure 8.3 A lattice of security labels.

Implementing the Lattice

- In order for information systems to support this, we need to distill the essence of:
  - classifications,
  - clearances, and
  - labels

  into a security policy that we can then use to drive security targets, implementation, and evaluation.
- As it happens, the Bell- LaPadula model appears to go across more or less unchanged.
- We still have information flows between High and Low as before, where High is a compartment that dominates Low.
- If two nodes in a lattice are incompatible — as with ‘Top Secret’ and ‘Secret Crypto’ in Figure — then there should be no information flow between them at all.

The Chinese Wall

- The second model of multilateral security is the Chinese Wall model, developed by Brewer and Nash.
- Its name comes from the fact that financial services firms such as investment banks have internal rules designed to prevent conflicts of interest, which they call Chinese Walls.

The Chinese Wall

- The model's scope is wider than just investment banking.
- Many professional and services firms have clients who may be in competition with each other:
  - software vendors
  - advertising agencies
  - accountants

  are other examples.
- A typical rule is that "a partner who has worked recently for one company in a business sector may not have access to the papers of any other company in that sector."
- So an advertising copywriter who has worked on, say, the Shell account, will not be allowed to work on any other oil company's account for some fixed period of time.

Separation of duty

- The Chinese Wall model thus features a mix of free choice and mandatory access control:

  a partner can choose which oil company to work for, but once that decision is taken their actions in that sector are completely constrained.

- It also introduces the concept of *separation of duty* into access control
  - A given user may perform transaction A or transaction B, but not both.

Chinese Walls in Practice

- In practice, however, Chinese Walls still are implemented using manual methods.
- One large software consultancy has each of its staff maintain an "unclassified" curriculum vitae containing entries that have been sanitized and agreed with the customer.
- A consultant's manager should be aware of possible conflicts, and not forward the CV to the client if in doubt;
  - If this fails, the client can spot potential conflicts from the CV.
  - And if this also fails, then the consultant is duty bound to report any potential conflicts as soon as they appear.

The BMA Model

- Perhaps the most important, interesting, and instructive example of multilateral security is found in medical information systems.
- People are arguing about whether privacy norms will have to be radically revised as genetic data become widely available.
- In Iceland, for example, a project to build a national medical database that will incorporate not just medical records but also genetic and genealogical data, so that inherited diseases can be tracked across generations, has caused an uproar.

Threat Model to Medical Privacy

- Currently, the main threat to medical privacy is social engineering. The typical attack on medical record privacy comes from a private detective who phones a doctor's office or health insurer with a plausible tale:

  Hello, this is Dr. Burnett of the cardiology department at the Conquest Hospital in Hastings. Your patient Sam Simmonds has just been admitted here in a coma, and he has a funny-looking ventricular arrythmia. Can you tell me if there's anything relevant in his record?

- In June 2000, millionaire British government minister Lord Levy was acutely embarrassed after someone called the tax office pretending to be him and found out that he'd only paid £5000 in tax the previous year. ☺

Defending against social engineering

- In 1996, an experiment was done in England whereby the staff at a health authority were trained to screen out such false pretext telephone calls.
- The most important element of the advice they were given was that they were to always call back — and not to a number given by the caller, but to the number in the phone book for the hospital or other institution where the caller claimed to work.
- It turned out that some 30 telephone enquiries a week were counterfeit.

Is training enough?

- Training staff in this way is more important than most technical protection measures.
- But the best staff training in the world won't protect a system in which too many people see too much data.
- There will always be staff who are careless or even crooked; and the more records they can see, the more harm they can do.
  - In one high-profile case, a convicted child rapist working as an orthopedic technician at Newton-Wellesley Hospital in Newton, Massachusetts, was caught using a former employee's password to go through the records of 954 patients to get the phone numbers of girls to whom he then made obscene phone calls. He ended up doing jail time. There are many more incidents of a less dramatic nature.

Ad hoc measures

- There are many ad hoc measures that hospitals can take to improve the protection of existing systems.
- One of the most effective is to keep the records of former patients in a separate archive, and give only a small number of admissions staff the power to move records from there to the main system.
- Another is to introduce a *honey trap*, a number of bogus records with celebrity names. Reportedly, one Boston hospital uses "medical records" with the names of Kennedy family members for this purpose:
  - staff who browse them can be identified and disciplined.
- However, a patchwork of ad hoc measures isn't a good way to secure a system. We need a proper access control policy, thought through from first principles and driven by a realistic model of the threats. Which policy is appropriate for healthcare?

Security Policy

- This question faced the British Medical Association (BMA) in 1995. The U.K. government had introduced an IT strategy for the National Health Service whose security policy was multilevel.
- The idea was that AIDS databases would be at a level corresponding to 'Secret'; normal patient records at 'Confidential'; and administrative data, such as drug prescriptions and bills for treatment, at 'Restricted'.
- It was soon realized that this wasn't going to work.
- For example, how should a prescription for AZT be classified?
  - It's a drug prescription, so it should be 'Restricted'; it identifies a person as HIV positive, so it must be 'Secret'. So all the 'Secret' AZT prescriptions must be removed from the 'Restricted' file of drug prescriptions.
- The same goes for most of the other prescriptions, as they identify treatments for named individuals and so should be 'Confidential'. But then what use will the file of prescriptions be to anybody?

A Policy Example

The policy consists of nine principles:

1. Access control: each identifiable clinical record shall be marked with an access control list naming the people or groups of people who may read it and append data to it.
   - The system shall prevent anyone not on the access control list from accessing the record in any way
2. Record opening: a clinician may open a record with herself and the patient on the access control list.
   - Where a patient has been referred, she may open a record with herself, the patient and the referring clinician(s) on the access control list

A Policy Example

3. Control: One of the clinicians on the access control list must be marked as being responsible.
   - Only she may alter the access control list, and she may only add other health care professionals to it.

4. Consent and notification: the responsible clinician must notify the patient of the names on his record's access control list when it is opened, of all subsequent additions, and whenever responsibility is transferred.
   - His consent must also be obtained, except in emergency or in the case of statutory exemptions

A Policy Example

5. Persistence: no-one shall have the ability to delete clinical information until the appropriate time period has expired.

6. Attribution: all accesses to clinical records shall be marked on the record with the subject's name, as well as the date and time.
   - An audit trail must also be kept of all deletions.

7. Information flow: Information derived from record A may be appended to record B if and only if B's access control list is contained in A's.

A Policy Example

8. Aggregation control: there shall be effective measures to prevent the aggregation of personal health information.
   - In particular, patients must receive special notification if any person whom it is proposed to add to their access control list already has access to personal health information on a large number of people.

9. Trusted computing base: computer systems that handle personal health information shall have a subsystem that enforces the above principles in an effective way.
   - Its effectiveness shall be subject to evaluation by independent experts.

Comparative Analysis

- Which of these three models—lattice, Chinese Wall and BMA — should be used in a given application?
- The lattice model on its own isn't enough, as it shows how to isolate compartments but not how to manage information flows between them.
- Both BMA and Chinese Wall tackle this problem, but BMA is as decentralized as possible, while in Chinese Wall the assignment of access rights is centralized.
- The fundamental policy decision is whether or not to centralize.
  - Can you cope better with lots of little traitors or with one big traitor?
- Medics, lawyers, and other professionals prefer the former, while spies seem to prefer the drama of the latter.

Part I Readings

- Ross Anderson, Security Engineering
  - Chapter 8, Sections 8.1 and 8.2.

Part II Outline

Banking and Bookkeeping

- Introduction
- How Bank Computer Systems Work
- Wholesale Payment Systems
- Automatic Teller Machines

Introduction

- Bookkeeping was for many years the main business of the computer industry, and banking was its most intensive area of application.
- Personal applications such as Netscape and Powerpoint might now run on more machines, but accounting is still the critical application for the average business.
- So the protection of bookkeeping systems is of great practical importance.
- It also gives us a well-understood model of protection in which confidentiality plays almost no role, but where the integrity of records (and their immutability once made) is of paramount importance.

Origins of bookkeeping

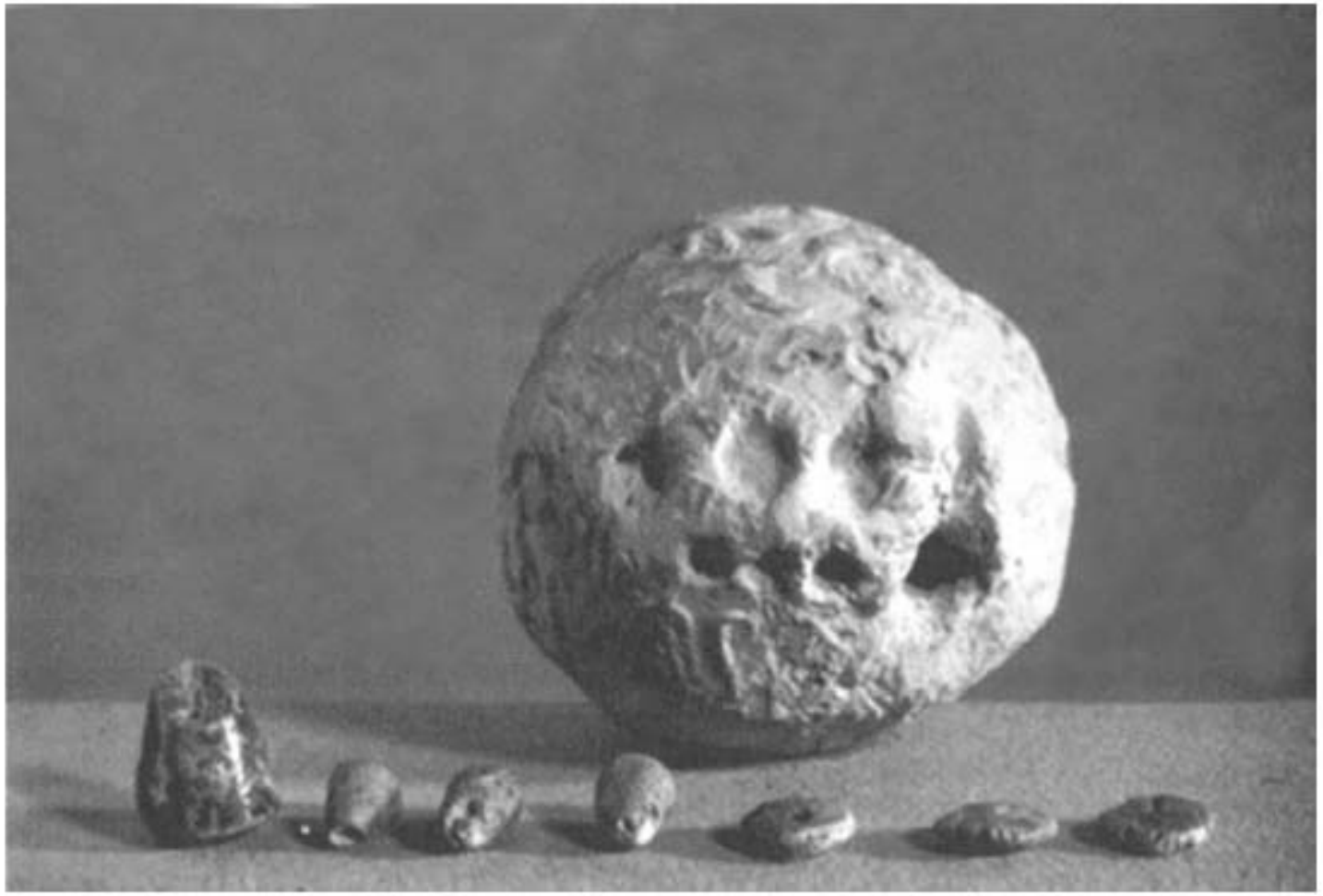

Figure 9.1 Clay envelope and its content of tokens from Susa, Iran, ca. 3300 BC (courtesy Denise Schmandt-Besserat and The Louvre Museum).

Double-entry Bookkeeping

- The idea behind double-entry bookkeeping is, like most hugely influential ideas, extremely simple.
- Each transaction is posted to two separate books, as a credit in one and a debit in the other.
- For example, when a firm is paid $100 by a debtor, the amount is entered as a debit in the accounts receivable book (the firm is now owed $100 less) and as a credit in the cash account book (the firm now has $100 more cash).
- At the end of the day, the books should *balance*, that is, add up to zero;
  - the assets and the liabilities should be equal.

Dual Control

- In all but the smallest firms, the books will be kept by different clerks, and have to balance at the end of every month (at banks, every day).
- By suitable design of the ledger system, we can see to it that each shop, or branch, can be balanced separately.
- Thus most frauds will need the collusion of two or more members of staff; and this principle of *split responsibility*, also known as *dual control*, is complemented by audit.

Double entry and software

- Many computer systems are used for bookkeeping tasks, and implement variations on the double-entry theme.
- However, the control is often illusory.
- The double-entry features may be implemented only in the user interface, while the underlying file formats may have no integrity controls.
- And even if the ledgers are kept on the same system, someone with root access — or with physical access and a debugging tool — may be able to change two or more records so that the balancing controls are bypassed.

Part II Outline

Banking and Bookkeeping

- Introduction
- How Bank Computer Systems Work
- Wholesale Payment Systems
- Automatic Teller Machines

How Bank Computer Systems Work

- A typical banking system has a number of data structures.
- There is an *account master file*, which contains each customer's current balance together with previous transactions for a period of perhaps 90 days.
- A number of *ledgers*, which track cash and other assets on their way through the system.
- Various *journals*, which hold transactions that have been received from teller stations, cash machines, check sorters, and so on, but not yet entered in the ledgers.
- An *audit trail* that records which staff member did what and when.

Bank processing software

- The processing software that acts on these data structures will include a suite of overnight batch-processing programs, which apply the transactions from the journals to the various ledgers and the account master file.
- The online processing will include a number of modules that post transactions to the relevant combinations of ledgers.
- For example, when a customer pays $100 into a savings account, the teller will make a transaction that records a debit to the savings account ledger of $100 (the bank now has an increased liability to the customer), while crediting the same amount to the ledger recording the amount of cash in the drawer.
- The fact that all the ledgers should always add up to zero provides an important check; if the bank (or one of its branches) is ever out of balance, an alarm will go off and people will start looking for the cause.

The Clark-Wilson Security Policy Model

- Although such systems have been in the field since the 1960s, a formal model of their security policy was only introduced in 1987, by David Clark and David Wilson (the former was a computer scientist, and the latter an accountant).
- In their model, some data items are constrained so that they can be acted on only by a certain set of transformation procedures.

The Clark-Wilson Security Policy Model

- More formally, there are special procedures whereby data can be input — turned from an *unconstrained data item*, or UDI, into a *constrained data item*, or CDI;
- *Integrity verification procedures* (IVPs) to check the validity of any CDI (e.g., that the books balance).
- *Transformation procedures* (TPs), which may be thought of in the banking case as transactions that preserve balance.
  - In the general formulation, they maintain the integrity of CDIs.

The Clark-Wilson Rules

1. The system will have an IVP for validating the integrity of any CDI.
2. The application of a TP to any CDI must maintain its integrity.
3. A CDI can only be changed by a TP.
4. Subjects can only initiate certain TPs on certain CDIs.
5. Triples must enforce an appropriate separation of duty policy on subjects.
6. Certain special TPs on UDIs can produce CDIs as output.
7. Each application of a TP must cause enough information to reconstruct it to be written to a special append-only CDI.
8. The system must authenticate subjects attempting to initiate a TP.
9. The system must let only special subjects (i.e., security officers) make changes to authorization-related lists

Remarks about Clark-Wilson

- A number of things bear saying about Clark-Wilson.
- First, unlike Bell-LaPadula, Clark-Wilson involves maintaining state.
  - This is usually necessary for dual control as you have to keep track of which transactions have been partially approved — such as those approved by only one manager when two are needed.
- Second, the model doesn't do everything.
  - It captures the idea that state transitions should preserve an invariant, such as balance, but not that state transitions should be correct. Incorrect transitions, such as paying into the wrong bank account, are still allowed.

Remarks about Clark-Wilson

- Third, Clark-Wilson skips the hardest question, namely: how do we control the risks from dishonest staff?
  - Rule 5 says that "an appropriate separation of duty policy" must be supported, but nothing about what this means.

Separation of Duties

- There are basically two kinds of separation of duty policy: dual control and functional separation.
- In dual control, two or more different staff members must act together to authorize a transaction.
- The classic military example is in nuclear command systems, which may require two or three officers to turn their keys simultaneously.

Separation of Duties

- With functional separation of duties, two or more different staff members act on a transaction at different points in its path.
- The classic example is corporate purchasing.
  - A manager makes a purchase decision and tells the purchasing department
  - A clerk there writes a purchase order
  - The store clerk records the arrival of goods
  - An invoice arrives at accounts
  - The accounts clerk correlates it with the purchase order and the stores receipt, and cuts a check
  - The accounts manager signs the check.

What Goes Wrong

A typical computer crime case:

- A bank had a system of suspense accounts, which would be used temporarily if one of the parties to a transaction could not be identified (such as when an account number was entered wrongly on a funds transfer).
  - This was a workaround added to the dual control system to deal with transactions that got lost or otherwise couldn't be balanced immediately.
  - As it was a potential vulnerability, the bank had a rule that suspense accounts would be investigated if they were not cleared within three days.
- One of the clerks exploited this by setting up a scheme whereby she would post a debit to a suspense account and an equal credit to her boyfriend's account; after three days, she would raise another debit to pay off the first.

What Goes Wrong

- In almost two years, she netted hundreds of thousands of dollars.
- The bank negligently ignored a regulatory requirement that all staff take at least 10 consecutive days' vacation no more than 15 months from the last such vacation.
- In the end, she was caught when she could no longer keep track of the growing mountain of bogus transactions.

Part II Outline

Banking and Bookkeeping

- Introduction
- How Bank Computer Systems Work
- Wholesale Payment Systems
- Automatic Teller Machines

SWIFT

- The Society for Worldwide Interbank Financial Telecommunications (SWIFT) was set up in the 1970s by a consortium of banks to provide a more secure and efficient means of sending payment instructions between member banks.
- It can be thought of as an email system with built-in encryption, authentication, and nonrepudiation services.

SWIFT

- Authenticity of messages is assured by computing a message authentication code (MAC) at the sending bank and checking it at the receiving bank.
- Formerly, the keys for this MAC were managed end-to-end.
  - whenever a bank set up a relationship overseas, the senior manager who negotiated it would exchange keys with her opposite number, whether in a face-to-face meeting or afterward by post to each other's private addresses.

Architecture of SWIFT

- Banks in each country sent their messages to a *regional general processor* (RGP), which logged them and forwarded them to SWIFT, which also logged them and sent them on to the recipient bank via the RGP in its country, which also logged them.
- The RGPs were generally run by different facilities management firms.
- Thus, a bank (or a crooked bank employee) wishing to dishonestly repudiate a done transaction — or claim that one had been done when it hadn't — would have to subvert not just SWIFT itself, but also two independent local contractors (in order to alter their log entries).
- Logs can be a powerful evidential resource, and are much easier for judges to understand than cryptography.

Architecture of SWIFT

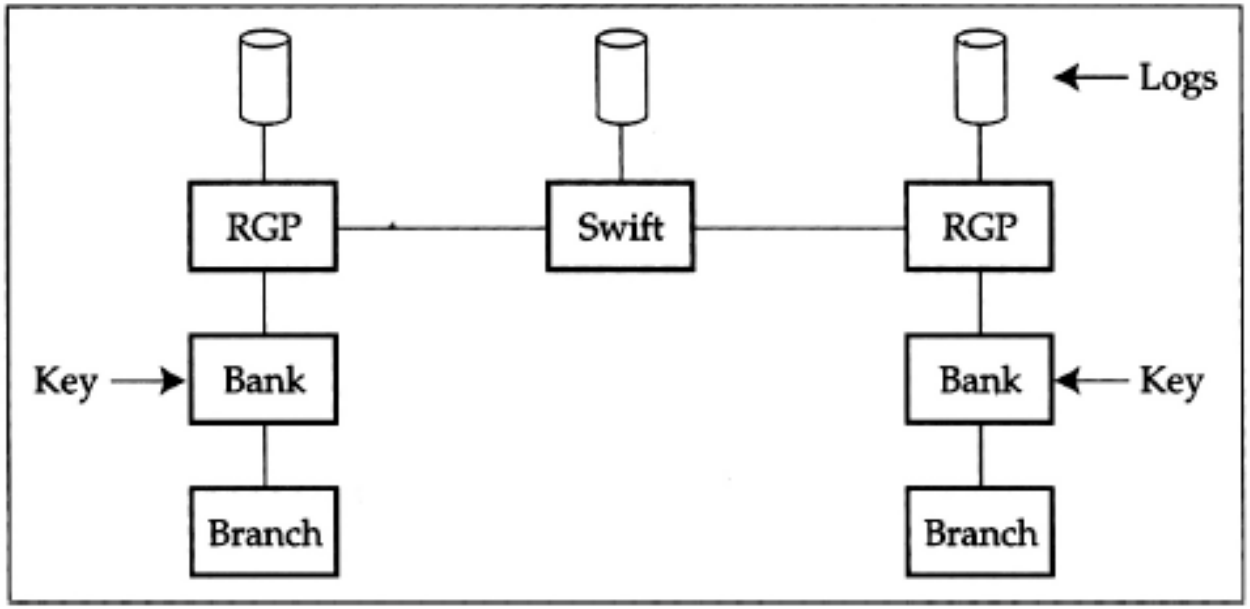

Figure 9.2 Architecture of SWIFT.

What Goes Wrong

- SWIFT I ran for 20 years without a single report of external fraud. In the mid-1990s, it was enhanced by the addition of public key mechanisms.
- The typical attack comes from a bank programmer inserting a bogus message into the processing queue.
  - It usually fails because he does not understand the other controls in the system or the procedural controls surrounding large transfers.
- Other possible technical attacks — such as inserting Trojan software into the PCs would also run up against these controls.
- **In fact, most large-scale bank frauds that "worked" have not used technical attacks but exploited procedural vulnerabilities,**

Procedural Vulnerabilities 1

- The classic example is a letter of guarantee.
- It is common enough for a company in one country to guarantee a loan to a company in another.
- This can be set up as a SWIFT message or even a paper letter. But as no cash changes hands at the time, the balancing controls are inoperative.
- If a forged guarantee is accepted as genuine, the "beneficiary" can take his time borrowing money from the accepting bank, laundering it, and disappearing.
- Only when the victim bank realizes that the loan has gone, and tries to call in the guarantee, is the forgery discovered.

Procedural Vulnerabilities 2

- Perhaps the best-known funds transfer fraud occurred in 1979 when Stanley Rifkin, a computer consultant, took over $10 million from Security Pacific National Bank.
- He circumvented the money-laundering controls by agreeing to buy a large shipment of diamonds from a Russian government agency in Switzerland.
- He got the transfer into the system by observing an authorization code used internally when dictating transfers to the wire transfer department, and simply used it over the telephone (a classic example of dual control breakdown at a system interface).
- He even gave himself extra time to escape by doing the deal just before a U.S. bank holiday. Where he went wrong was in not planning what to do after he collected the stones.
- If he had hidden them in Europe, gone back to the United States, and helped investigate the fraud, he might well have got away with it; as it was, he ended up on the run and got caught.

Part II Outline

Banking and Bookkeeping

- Introduction
- How Bank Computer Systems Work
- Wholesale Payment Systems
- Automatic Teller Machines

ATMs

- Another reason that dual control — although necessary — is not sufficient, emerges from the study of “phantom withdrawals” — complaints of unauthorized cash withdrawals from *automatic teller machines* (ATMs).
- ATMs, also known as cash machines, have been one of the most influential technological innovations of the twentieth century.
- Quite apart from their social and economic impact, they are just as important to the security engineer both as a source of technology and as a case study.

ATMs as a “killer app”

- ATMs were the first large-scale retail transaction processing systems. They have been around since 1968.
- The technology developed for them is now also used in terminals for *electronic funds transfer at the point of sale* (EFTPOS, or just POS) in shops.
- Modern block ciphers were first used on a large scale in ATM networks, to generate and verify PINs in secure hardware devices located within the ATMs and at bank computer centers.
- This technology, including block ciphers, tamper-resistant hardware, and the supporting protocols, ended up being used in many other applications, from postal franking machines to lottery ticket terminals.
- ATMs were the “killer app” that got modern commercial cryptology off the ground.

ATM Basics

- Many ATMs operate using some variant of a system developed by IBM for its 3614 series cash machines in the mid-1970s.
- This uses a secret key, called the *PIN key*, to encrypt the account number, then decimalize it and truncate it.
- The result of this operation is called the *natural PIN*; an offset can be added to it to give the PIN the customer must enter.
- The offset has no real cryptographic function; it just enables customers to choose their own PIN.

ATM Basics

Account number N (on the mag stripe):	8807012345691715
PIN key KP:	FEFEFEFEFEFEFEFE
Result of DES $\{N\}_{KP}$:	A2CE126C69AEC82D
$\{N\}_{KP}$ decimalized:	0224126269042823
Natural PIN:	0224
Offset:	6565
Customer PIN:	6789

Figure 9.3 IBM method for generating bank card PINs.

Dual Control in ATMs

- Dual control is implemented in this system using tamper-resistant hardware.
- A cryptographic processor, often called a ***security module***, is kept in the bank's central computer room. It will perform a number of defined operations on customer PINs and on related keys in such a way that:
- Operations on the clear values of customer PINs, and on the keys or other material needed to compute them or used to protect them, are all done in tamper-resistant hardware and the clear values are never made available to any single member of the bank's staff.
  - For example, the cards and PINs are sent to the customer via separate channels.
  - The cards are personalized in a facility with embossing and mag-strip printing machinery.
  - The PIN mailers are printed in a separate facility containing a printer attached to a security module.

ATM: the terminal master key

- A *terminal master key* is supplied to each ATM in the form of two printed components, which are carried to the branch by two separate officials, input at the ATM keyboard, and combined to form the key.
- Similar procedures are used to set up keys between banks and network switches such as VISA.
- If ATMs are to perform PIN verification, the PIN key is encrypted under the terminal master key, then sent to the ATM.
- If the PIN verification is to be done centrally over the network, the PIN is encrypted under a key that is set up using the terminal master key, and sent from the ATM to the security module for checking.

ATM: the terminal master key

- If the bank's ATMs are to be networked with other banks', then one uses transactions that will take an encrypted PIN from one source (such as encrypted under an ATM key), decrypt it, and re-encrypt it for its destination (such as using a key shared with VISA).
- This *PIN translation* function is done entirely within the hardware security module, so that clear values of PINs are never available to the bank's programmers.

ATMs difficulties

- But extending the dual control security policy from a single bank to tens of thousands of banks worldwide, as modern ATM networks do, proved not to be completely straightforward:
- When people started building ATM networks in the mid-1980s, many banks used software encryption rather than hardware security modules to support the machines.
  - So in theory, any bank's programmers might get access to the PINs of any other bank's customers. The remedy was to push through standards for security module use.

Large networks of ATMs

- It is not feasible for 10,000 banks to share keys in pairs, so each bank connects to a switch provided by an organization such as VISA or Cirrus, and the security modules in these switches translate the traffic.
- The switches also do accounting, and enable banks to settle their accounts for each day's transactions with all the other banks in the system, by means of a single electronic debit or credit.
- The switch is highly trusted; if something goes wrong, the consequences could be severe.

Attacks to ATMs

- There are many other ways in which ATM networks can be attacked in theory.
- For example, they mostly use single-key DES encryption, even for top-level keys, and DES can now be broken by exhaustive keysearch.
- However, one of the interesting things about these systems is that they have now been around long enough, and have been attacked enough by both insiders and outsiders, to give us a lot of data points on how such systems fail in practice.

What Goes Wrong with ATMs

- The engineers who designed ATM security systems in the 1970s and 1980s assumed that criminals would be relatively sophisticated, fairly well informed about the system design, and rational in their choice of attack methods.
- In addition to worrying about the many banks that were slow to buy security modules, and about the implementation loopholes such as omitting authentication codes on authorization responses, they agonized over whether the encryption algorithms were strong enough, and whether the tamper-resistant boxes were resistant enough.
- They were afraid that a maintenance engineer could disable the tamper sensing circuitry on one visit, and extract the keys on the next.

What Goes Wrong with ATMs

- They worried whether the random-number generators used to manufacture keys were random enough.
- And a very serious concern was that they just couldn't enforce dual control properly. Bank managers considered it beneath their dignity to touch a keyboard, so rather than entering the ATM master key components themselves after a maintenance visit, most of them would just give both key components to the ATM engineer.
- They believed that sooner or later a repairman would get his hands on a bank's PIN key, forge cards in industrial quantities, close down the whole system, and wreck public confidence in electronic banking.

Causes of phantom withdrawals

- The bulk of the actual phantom withdrawals, however, have one of the following three simple causes:
  - *Simple processing errors account*
  - *Thefts from the mail*
  - *Frauds by bank staff*

Simple processing errors

Simple processing errors account for a lot of disputes.

- With U.S. customers making something like 5 billion ATM withdrawals a year, even a system that makes only one error per 100,000 transactions will give rise to 50,000 disputes a year.
- In practice, the error rate seems to lie somewhere between 1 in 10,000 and 1 in 100,000.
- One source of errors tracked down was that a large bank's ATMs would send a transaction again if the network went down before a confirmation message was received from the mainframe;
  - Periodically, the mainframe itself crashed, and "forgot" about open transactions.

Thefts from the mail

Thefts from the mail are also huge.

- They are reckoned to account for 30 percent of all U.K. payment card losses, but most banks' postal control procedures are dismal.
- For example, a customer asked the bank for an increased card limit: the bank sent not one, but two, cards and PINs through the post.
- These cards arrived only a few days after intruders had got hold of the apartment mail and torn it up looking for valuables.
- It turned out that this bank did not have the systems to deliver a card by registered post. (The person had asked them to send the card to the branch for him to pick up, but someone at the branch had simply readdressed the envelope to the person.)
- Since then, many banks have found that better postal controls are the one way to solve these problems.

Frauds by bank staff

Frauds by bank staff appear to be the third major cause of phantoms.

- For example, in Paisley, Scotland, an ATM repairman installed a portable computer inside an ATM to record customer card and PIN data, then went on a spending outburst with forged cards.
- In London, England, a bank stupidly used the same cryptographic keys in its live and test systems;
  - The maintenance staff found out that they could work out customer PINs using their test equipment, and started offering this as a service to local criminals at £50 a card.
- Such frauds are particularly common in countries such as Britain, where banks had for many years a policy of denying that their cash machines could possibly make an error. Bank staff knew that customer complaints would be stonewalled rather than investigated.

Online processing of ATMs

- These failures are all very much simpler and more straightforward than the ones the engineers had worried about.
- In fact, the only fraud they had anticipated, and that happened to any great extent, came from the practice (common in the 1980s) of letting ATMs process transactions while the network was down or the central mainframe was offline.
- Though this was convenient — it meant 24-hour service — criminals, especially in Italy and England, learned to open bank accounts, duplicate the cards, then use them to withdraw money simultaneously from a large number of ATMs overnight when the network was down.
- Such frauds led most banks to make ATM operation online only by the mid-1990s.

Copying the account number

- A favorite modus operandi was for criminals to stand in ATM queues, observe customers' PINs, pick up the discarded ATM tickets, copy the account numbers from the tickets to blank cards, and use these to loot the customers' accounts.
- This trick was first reported in New York in the mid-1980s; it was still working in the San Francisco Bay Area in the mid-1990s.
- Yet there are many simple countermeasures, such as incorporating extra data on the mag strip, or just not printing the full account number on the ticket.

Programming Errors

- One bank's systems had this feature: when a telephone card was entered at an ATM, it believed that the previous card had been inserted again.
- Crooks stood in line, observed customers' PINs, and helped themselves.
- This seems to have been an obscure programming error involving the card reader's error handler; one can't expect all such errors to be found during testing.☺

Programming Errors

- One make of ATM would output 10 banknotes from the lowest-denomination nonempty cash drawer whenever a certain 14-digit sequence was entered at the keyboard.
- One bank printed this sequence in its branch manual, and three years later there was a sudden spate of losses.
- These went on until all the banks using the machine put in a software patch to disable the transaction.

Simple Programming Error ☺

- One small institution issued the same PIN to all its customers, as a result of a simple programming error.

Disregarding prudent procedures

- Some banks show a complete disregard for prudent procedure.
- For example, someone went into a branch of her bank and said that she'd forgotten her PIN.
- The teller helpfully printed her a new PIN mailer from a printer attached to a PC behind the counter.
- There were no visible dual controls.
- Worse, this was not the branch where she kept the account. Nobody knew her and the only identification she offered was a bank card and her checkbook.
- When procedural controls are so lax that anyone can walk in off the street and get a PIN for a random customer account, no amount of encryption technology will do much good.

Installing False Terminals

- A rapidly growing modus operandi is to use false terminals to collect customer card and PIN data.
- Attacks of this kind were first reported from the United States in 1988 where crooks built a vending machine that would accept any card and PIN, and dispense a packet of cigarettes.
- They put their invention in a shopping mall, and harvested PINs and magnetic strip data by modem.

Conclusions

- From Ross Anderson:
  - *In conclusion, the main thing we did wrong when designing ATM security systems in the early to mid-1980s was to worry about criminals being clever; we should rather have worried about our customers — the banks' system designers, implementers, and testers—being stupid.*

Conclusions

- Crypto is usually only part of a very much larger system.
- It gets a lot of attention because it is mathematically interesting; but as correspondingly little attention is paid to the "boring" bits such as training, usability, standards, and audit, it's rare that the bad guys have to break the crypto to compromise a system.
- It's also worth bearing in mind that there are so many users for large systems, such as ATM networks, that we must expect the chance discovery and exploitation of accidental vulnerabilities that were simply too obscure to be caught in testing.

Part I Readings

- Ross Anderson, Security Engineering
  - Chapter 9.

Good Luck

- Hope you enjoyed the course!
- See you in the FINAL EXAM
  - And Project Discussion ☺!