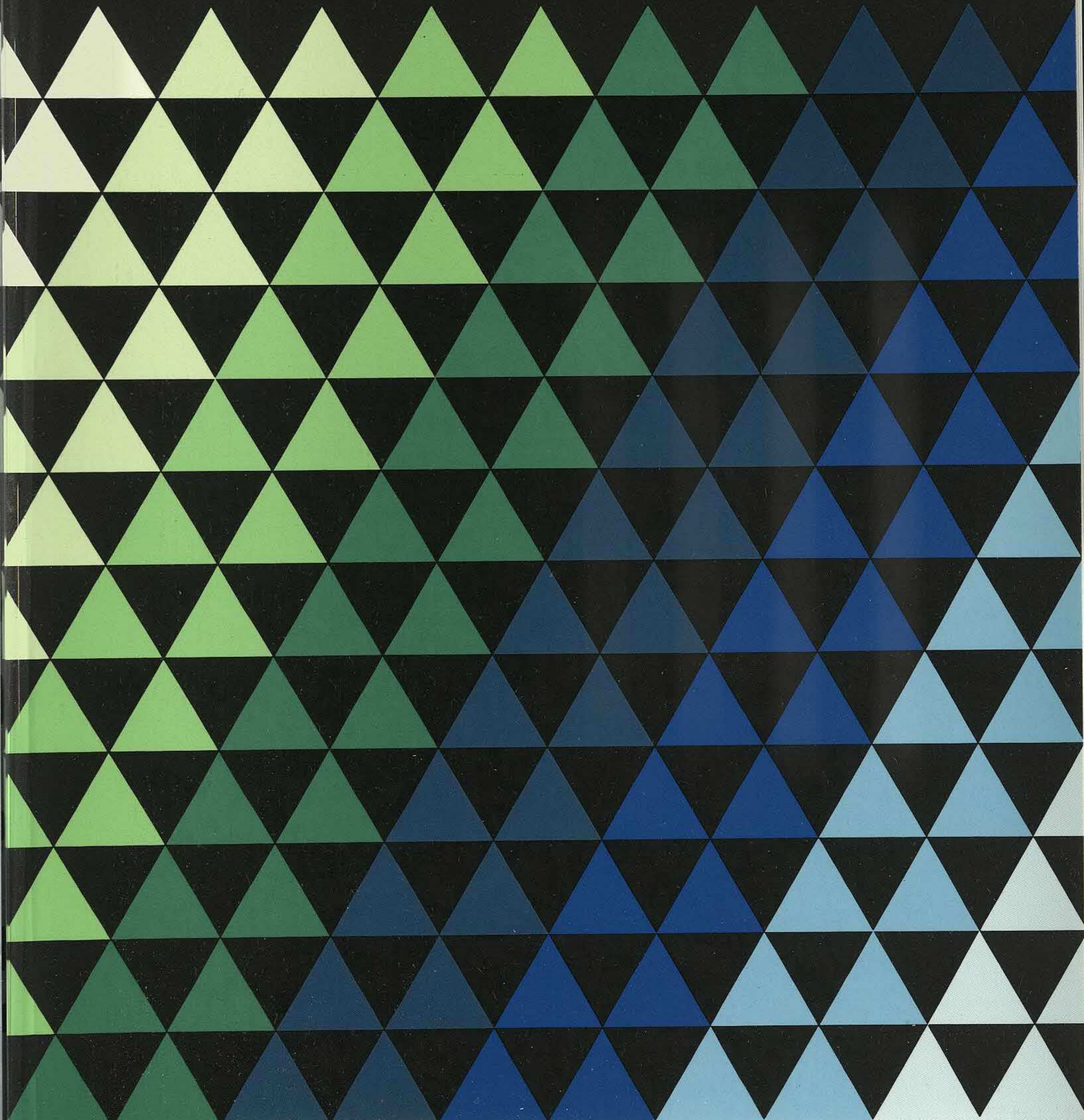
Systems Security

BUTLER COX
FOUNDATION

Research Report 76, August 1990



BUTLER COX
FOUNDATION

Systems Security

Research Report 76, August 1990

Butler Cox plc

LONDON
AMSTERDAM MUNICH PARIS

Published by Butler Cox plc
Butler Cox House
12 Bloomsbury Square
London WC1A 2LL
England

Availability of reports
Members of the Butler Cox Foundation receive three copies of each report upon publication; additional copies and copies of earlier reports may be purchased by members from Butler Cox.

Photoset and printed in Great Britain by Flexiprint Ltd., Lancing, Sussex.

BUTLER COX FOUNDATION

Systems Security

Research Report 76, August 1990

Contents

A Management Summary of this report has been published separately and distributed to all Foundation members. Additional copies of the Management Summary are available from Butler Cox.

BUTLER COX FOUNDATION

Report synopsis

This report is concerned with the security necessary to protect the confidentiality, integrity, and availability of information. Total security would, however, be prohibitively expensive and unacceptably restrictive. Systems managers are therefore obliged to make trade-offs, with a view to achieving acceptable security at a reasonable cost. To do this, they need to be constantly developing and improving the reliability of methods for identifying threats, and for preventing and detecting breaches of security. It is a long-term process because political, business, legal, and cultural changes mean that security procedures can never be considered permanent.

Chapter 1

The need for a systems security policy

The subject of systems security attracts a lot of attention in the press and other media. Scare stories abound about multimillion-dollar financial frauds, malevolent viruses, and the devastating exploits of computer hackers. While most organisations are never likely to experience all these kinds of calamities, systems security is, quite rightly, a matter of growing concern worldwide. Most systems managers are well aware that threats to security *are* real — that inadequate security leads to losses of assets, money, time, effort, reputation, and competitive edge. Everyone would like security to be better.

It is, however, very difficult to assess the true nature of the threats to systems security. Such threats are exceedingly diverse in nature. They may derive from obvious causes like fire and water, from accident, from error, from some form of sabotage, or from unidentified risks. What makes the subject especially complicated, however, is that each threat can affect every systems asset within the organisation. There are therefore potentially thousands of risks.

Almost as many measures might be put in place in an effort to counteract these potential threats, some of which are extremely unlikely to occur, but would have a catastrophic effect if they did. The complex nature of the potential threats to systems security and the way in which they interact mean that security will be only as good as the weakest link in any set of measures put in place to counteract them — problems will occur where interfaces are weak, hackers will work around the existing measures that protect computers, and burglars will not bother to pick a good lock if they can quickly kick down a weak plasterboard wall.

While the dramatic stories in the media do nothing to encourage a rational approach to systems security, bringing together all the elements that are essential for good security does require a coherent and comprehensive policy, endorsed by senior management, and reviewed periodically in the light of political, business, and cultural changes. With such a policy in place, systems managers will be in a position to allocate sensible priorities to the various aspects of security, to control costs, to assign appropriate responsibilities, and to be confident in the knowledge that as changes occur in the business environment, security is maintained.

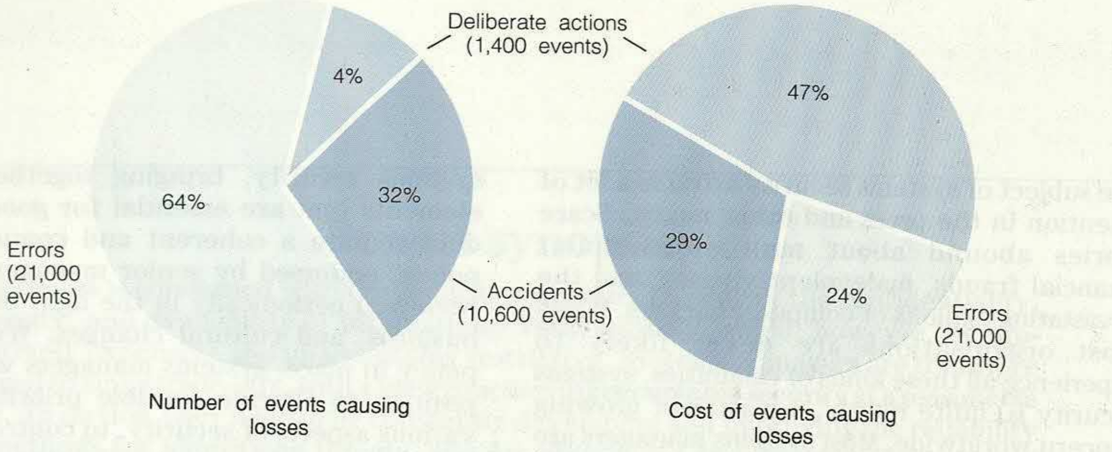
Security threats are real

Losses resulting from breaches of security may be divided into three main categories — deliberate actions (fraud and sabotage), accidents, and errors. Figures on the incidence of such events and the losses resulting from them are extremely difficult to come by. A French insurance organisation, Les Sociétés d'Assurance contre l'Incendie et les Risques Divers (APSAIRD), does, however, regularly report losses in France. During 1988, these amounted to some FFr8,130 million ($1.3 billion). The number of events causing these losses and the costs incurred in each category are illustrated overleaf in Figure 1.1.

The 1988 figure represents a 3 per cent increase over 1987 and a 560 per cent increase over 1984. We believe that the enormous increase over 1984 is primarily due to a widening of the definition of a computer-related loss. As it became more and more unusual to transact business without the involvement of a computer, a larger proportion of losses became classified as 'computer-related'. The same phenomenon is evident in the United Kingdom, where losses are estimated to have increased by a factor of 10 over the last six years. In the

Figure 1.1 Threats to systems security are real — tens of thousands of events cause losses each year, at an enormous cost

The pie charts show the number and cost of events causing losses in France in 1988. We believe that the figures are reasonably indicative of what happens in other European countries.



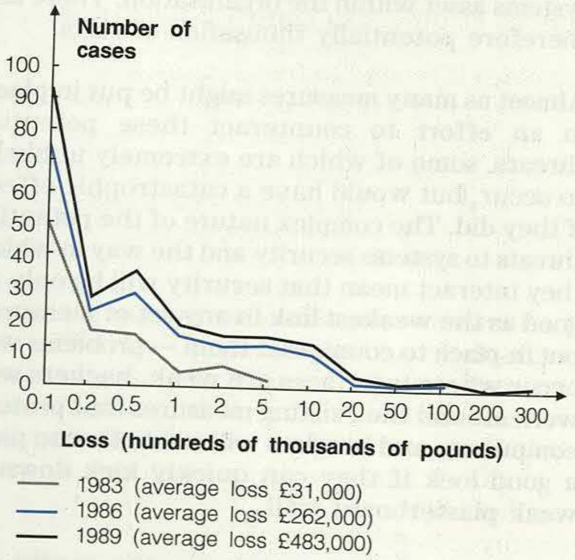
(Source: APSAIRD, 1988)

United States, however, where loss reporting has been of a much more open nature for many years, and where the definition of computer-related loss is well established, losses are estimated to have increased by a factor of only 2 to 3 over a similar period. We believe that the current low annual increase in computer-related loss (3 per cent) reported by APSAIRD is a reflection of the increased awareness in Europe of the need for systems security and the effectiveness of the countermeasures applied. Nevertheless, we suspect that the rate of increase in Europe is likely to climb nearer to the American rate of increase, which is currently about 20 per cent a year.

That losses also occur in other European countries is indisputable, and APSAIRD's figures are likely to be indicative of what happens elsewhere. Miscellaneous statistics and reports appear in the press. Many more incidents undoubtedly take place, but the victims, understandably, choose to remain silent.

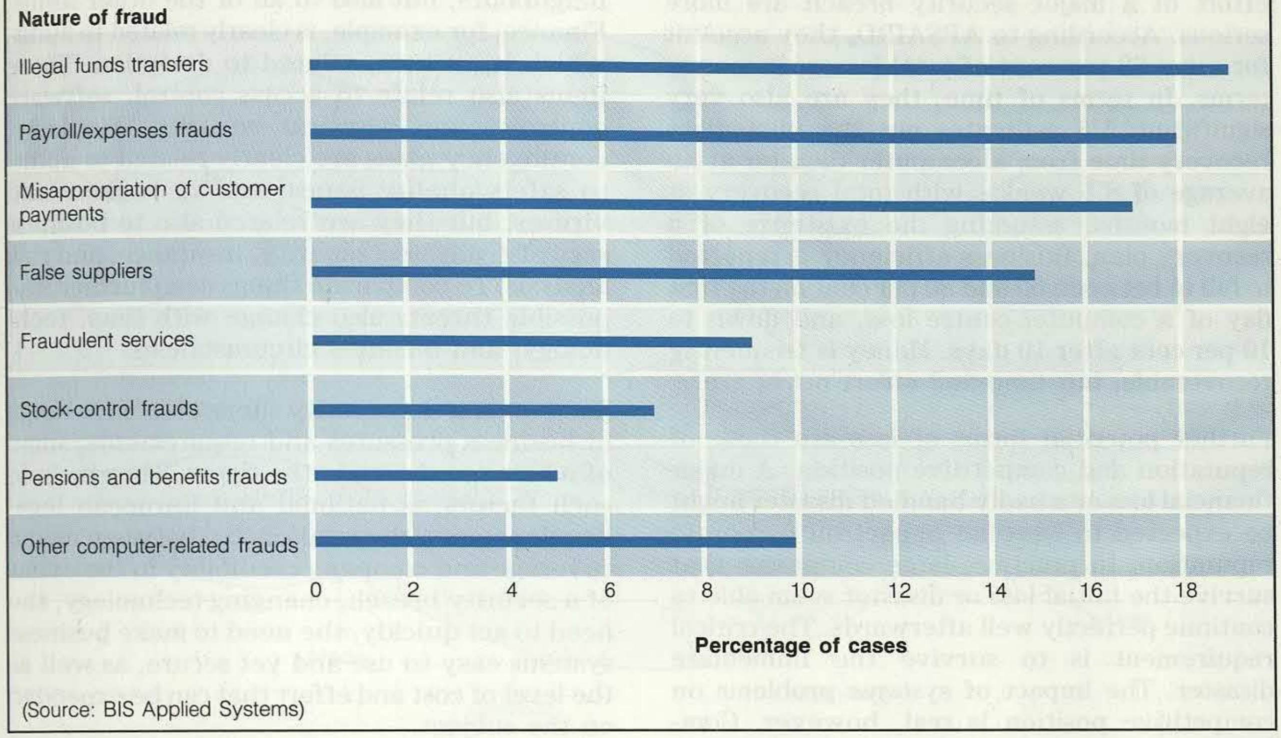
In the United Kingdom, *fraud* (including computer-related fraud) is reported to cost between £400 million ($700 million) and £2.5 billion ($4 billion) annually. The distribution of losses specifically due to computer-related fraud is illustrated in Figure 1.2 (size of loss) and Figure 1.3 (nature of fraud). In France, some 70 cases of computer-related crime were reported in 1989, including a loss of FFr10 million ($1.7 million) by a stockbroking company. In Switzerland, Manufacturers Hanover reported

Figure 1.2 Very few cases of computer-related fraud in the United Kingdom in 1989 involved more than £1 million ($1.75 million)



(Source: BIS Applied Systems)

Figure 1.3 In the United Kingdom, the most common types of computer-related fraud in 1988 were illegal funds transfers and payroll frauds



(Source: BIS Applied Systems)

a write-off of $50 million due to fraudulent foreign-exchange transactions. During the period 1987 to 1989, attempts were made to defraud major Swedish corporations of some Skr580 million ($95 million). Much more common, however, are smaller-scale, less spectacular frauds, carried out over longer periods of time, as Figure 1.2 illustrates.

Sabotage is related to fraud in the sense that it is deliberately instigated. Within Europe, some 60 computer-related incidents of sabotage have been reported over the last 10 years. Some of them were extremely violent, resulting in major damage to, or total destruction of, computer sites.

Accidents may be caused by people, or be the result of natural disasters. During the winter storms of 1989, both the Rhine and Seine rivers flooded, causing damage to computer installations. Burst water mains in London had a similar effect. High winds interrupted power supplies and halted computer systems. Fire destroyed one of Digital Equipment Corporation's customer-service administration offices, causing damage estimated at £38 million ($65 million), and during October 1989, earthquakes damaged computer and power installations in San Francisco and over 70 buildings in the Bay area were damaged beyond repair.

Errors can create problems throughout a business and may lead to real or potential security problems. In one episode that was described to us during the research for this report, an illusive software fault blocked access to a user's complete database for two days. This fault severely restricted batch and online operations and created concern at board level. Back-up procedures would have been of only limited help because the problem was not in the hardware. The user could do nothing but wait until the problem was fixed. Another well publicised error occurred during January 1990 when a software fault seriously degraded AT&T's long-distance telephone system in the United States. Again, very little could be done until technicians were able to devise a temporary solution. Software flaws were also implicated in the failure of a US Therac medical linear accelerator system, resulting in the deaths of three people.

Direct costs are, of course, only one of the penalties of such breaches of systems security. The consequential losses in money, time, and effort of a major security breach are more serious. According to APSAIRD, they account for some 63 per cent of total losses, in money terms. In terms of time, they are also very significant: US estimates put the immediate recovery time from a computer disaster at an average of 8.5 weeks, with total recovery at eight months, assuming the existence of a recovery plan. Business efficiency is reported to fall to between 60 and 80 per cent on the first day of a computer-centre loss, and down to 10 per cent after 10 days. Money is frequently recoverable, but time and effort never are.

Further potential forms of loss are those of reputation and competitive position. A major financial loss or a badly handled disaster might be expected to have an impact on corporate reputation. In practice, large companies that survive the initial loss or disaster seem able to continue perfectly well afterwards. The critical requirement is to survive the immediate disaster. The impact of systems problems on competitive position is real, however. Companies can lose customers or have their business otherwise affected by bad service or lack of systems availability.

Assessing threats to systems security is difficult

Systems security is a very complex subject. Threats and the measures to counteract them are very varied, they interact with each other, and they cross disciplinary and functional boundaries. While there is a very low probability of some of the potential threats ever occurring, the damage could be catastrophic if such an event were to occur. There are almost as many possible measures that might be put in place to counteract these potential threats, but very little guidance is available in the form of international standards or even in guidelines on the subject.

Threats to security are very varied

Some of the potential threats to systems security and possible countermeasures are illustrated in Figure 1.4. It shows security as an apparently loose collection of disparate items that are drawn together simply because they have some influence on security. Each facet of security is connected in some way, not only to its neighbours, but also to all of the other items. Finance, for example, is clearly related to audit, which is, in turn, related to detection. These items also relate to access control, software controls, and physical security. Similarly, contingency plans are clearly related to audit, to safety/quality issues, and to hackers and viruses, but they are related also to business security, physical security, insurance, and risk analysis. To complicate things even further, the possible threats also change with time, technology, and business circumstances.

Surrounding the security subject is a wide range of business pressures and requirements, some of which are shown in the figure. They include such factors as national and European legal developments, the relationship between media coverage and company credibility in the event of a security breach, changing technology, the need to act quickly, the need to make business systems easy to use and yet secure, as well as the level of cost and effort that can be expended on the subject.

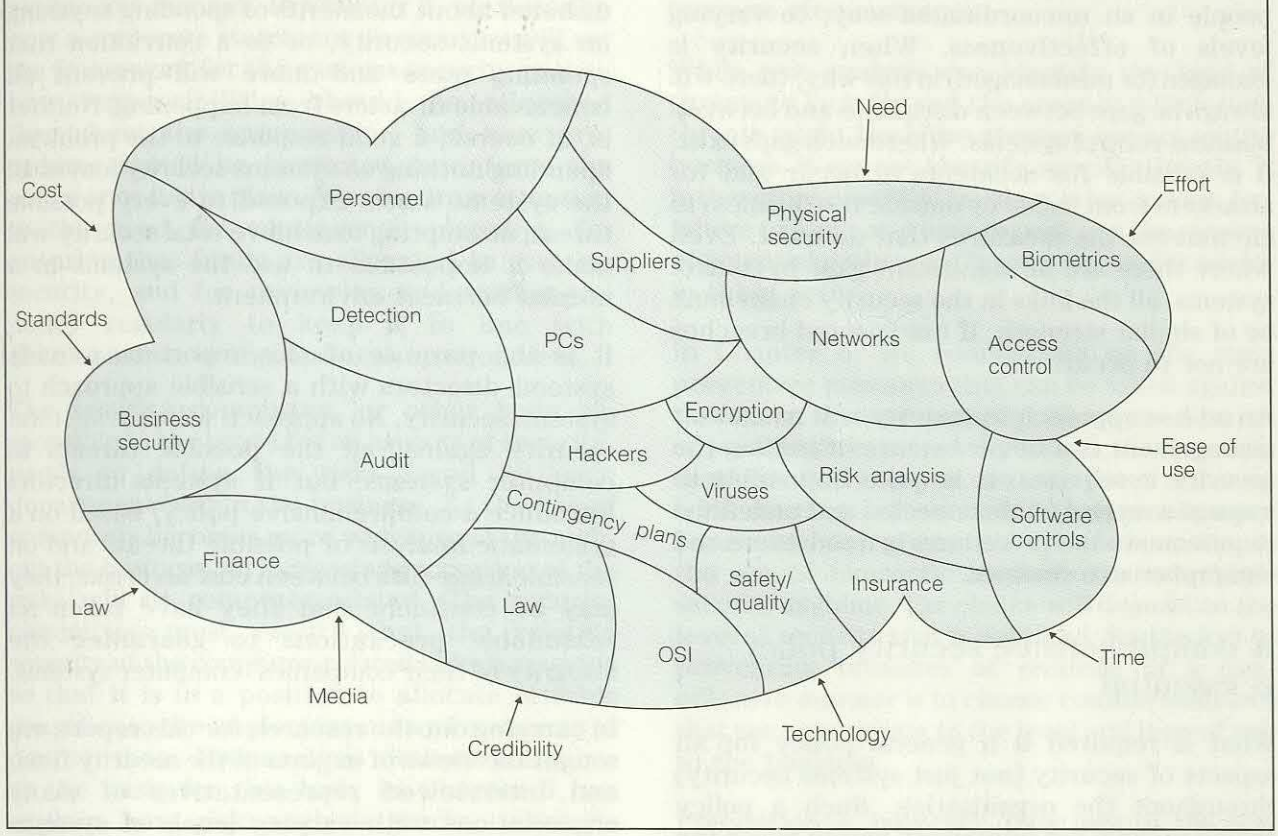
Many threats are of very low probability

Some potential threats have an extremely low probability and are practically impossible to prevent, but would have a devastating effect if they were to occur. The probability of an airliner crashing on a particular building, for example, is extremely small, but would be totally disastrous. Risk analysts call this level of catastrophic result from a low-probability event the zero-infinity dilemma (ZID). The danger of the ZID is that it can inhibit action on those activities about which something can be done. Deciding whether or not a risk is sufficiently serious to justify countermeasures depends almost entirely on the business environment. What is a serious risk in banking may be an acceptable risk in manufacturing, for example.

Many countermeasures are possible

Over the years, systems managers have learnt how to deal with many of the threats faced by business. A wide range of technical and non-technical countermeasures is possible. Some are

Figure 1.4 Security has many facets that have to be managed within the constraints imposed by business pressures and business requirements



complex and expensive; others require management or legal support. Few, however, are totally effective in isolation.

Countermeasures may be installed for three purposes — prevention, detection, and mitigation. To be complete, any practical set of countermeasures should include an element of each. In this way, management can attempt to prevent threats becoming risks, to detect risks if they occur, and should a risk occur, either to recover from the event or to limit the damage. Underlying each kind of countermeasure is a series of enabling techniques that can either singly, or in combination, be used to create countermeasures. Examples of these techniques include encryption, monitoring, data analysis, personnel policies, and software alarms.

Existing standards do not provide much useful guidance

For many years, European standards bodies have been concerned with standards for computers. Building on this, work is currently underway at CEN/CENELEC (the European Committee for Standardization and the European Committee for Electro-technical Standardization) and at the office of DG XIII at the European Community on the standardisation of electronic business procedures and their impact on security, risk analysis, and the need for policies, and perhaps even legislation. This work is an indication of the importance of systems security issues within international business, but much of it may take a long time to become commercially significant. In the meantime, there are no standard approaches to systems security, and the available methods are considered by many systems managers to be over-simple and incomplete.

Ad hoc measures are not adequate

The subject of systems security includes everything from locks on doors, staff honesty, and database-access control, to fire protection, the selection of buildings, and the political and cultural environment. These different aspects of security are usually handled by different people in an uncoordinated way, to varying levels of effectiveness. When security is managed (or mismanaged) in this way, there will always be gaps between disciplines and between business-control systems. Where such gaps exist, it is possible for accidents to occur and for attackers from inside or outside the business to circumvent the measures that do exist. Even where there are no significant gaps in control systems, all the links in the security chain must be of similar strength, if unexpected breaches are not to occur.

An ad hoc approach to security will mean that management can never be sure of getting the security it requires. It is quite impossible to manage a myriad of unconnected and undefined requirements and procedures in a consistent and comprehensive manner.

A comprehensive security policy is essential

What is required is a general policy for all aspects of security (not just systems security) throughout the organisation. Such a policy should provide the framework within which the detailed analysis of risks, the implementation of appropriate countermeasures, the maintenance of an appropriate level of security, and the testing of security procedures, can be put in place and controlled. The role of senior management is to initiate the preparation of such a policy, and to provide guidance about specific concerns and priorities that it should have. Senior management should also be prepared to endorse the policy to give it strength and effectiveness where required. The policy will need to be reviewed periodically at senior management level in the light of political, business, legal, and cultural changes. The systems security policy should then be developed within the framework of the overall policy on security.

Purpose and structure of the report

Systems security is a subject that is widely covered in the media, but the emphasis is usually on dramatic stories about the subversion of vastly expensive computer systems by gangs of master criminals, or by juveniles armed with a personal computer. The mismatch between such reports and the experience of most businesses tends to lead either to sceptical disbelief about the merits of spending anything on systems security, or to a conviction that spending more and more will prevent all conceivable disasters from happening. Neither is, of course, a valid response to the problem. Spending nothing on systems security leaves all the systems assets exposed to every possible threat; attempting to achieve total security will make it impossible to use the systems in a normal business environment.

It is the purpose of this report to provide systems directors with a sensible approach to systems security. No approach will provide total security against all the possible threats to computer systems, but if systems directors introduce a comprehensive policy, based on a systematic analysis of possible threats and on sensible trade-offs between cost and risks, they may be confident that they have taken all reasonable precautions to guarantee the integrity of their companies' computer systems.

In carrying out the research for this report, we sought the views of experts in the security field, and interviewed representatives of many organisations with varying levels of systems security provisions in place. We also reviewed a wide range of the technical and specialist literature on the subject of computer security. A selected bibliography is included at the end of the report for those who wish to delve more deeply into some of the more detailed aspects of systems security. The scope of the research and details of the research team are described in Figure 1.5.

We have also built on the findings in our previous report on computer security (Report 51, *Threats to Computer Systems*), which was published in 1986. Since then, the subject has attracted a great deal more attention. Legislation has been enacted in many countries to try to restrict computer-related crime, and there are plenty of techniques available to counter threats to systems security. The main concern now is how to manage them effectively.

It is imperative that organisations have a corporate security policy in place, which sets the guidelines for preparing the specific systems security policy. Within the context of a corporate policy, a corporate security ethic can be established and all aspects of systems security can be designed in a consistent and comprehensive manner. In Chapter 2, we describe how a corporate statement on security will set the framework for the systems security policy, how responsibilities should be allocated throughout the organisation, and how procedures should be instituted to enforce the policy once it is in place. We also draw attention to the need for contingency planning, for ensuring that funds are allocated to systems security, and for reviewing and testing the policy regularly to keep it in line with developments in the business environment.

The security committee, or other body responsible to the board for all aspects of security, needs to define the risks faced by each department within the business. As the use of computers becomes more widespread throughout the business, an increasing proportion of the risks will be computer-related. The security committee must ensure that the relative severity of the computer-related risks is assessed so that it is in a position to allocate sensible priorities to the implementation of measures to counter them. Risk analysis methods will serve as the basis for this task. In Chapter 3, we explain how risk analysis works, what its advantages and disadvantages are, and what factors ought to be considered in selecting an appropriate method for a particular set of business circumstances.

While risk analysis can identify the kind of threats to look for and the areas in which such threats might have the greatest impact on the business, it cannot identify specific threats. It is therefore essential to know what to look for, where to look, and how to seek out the sources of potential problems. Chapter 4 provides advice in these areas.

In Chapter 5, we concentrate on the main preventive measures that can be taken against threats to systems security. Since the sources of possible threats are very varied, the measures that can be taken to counter them are very wide-ranging — from making passwords more secure and using tokens for basic security, to the use of biometric methods for very-high-security systems. The choice will depend on the level of security that is required, for the key to preventing breaches of security in a cost-effective manner is to choose countermeasures that are appropriate to the level and type of risk to the business.

Prevention is, however, only part of the task of ensuring that systems are secure. Since no

Figure 1.5 Research team and scope of the research

The research for the report was carried out between November 1989 and June 1990 and was led by Roger Hart, a senior consultant with Butler Cox in London. He is a specialist in systems specification and telecommunications and has experience in the design of systems software. He was assisted, in particular, by:

— Simon Forge, a principal consultant in Butler Cox's Paris office, with considerable experience in the field of systems development and security-related issues.

— Lothar Schmidt, a senior consultant in Butler Cox's Munich office.

— Robin Sherman, an associate of Butler Cox, and a specialist consultant in the field of systems security.

— Adrian Norman, an associate of Butler Cox, who has, for many years, specialised in systems security issues.

Further research was carried out by John Cooper (Australia), Loredana Carpinella (Italy), Per Hansen (Sweden), and Onno Schroder (Netherlands).

As well as conducting an extensive review of the published literature, we conducted interviews with specialists in systems security, suppliers of security systems and software, and Foundation members, many of whom replied to the questionnaire sent out at the beginning of the research. Subsequently, we held a series of workshops in the United Kingdom, France, and the Netherlands to identify the primary sources of threats and ways of eliminating or controlling them. We followed this up with postal questionnaires to members in France and Germany, and telephone interviews with members in the United Kingdom.

We interviewed experts in security in France, Germany, Austria, Italy, Sweden, the United States, and the United Kingdom. In Germany, we spoke to a leading member of the Chaos Club (one of the best-known groups of hackers), and in the United Kingdom, to investigators from within the police and telecommunications-investigation agencies, to whom special thanks are due.

We also drew on the experience of our in-house experts and on our consulting work in the telecommunications and systems development areas. To confirm some of the claims made in the literature, we made use of dial-up modem services to access bulletin boards.

system can ever be totally secure without becoming unusable, it is important that measures are also in place to detect breaches that do occur. There must also be plans for responding quickly to any breaches of security, so that the damage caused by accidents or malpractice can be minimised. Advice on these two aspects of systems security is given in Chapter 6.

Chapter 2
Developing a corporate-wide security policy

In Chapter 1, we identified the need for a corporate-wide policy on overall security (including systems security), to ensure that all security issues are fully covered, that security is given the attention it deserves within the organisation, and that everyone understands what is required. In this chapter, we describe how a corporate statement on security will set the framework for the policy, how responsibilities should be allocated throughout the organisation, and how procedures should be instituted to enforce and test the policy once it is in place. We also draw attention to the need for contingency planning, for ensuring that funds are allocated to security (to ensure that the policy can continue to be properly implemented), and for reviewing the policy regularly to keep it in line with developments in the business environment and technology.

Issue a corporate statement on security

A security policy document should contain a corporate statement on overall security, with the objective of creating a security ethic within the company. In this sense, security is like quality; it is fundamentally an attitude of mind, and must permeate the activities of everyone in the organisation. The document should explain that the purpose of a corporate policy is to protect both the company and those whom it employs. It should make it clear that the policy applies to everyone, to a greater or lesser extent, and define the legal duties that everyone has. These will, of course, vary from country to country, but will normally include relevant corporate legislation, industry-specific legislation, and (for the systems security aspects of the policy) the implications of data-protection legislation.

The policy document should explain the degree of protection to be given to each of the company's assets. Each asset, and particularly a data asset, has three security-related properties — availability, confidentiality, and integrity. What is required in terms of *availability* can be fairly simply addressed by deciding how long the organisation could manage without a particular asset. The appropriate levels of *confidentiality* and *integrity* are decided by assessing what would happen if confidentiality were breached or if some of the records were wrong. Obviously, the degree of emphasis given to each of these properties may vary — for example, the lack of availability or integrity of a stock file may be disastrous for a manufacturer, while its confidentiality may be relatively unimportant; the confidentiality of a medical record, on the other hand, may be of greater importance than its availability. The appropriate levels of protection for each type of asset must be specified in the policy document.

The corporate statement on overall security should also make it clear where the responsibility lies for enacting the policy. Security may be controlled and audited from the centre, with security specialists defining the type and level of security to be applied. Alternatively, an 'ownership' approach may be adopted, in which those responsible for the asset also define the level of security required, and pay for it. The choice of approach will depend on corporate management style, the risks involved, and the technical skills required.

Allocate responsibilities at all levels

Responsibility for security in its widest sense clearly belongs at board level. The legal responsibility of board members for security is unclear, however. In the United States, security-related claims against directors and officers are increasing at a rate of 15 to 20 per cent annually. In West Germany, company officers may be made liable for negligent management.

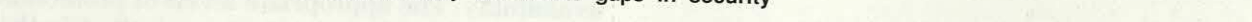
There is no need for the whole board to become involved. One member should chair a committee of experienced managers to formulate a comprehensive policy. These managers must know the business well, and ideally, have worked in several functional areas. The critical requirements are to provide control over, and support for, security, to set priorities, to draw attention to the residual risks arising from the agreed policy, to review operational changes to the policy, and to see that the policy is reviewed regularly.

Responsibility must be delegated downwards from the committee, with everyone bearing at least some basic level of responsibility. The systems security policy will obviously form an important part of the overall security policy, and clearly, some departments and staff will bear a heavier burden than others in this area — for example, internal audit, legal, and systems will have specific responsibilities for the systems-related aspects of the security policy. These responsibilities will be defined by the security committee on the basis of the results of the risk analysis that it will commission. Each department must be able to implement its part of the security policy, and each must be aware of the procedures and assets that will be audited as part of a regular security-audit process. Responsibilities for carrying out the security audit and its timing should also be defined.

The policy committee must consider how overall security is to be managed. Only very large organisations can justify a specialist security manager, responsible for physical security, systems security, document security, and so on. In most organisations, these tasks are shared among the building-management, computer-management, business-management, and personnel functions. If these are left as separately managed functions, security may prove to be unsatisfactory; they must be coordinated to avoid incomplete coverage, and hence, gaps in security. Figure 2.1 illustrates the point.

The figure also shows that the security policy and the rules and procedures for each individual area of security need to be audited and revised. The audit process may, at one extreme, take the form of checklists that are used to verify that

Figure 2.1 Fragmentation of responsibility will create gaps in security



the rules and procedures are being adhered to and are effective. At the other extreme, the policy, rules, and procedures may be formally reviewed by security specialists. The policy committee should ensure that the most appropriate form of security audit is carried out at regular intervals, and that any necessary revisions are made to the policy, rules, and procedures.

The risk analysis process, described in Chapter 3, will indicate whether there is a need for systems security specialists. In 70 per cent of the organisations we studied, the management of mainframe and minicomputer security was a full-time job for at least one member of staff. Simply managing a mainframe access-control package can occupy nearly all a specialist's time. Indeed, in the banking sector, there is often a whole department of systems security specialists.

Specify how the policy will be enforced

Because overall security is such an important matter, some sanctions must be put in place for failure to abide by the security measures. Ultimately, serious breaches of security may demand dismissal of staff or require legal action. This imposes a responsibility on those managing security to ensure that people are aware of the sanctions and that the preventive, audit, and detection mechanisms are sound enough to justify the sanctions.

In practice, organisations will need to institute a graduated series of enforcement procedures and measures. In one organisation we spoke to, the role of the systems security manager was to 'keep a fatherly eye' on users, and users, in turn, saw the security measures as an aid to good practice. At the other extreme, in an organisation with a more draconian approach to systems security, users were alienated from the security function. Because of the lack of trust, potential security breaches went unreported because no-one was prepared to admit to mistakes. A third organisation had created an unnecessary administrative layer in its approach to systems security by insisting that the personal computer support group report virus outbreaks to the security team, when members of the group could very well have cured them themselves. This approach led to disgruntled users and a conflict of loyalties within the personal computer support group.

Both the 'strict discipline' and the 'fatherly eye' approaches to enforcement should be reflected in those aspects of the policy that relate to systems security. In this way, systems users will be aware that serious offences are punishable, but that making errors or being responsible for omissions will not be treated as serious offences but will indicate where help and guidance are required.

Plan for contingencies

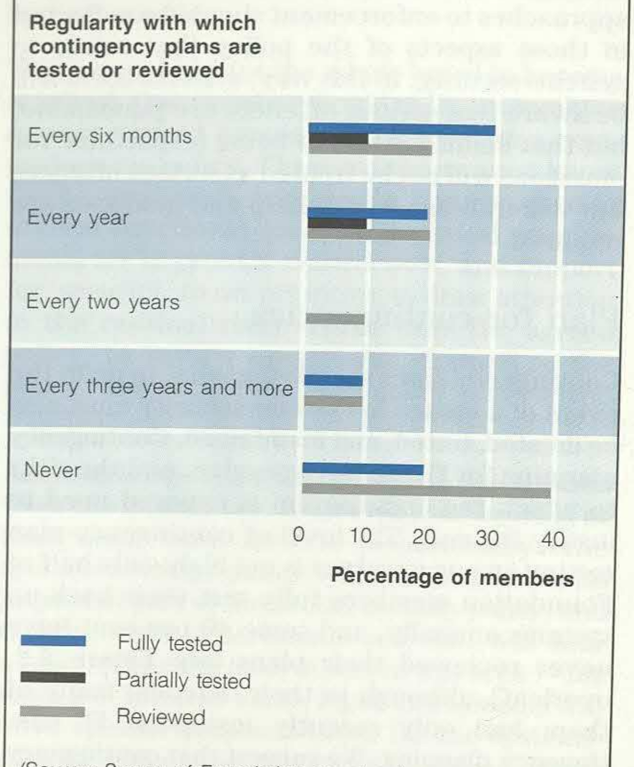
Contingency plans specifying what to do in the event of a breach in systems security must also be created, tested, and maintained. Contingency planning for IT can be expensive, and the risks to which the organisation is exposed need to justify the cost. The level of contingency-plan testing among members is not high: only half of Foundation members fully test their back-up systems annually, and some 40 per cent have never reviewed their plans (see Figure 2.2, overleaf), although in their defence, many of them had only recently instituted IT contingency planning. We suggest that contingency plans should be exercised every year, as a minimum, and reviewed every two years.

As part of the contingency-planning process, the procedures to be followed in the event of security-related incidents such as virus infection, the accessing of networks by hackers, and suspicion of fraud should be specified. Following problems caused by hackers and virus attacks, Internet in the United States has now formed a Computer Emergency Response Team with a 24-hour hotline for users to report problems. Network users must know who to contact if such an event should occur, and what to do if a breach of security is discovered immediately prior to a weekend or a public holiday.

Budget for security

Security costs money, time, and effort. Our research indicates that expenditure on systems security ranges from 1 to 2 per cent of total systems expenditure in manufacturing and retail organisations to 12 per cent in some

Figure 2.2 Only half of Foundation members fully test their contingency plans annually, and some 40 per cent have never reviewed them



(Source: Survey of Foundation members)

banks. These costs must be weighed carefully against the associated risks. Where departments or business groups are required to implement security measures, they must either fund them themselves, or agree to carry out the measures in return for funding from the centre.

The cost of implementing security measures need not be onerous, however. One Foundation member we spoke to had a security review conducted by consultants, who found that document security was weak. Apart from the cost of the security study, the only further cost was a few shredding machines that were funded out of day-to-day expenditure.

Nevertheless, some measures are quite costly. The cost of installing and implementing a comprehensive access-control package for a mainframe can run into several man-years of effort. Although the purchase and licensing costs are moderate, they are dwarfed by the costs of the effort required to set the package up to achieve good security in a specific installation.

Review the policy regularly

The security policy will be drafted in the context of current social, political, legal, business, and technological factors, and changes in any of these will require changes in operating practices, which, in turn, should initiate changes in the security policy. In this way, the policy can be kept up to date by the security committee on a short-term basis. The policy should also be independently reviewed every four or five years so that board-level concerns and longer-term business and technology changes are not overlooked. There should be no need to repeat the risk analysis in every functional area at each of these reviews, because modifications will have been made, in the interim, in line with operating needs. The risk analysis should be redone only in areas where it is essential at the time of the review.

The procedure for carrying out the initial risk analysis in the systems area is described in Chapter 3. It will provide the basis for specifying appropriate levels of protection for the organisation's systems assets, and for allocating individual responsibilities for aspects of systems security within the organisation.

Chapter 3
Carrying out risk analysis

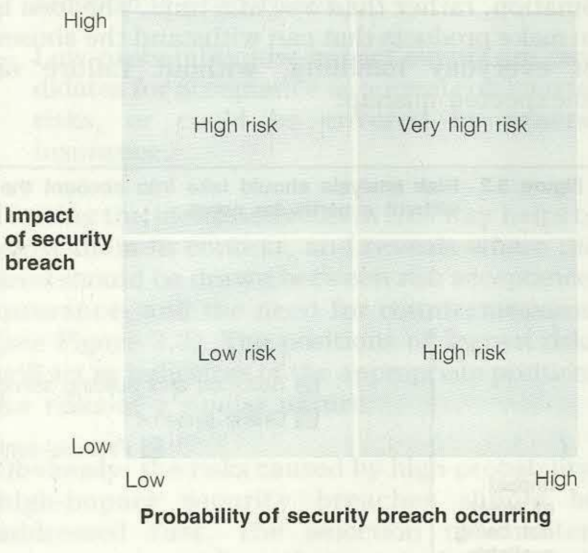
In Chapter 2, we alluded to the use of risk analysis, which the security committee would use as a basis for defining the risks faced by each part of the business, and for comparing those risks to assess their relative severity. In this chapter, we explain how risk analysis works, how it can be applied in the systems area, what its advantages and disadvantages are, and what factors ought to be considered in selecting an appropriate method for a particular set of business circumstances.

Risk analysis works in a similar way to quality control

Risk analysis is used to identify and document an organisation's assets, the threats to which those assets are subjected, and the risks that would result from a breach of security that would occur if the threats are allowed to act upon the assets. The severity of a risk can be estimated by multiplying the impact of a breach of security by the probability of the security breach occurring. The impact will depend on the value of the asset, the business implications of the asset not being available, and the time required to replace or repair the asset. The probability of the security breach occurring will depend on the frequency at which the threat is likely to occur and the vulnerability of the asset. These complex relationships are summarised in Figure 3.1.

The severity of a risk can be reduced by taking countermeasures to reduce either the probability of a security breach occurring or the impact of a security breach, should it occur. For example, the business may depend on a computer system (the asset), which could be threatened by water (the threat) getting into the computer room. A breach of security would occur if the computer room were to become flooded, which would mean that critical business-support systems could not be run until the computer had been repaired or replaced (the impact). Suitable countermeasures would include a drainage system (to minimise the probability of the security breach occurring) and a back-up site (to minimise the impact of the security breach should it occur). The choice of countermeasures will depend on their feasibility, reliability, effectiveness, and cost, the probability of the security breach occurring, and the impact of the breach if it does occur.

Figure 3.1 The severity of a risk is determined by the probability of a security breach occurring and its impact


High
High risk
Very high risk
Impact of security breach
Low risk
High risk
Low
Low
High
Probability of security breach occurring

Security breach = Threat acting upon an asset

Risk = (Impact of security breach) x (Probability of breach occurring)

Impact is a function of value of asset, business implications of asset not being available, and time required to replace or repair asset

Probability is a function of the frequency of the threat and the vulnerability of the asset

The process of risk analysis is very similar to the process of designing for quality. Quality is generally defined as fitness for purpose and has attributes such as value-for-money, product robustness, and reliability. Security, too, must deliver value-for-money, be robust in the face of threat, work when required, and cover all threats. Like quality, security aims to achieve zero defects, so it is worth examining how quality management achieves this goal.

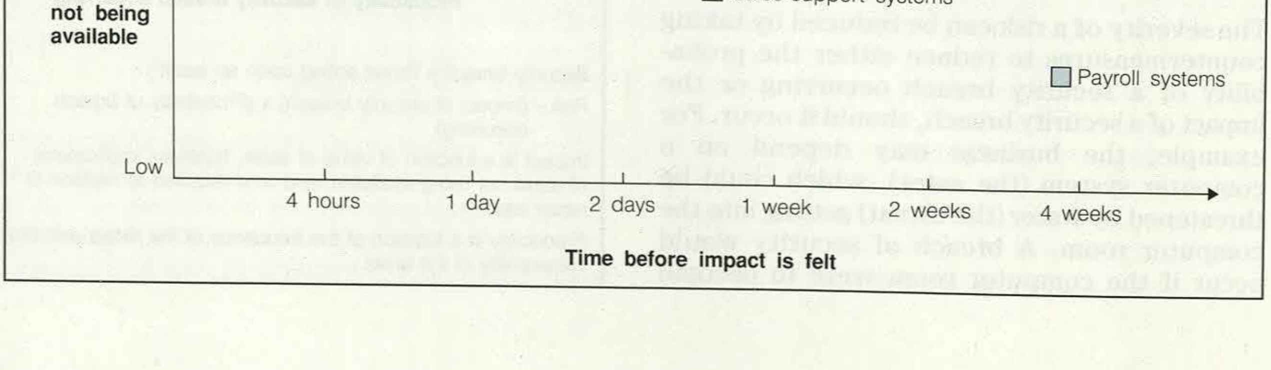
Shigeo Shingo, who taught production engineering at Toyota, and Genichi Taguchi, winner of a prestigious award for his contribution to Japanese industrial standards, are widely respected for their work in zero-defect manufacture and design. Shingo has proposed a checklist approach, designed to eliminate defects by continual product and process checking. The checking procedures are sometimes manual, but are more often embedded in automated procedures and tools. Improvements are fed back to the checking procedures as soon as any residual errors are detected. The aim of Taguchi's work is to achieve highly robust solutions at an affordable cost. In essence, his approach is to examine combinations of solutions to test their effectiveness in combination, rather than one at a time. The idea is to make products that can withstand the abuses of everyday handling, without failure or unexpected mishap.

Risk analysis emulates this approach, by formulating a checklist of the threats, assets, and risks to the business. This checklist can be used when designing information systems and for subsequent audit work. Risk analysis also seeks to identify the most suitable set of countermeasures — that is, those that are most effective over the widest range of threats. The approach can be expected to address the two main aspects of security — the control of threats due to accidents and errors, which account for some 95 per cent of events and over 50 per cent of losses, and the control of threats due to deliberate actions, such as fraud and damage. The means of identification, detection, and elimination may be different, but the central management technique remains the same.

For the purposes of risk analysis, assets are valued either absolutely, in money terms, or on a scale of value, from low to crucial. The impact of any given asset's not being available must take account of the length of time that the business could continue to function without that asset. Figure 3.2 gives some examples and shows that some assets should be given a very high risk analysis value, whatever their monetary value. The risk analysis values should also be consistent with the criticality of the assets to the associated business function.

In the risk analysis process, the threats, the assets they threaten, and the risks they produce

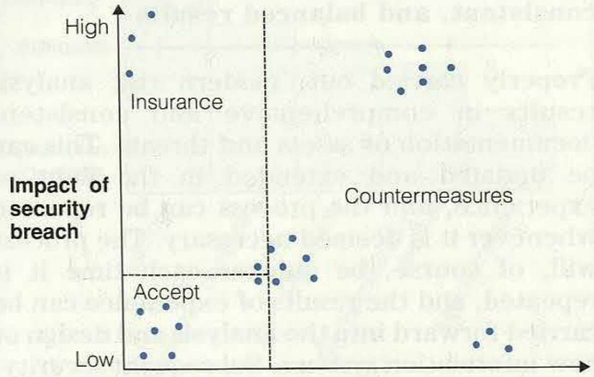
Figure 3.2 Risk analysis should take into account the length of time that the business could continue to operate without a particular asset



are documented. The procedure is repeated for all significant assets and threats. This means that a large amount of information has to be collected and tabulated. To do this efficiently and effectively, both top-down and bottom-up approaches are necessary. For some installations, the smallest valid asset/threat combination might be a minicomputer; for another, it might well be as small as the quality of hinges on an access door. To avoid too much irrelevant detail, a top-down approach, which establishes the framework for analysis, is essential. Tabulation stops when threat/asset combinations start to fall below previously agreed risk-acceptance criteria. Of course, some threats arise from highly specific technical causes, such as tiny flaws in operating systems, small weaknesses in networking, and so on. A 'bottom up' approach will be more appropriate in such instances, and software and telecommunications specialists will need to be involved to identify the threats and vulnerabilities.

Once an analysis of the potential threats, and the assets on which they have an impact has been conducted, the level of risk involved can be assessed. One way of doing this is to plot the risks according to the probability of the security breach occurring and the impact of the breach should it occur (as depicted in Figure 3.3). The size of each impact is assessed, either in monetary or other terms — for example, the importance of the asset to the business, or the effect of its loss on the operations of the company. The probability of a breach in security occurring is more difficult to assess, however, because it depends on the vulnerability of the asset, the frequency at which the threat is likely to occur, and the effectiveness of the measures in place to counter the threat. For example:

Figure 3.3 Plotting risks according to the probability of a breach occurring and the impact of the breach helps to determine where counter-measures or insurance are appropriate, and where the risk can be accepted



— The likelihood of fire is equally low for large and small buildings, and the countermeasures are similar and equally effective. The frequency of the threat does not increase with the value of the asset. Thus, fire is normally a low-probability/high-impact risk, and is usually countered by insurance and measures to protect life and property.

— Where large sums of money are involved, the probability of theft or fraud and the impact of such an event increase with the value of the asset. This represents a high-probability/high-impact situation, and strong countermeasures will be required.

— Security breaches that fall into the high-probability/low-impact area also justify countermeasures, if only to reduce the frequency of occurrence of threats to acceptable proportions.

— Low-probability/low-impact events are candidates for acceptance as normal commercial risks, or could be covered by general insurance.

Plotting the identified risks in this way helps to show them in context, and reveals where the lines should be drawn between risk acceptance, insurance, and the need for countermeasures (see Figure 3.3). The positions of known risks will act as indicators of the appropriate positions for risks of a similar nature.

Obviously, the risks caused by high-probability, high-impact security breaches should be addressed first. The selection of countermeasures depends on three main factors: their effectiveness, their robustness, and their cost. The set of countermeasures chosen should be effective across as wide a range of threats as possible, and not be easily subverted or rendered ineffective through human or other errors. Well designed security procedures are robust, in that they provide a consistent level of protection against the threats, and can cope with a changing environment. In practice, the number of available countermeasures is quite limited, because only a few feasible options exist, and cost plays a major part.

Because one countermeasure, or a set of countermeasures, can address more than one threat and therefore more than one risk, each risk should be re-assessed as new countermeasures are added, to see if further countermeasures are still required. However, the addition of a new countermeasure can also create new risks, and these should be plotted as well to ensure that they fall into the low-probability region. For example, the use of an encrypting device means that there is one more piece of equipment that can fail.

Some organisations (banks, for example) will need to apply the strongest countermeasures for each of the threats, regardless of the cost. In others, however, applying the strongest countermeasures for each threat would be expensive and not very efficient, because many of the countermeasures will apply to several of the threats. In some respects, selecting a set of countermeasures that will provide an *adequate* level of security at a reasonable cost is rather like selecting the optimum mix of the mechanical parameters for a car's suspension system — stiffness of springs and shock absorbers, type of tyres, tyre pressures, and so on. The suspension system has to be designed to cope with a wide range of conditions and drivers, and is likely to be subjected to considerable abuse throughout its life.

In this type of situation, each component can have a wide range of characteristics, and there are hundreds of different possible combinations. Computer simulations take the designer only so far; the final, optimum combination can be found only by carrying out physical tests. To try out each possible combination in order to find the optimum is much too time-consuming and expensive. Taguchi has a method for reducing the number of combinations that need to be considered in order to find the optimum combination (details of the method can be found in the proceedings of *ISATA 88*, which is listed in the bibliography). We believe that the principles of this method can be applied to determine the optimum set of countermeasures against threats to systems security. The method is relevant because the characteristics of security countermeasures are not easy to quantify for computer analysis.

When an appropriate set of countermeasures has been identified and applied, and insurance cover has been taken out against the most serious potential losses, some residual risks will remain. These should be documented, and the security committee must review the residual risks on a regular basis.

Modern risk analysis has advantages over earlier methods but requires special skills

Early attempts to analyse risks in the systems area centred on losses that might be caused by physical threats such as fire, water, and so on. The objective of these methods was to quantify the probable expected loss, known as the annual loss expectancy. This information could then be used to evaluate the effectiveness of protective measures, and to determine the level of insurance required. The limitations of such an approach are now recognised. The trend in Europe, in particular, is away from purely numerical approaches towards either a purely qualitative approach, or a combined quantitative and qualitative approach, based on easily measurable elements coupled with features that can only be estimated or that have a subjective element. Such an approach does, however, require different skills from those required for the earlier numerical approaches.

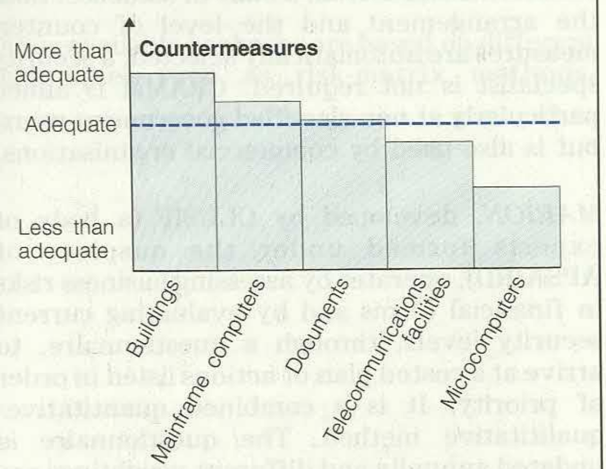
Risk analysis provides comprehensive, consistent, and balanced results

Properly carried out, modern risk analysis results in comprehensive and consistent documentation of assets and threats. This can be updated and extended in the light of experience, and the process can be repeated whenever it is deemed necessary. The process will, of course, be quicker each time it is repeated, and the results of experience can be carried forward into the analysis and design of new information systems. Subsequent security-audit work will be simplified too, because the documentation can readily be reviewed. This will reduce reliance on the skill and judgement of security auditors, and enable the task to be delegated to local user groups.

Use of modern risk analysis will ensure consistency throughout the organisation because all facets of the security problem — financial, physical, technical, and commercial — will be addressed in a similar way, by suitably skilled people. Threats and the impact resulting from these threats acting on the organisation's assets will be allocated weightings of importance on the same basis, so that sensible judgements can then be made about security priorities. Plotting the risks in the way described earlier makes it easier to take a broad and comprehensive view.

Risk analysis ensures that an appropriate balance is struck between the completeness of cover, the acceptance of residual risks, and value for money. Figure 3.4 illustrates the kind of mismatch between risks and countermeasures that might be brought to light as part of the risk analysis process. In the figure, building and mainframe computer security are more than adequate, while telecommunications and microcomputer security are inadequate. The reason for this imbalance is not to do with cost: the protection measures given to mainframe computer systems may be very expensive, while improving microcomputer security may cost very little by comparison. A more likely cause is the lack of a rational and balanced risk analysis.

Figure 3.4 Risk analysis will reveal where countermeasures are less than adequate



Risk analysis requires extensive and sometimes specialist resources

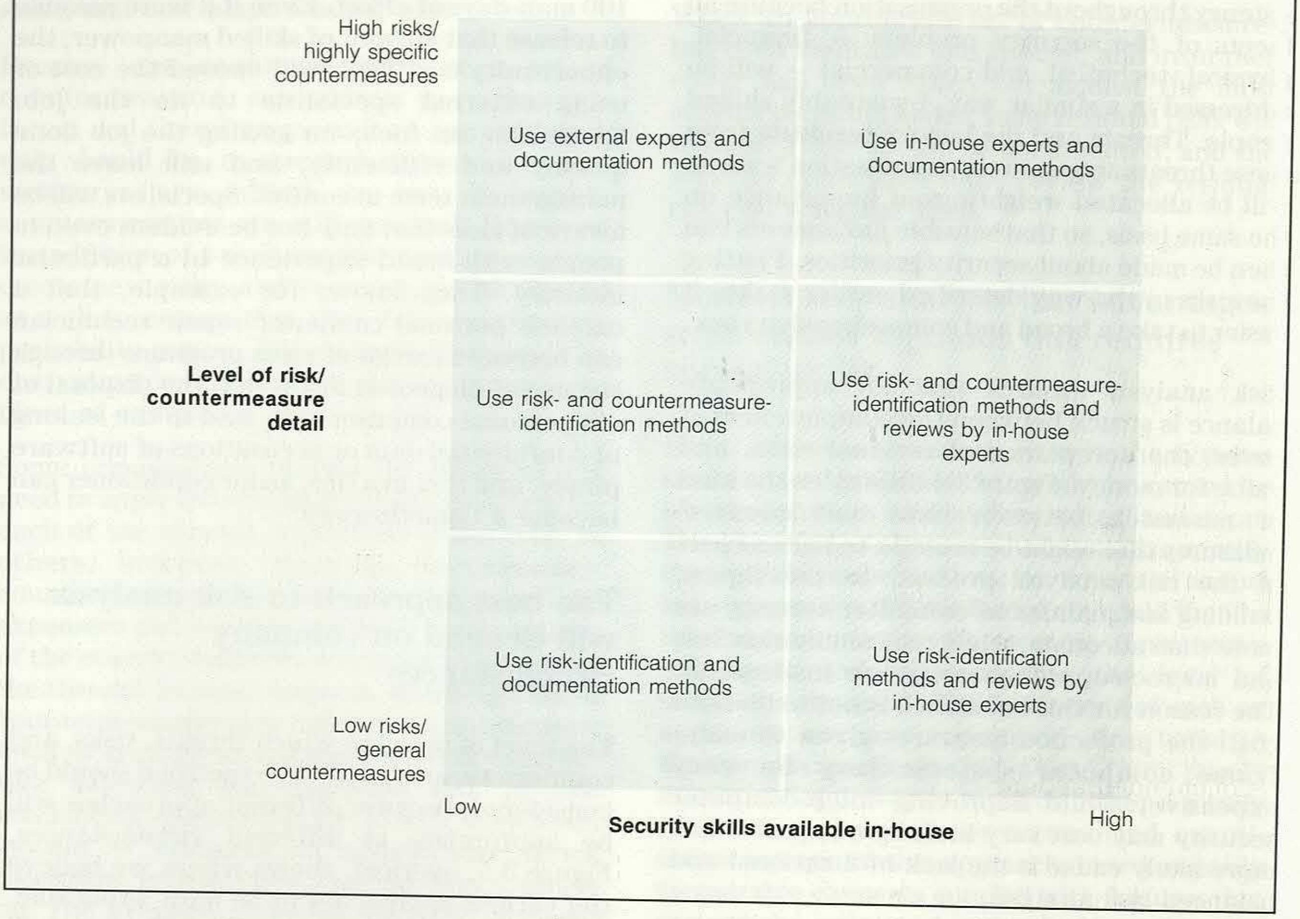
A large risk analysis study can take upwards of 100 man-days of effort. Even if it were possible to release that amount of skilled manpower, the opportunity cost may well exceed the cost of using external specialists to do the job. Specialists can focus on getting the job done quickly and efficiently, and still leave the management team in control. Specialists will be aware of risks that may not be evident even to people with broad experience in a particular industry. They know, for example, that a careless personal computer repair technician can become a carrier of virus programs through the use of diagnostic discs, that the disposal of old personal computers can lead to the leaking of confidential data or accusations of software piracy, and that in a fire, an air conditioner can become a flamethrower.

The best approach to risk analysis will depend on company circumstances

The level of detail to which threats, risks, and countermeasures need to be specified should be considered, because different approaches will be appropriate in different circumstances. Figure 3.5, overleaf, shows where we believe the various approaches to be most applicable. Their appropriateness in various situations will depend on what level of security skills is available in-house, and on how specific the identification of risks and countermeasures needs to be. At the lowest level are the organisations that face few serious threats and that require only general guidance about countermeasures. Their needs are likely to be satisfied by any of the available methods. At the middle level, organisations face some threats and need definite solutions. At the highest level are the organisations that face the most serious threats, like terrorism, high-value fraud, or disclosure risks. They require a highly detailed analysis and highly specific solutions, and will be obliged to use security experts, either from within or outside the organisation.

Figure 3.5 makes it clear that there is no single approach to risk analysis that will always be appropriate. During our research, we did,

Figure 3.5 The appropriateness of different approaches to risk analysis will depend on what security skills are available and on how specific the risks and the associated countermeasures are



Level of risk/ countermeasure detail	Low — Security skills available in-house	High
High risks/ highly specific countermeasures	Use external experts and documentation methods	Use in-house experts and documentation methods
	Use risk- and countermeasure-identification methods	Use risk- and countermeasure-identification methods and reviews by in-house experts
Low risks/ general countermeasures	Use risk-identification and documentation methods	Use risk-identification methods and reviews by in-house experts

however, identify four commercially available risk analysis software packages that can be used in particular circumstances:

CRAMM, which is an acronym for CCTA Risk Analysis and Management Methodology, was developed originally by the CCTA (the UK Government's computing and telecommunications agency), and is marketed by BIS Applied Systems, a UK systems consulting and training organisation. It works in three stages. All physical and data assets are valued on a scale of 1 to 10 for each of four events: disclosure of information, modification of information, destruction of the data or asset, and non-availability of the data or asset. Monetary values are normalised to fit the 1 to 10 scale. Items above a baseline value (usually 3) are analysed. Threats that may affect assets are assessed on a scale of 1 to 5, as are identified vulnerabilities. The results of these two evaluation steps are used by CRAMM's software to select suitable countermeasures from a built-in database. Both the arrangement and the level of countermeasures are automatically selected; a security specialist is not required. CRAMM is aimed particularly at non-classified government users, but is also used by commercial organisations.

MARION, developed by CLUSIF (a body of experts formed under the auspices of APSAIRD), operates by assessing business risks in financial terms and by evaluating current security levels, through a questionnaire, to arrive at a costed plan of actions listed in order of priority. It is a combined quantitative/ qualititative method. The questionnaire is updated annually and different weightings are applied to individual questions based on statistical analysis of APSAIRD'S incidents database. Cost information is also held within MARION's database and is used to compare current expenditures with industry norms. MARION is supported by software developed by PSI, a French software house. An English language version has been developed in conjunction with the UK office of Coopers & Lybrand Deloitte, a large accounting and consulting company. MARION has been used most in France, although it has been used in several other European countries, including Belgium, Italy, Switzerland, and the United Kingdom.

Riskpac, marketed in Europe by Computer Security Ltd, a UK specialist computer security services company, is based on a series of predetermined questionnaires and scoring mechanisms. A variety of questionnaires are available and apply to most security requirements including sector-specific areas, such as banking, insurance, and manufacturing, as well as subject areas such as personal computers, physical security, and communications. Riskpac is a qualitative method that scales risks from 'nominal' to 'catastrophic', on a scale of 1 to 5. An internal 'expert system' technique is used to evaluate questionnaire results and produce risk summaries, standards, risk profiles, and recommended actions.

SIVOR, an MS-DOS-compatible product, is offered by Siemens. This product is currently available in German and addresses physical security, and data and communications security, and allocates priorities to the identified risks. SIVOR holds an internal catalogue of some 792 questions to cover the majority of risk cases.

Other available packages are based on different techniques such as risk-matrix methods, Bayesian statistics (fuzzy logic), Delphi techniques, and incident databases. Each technique has its supporters, but none is yet in widespread use.

One company, Electricité de France/Gaz de France (EDF/GDF), has developed its own risk analysis method over several years. This specifically addresses issues faced by an electricity generating and supply utility and is built as an expert system. EDF/GDF has built its method into a personal computer system that is available to local systems managers for their own use, thereby ensuring that all risk planning is coherent. The package includes a risk simulator that recommends countermeasures through the construction of a 'menace dictionary' that is applied to a vulnerability analysis developed as part of the risk analysis. In this way, potential disasters can be predicted from the known 'menaces' and vulnerabilities. Each systems manager is required to analyse his own risks, to prepare a security plan, and to show how security addresses software development, data management, and the management of physical assets.

As EDF/GDF has acknowledged in the design of its own method, risk analysis will serve to draw attention to those areas within an organisation that are potentially vulnerable to breaches of security, and perhaps, indicate appropriate measures that might be put in place to counter them. It cannot, however, specify precisely where the risks lie or how particular threats might arise. This remains the responsibility of the systems manager. Chapter 4 provides guidance for systems managers on where the problems are most likely to appear.

Chapter 4

Identifying possible threats

While risk analysis can identify the kinds of threats to look for and the areas in which these threats might have the greatest impact on the business, it cannot identify specific threats. Of course, the aim of any security policy is to prevent the occurrence of problems in the first place, but potential problems rarely identify themselves openly. It is essential, therefore, to know what to look for, where to look, and how to seek out the sources of problems.

Threats derive from three main sources — accidents, errors, and deliberate actions. Natural disasters and accidents that have a catastrophic effect should be covered by contingency plans, and are outside the scope of this report. For the most part, however, accidents cause damage, inconvenience, and expense, but do not generally make newspaper headlines. The examples in Figure 4.1 illustrate the point. The emphasis in systems security should therefore be on identifying threats that arise from errors or deliberate actions.

Errors account for well over half of all events causing losses

By far the most common cause of problems is errors. The statistics from APSAIRD, which formed the basis for Figure 1.1, showed that errors produced some 21,000 events in France in 1988, and accounted for 24 per cent of the total losses attributable to breaches of systems security. These numbers are probably fairly typical for other European countries.

In the main, errors are 'people' problems. A certain amount of damage is caused by people making 'positive' mistakes — for example, entering information incorrectly, or deleting information inadvertently. A much greater source of potential errors, however, is human inertia — leaving in place inadequate measures, which themselves create potential risks, or failing to take account of changes in the business environment, and thus leaving the organisation vulnerable to new types of threats that could quite easily be guarded against.

Figure 4.1 Accidents are a major nuisance but their effects are not often devastating

An electrician made a minor wiring error when repairing an air conditioner. When the air conditioner was turned on, the fault caused an excessive supply voltage to some 20 terminals and PCs as well as a minicomputer. Systems were down for some 48 hours while repairs were made. The basic cause of the problem was a poorly designed power feed to both air conditioner and computer suite.

A member of staff tripped over a trailing lead supplying a local area network server. The lead was pulled out of its plug and the earth wire contacted the live power connection. The resulting power surge destroyed part of the server hardware.

Staff working on the power feed to an office floor failed to turn off the supply. The accident not only disrupted power to this office area, but destroyed a considerable number of communication ports, both on terminals and computer equipment.

Other problems have been caused by the mismanagement of access rights to computer systems. In a case we heard of, a user was able to delete not only his own files but also those of other people, without being aware of what he was doing. Examination of minicomputer file structures often reveals potential problems of this sort — accidents waiting to happen.

Inadequate measures are a source of potential risk

Measures that are simply inadequate or that do not work when required are a source of potential threat; they can be identified only by regular testing and review. For example, extensive fire precautions are usually applied to computer installations, and the complexity of such precautions may lead to problems:

- At one site, part of the fire-detection and water-sensing system had been disabled because it would randomly indicate a spurious problem and unnecessarily shut down the computer installation.
- At another site, the fire-suppressant control system was ineffectively linked to the computer room and, when a fault arose, failed to shut down the computer even though the air conditioning had been cut off, resulting in an overheating problem.
- In another instance, a computer room was destroyed, possibly because a smoke detector was not set to cut the computer-room power supply.

There have been several cases of water sprinklers associated with fire-prevention systems being triggered when there was no fire, flooding computer rooms and causing major damage. Water, in general, is frequently overlooked as a serious hazard. Burst pipes and overflowing washbasins have also caused serious damage to computer rooms and telecommunication facilities, and buildings have even collapsed owing to water trapped in an upper floor. A recent inspection of a computer room revealed water collecting under the floor. An air conditioning unit was leaking and its water sensor was not working properly.

These examples illustrate that the very act of attempting to counter threats can, if inadequately thought out or managed, result in damage. Only by continually being on the alert for these problems can management hope to avoid them. For example, if terrorism is considered to be a serious threat, the back-up site should be at least as secure as the primary site.

Other sources of threat occur when normal defences are weakened. For major computer installations, especially in the financial-services industry, the recovery situation itself represents a potential security threat. Recovery procedures can weaken the normally strong software protection, and this means that security could be compromised, even if only for a short time.

Failure to take account of change will leave the organisation vulnerable

Risks also arise because systems management fails to update existing countermeasures in the light of changed circumstances. Sometimes, this occurs because the changes happen very slowly and are not noticed. New threats can also arise because of technological change or new applications, or because parts of the business become exposed as a result of altered relationships with other businesses.

Gradual change outdates existing measures

The gradual obsolescence of equipment can create unnoticed threats. Should a disaster, such as a fire or a major breakdown, occur, recovery could be slow, or impossible if replacement equipment is no longer available. In such a case, the disaster-recovery procedures will involve upgrading to a new generation of equipment, and it is neither easy nor quick to do this.

The gradual loss of trained staff can also jeopardise emergency planning without anyone necessarily being aware that this is happening. One Foundation member tested his contingency plan and discovered that some 40 per cent of his staff had changed since the previous test had been conducted. Not enough people understood the contingency procedures, and the test failed. The solution was to test the contingency plan more frequently — every six months, in this case.

Technological change creates a need for new countermeasures

Technological advances, such as the growth of networking and departmental computing, and the extensive use of personal computers throughout most organisations, create a need for new measures to counter any added threats to their security.

Networking: Networking can have an impact on all three security-related properties of any organisation's systems assets — confidentiality, integrity, and availability. The most common problem is probably availability. The telecommunications links between sites can never be completely reliable and line failures are not uncommon. Organisations frequently seek to provide duplicate links and back-up mechanisms to compensate for this problem, but these measures can bring about their own problems. When selecting diverse routeing from network suppliers, it is wise to ensure that the cable routes are physically separate — at least the reach of a backhoe digger apart, for example. They are then most unlikely to be dug up by the same digger. Some companies we visited frequently use dial-up circuits when their leased-line links fail, but do not use modem dial-back or other forms of protection against hackers. The risk from failing to provide such basic protection is small, but it does exist.

One of the most worrying security aspects of networking is linking into other networks, because it is not usually practical to control security in other networks. It is therefore vital that an organisation controls access into its own network. One company that was obliged to connect to an international public network for business reasons regularly had hackers attempting to access its systems. Failures of password security by users of the public network service led to occasional unauthorised accessing of business accounts. Only stronger password measures in its own system and the use of authenticator devices (which are described in Chapter 5) prevented further damage.

Departmental computing: Departmental computers are designed to work within the general office environment, where unreliable power supplies can cause problems. It is not unusual to find that office cleaners remove minicomputer power plugs in order to run a floor cleaner, or that when a new photocopier is installed, it interferes with the departmental minicomputer. These are simple problems that are generally well understood.

Departmental computers are subject to other threats, however, mainly due to lack of awareness among users. Frequently, the standard initial field service and installation passwords built in to the system when it is supplied are not removed; users of departmental computers are rarely aware that they even exist. There are well documented cases of unauthorised people using these passwords to gain access to departmental computers.

The need for back-up data is sometimes forgotten, and users sometimes forget to make back-ups regularly, or to document them properly. Back-up data is often stored alongside the departmental computer. To lose a computer in a fire or a flood is a problem, but to lose both computer and data could be a catastrophe. None of these errors is particularly difficult to understand or to manage, so long as someone takes responsibility for seeing that they are attended to.

Personal computers: Apart from viruses, personal computers are subject to problems ranging from simple theft and the use of pirated software, to problems with the security of passwords and the security of data held on hard discs.

The very usefulness and portability of personal computers leaves them open to theft. People may walk in from outside and simply carry away unattended machines, or employees may carry out smaller, modern ones in a plastic carrier bag. The temptation for users to make unauthorised copies of software also leads to potential prosecution for software pirating. Recent prosecutions in Europe by the Federation Against Software Theft (FAST) have highlighted this problem.

Password security is not simply a matter of users ensuring that they prevent others seeing the passwords they are using. Clever hackers can retrieve passwords from the innards of a personal computer without the user being aware that his password has, in effect, been stolen. In the past, only programmers could retrieve passwords in this way because it involved a small program designed to copy the password into a file for later collection and use. The advent of personal computer 'keystroke filer' programs has also opened this route to some personal computer users. Some users simply put their passwords into a keystroke file and use it to log on to systems automatically. The dangers of this are obvious; anyone who can gain access to the user's personal computer can use the automatic log-on feature or copy his passwords. A more insidious use of the keystroke filer is to set it running on a personal computer without telling its user. All keystrokes, including passwords, can be saved to a disc file for later collection.

Data security can also be compromised because many personal computer users do not realise that using a 'delete' command does not necessarily physically remove data from a disc. The normal delete command within MS-DOS, for example, does not actually erase a data file but merely removes the directory's reference to it. Recovery programs work by rebuilding this linkage. Similarly, the normal hard disc format command does not actually delete data in the way a floppy disc format command does. Users have sold personal computers as surplus, believing that their data has been 'deleted', only to find that the data could be recovered. In fact, methods are available to erase files comprehensively so that even the most skilled software hacker cannot reconstruct them. Where extreme security is required, it might be better to destroy redundant floppy discs by cutting and burning them, and to destroy old hard disc units with a hammer.

New kinds of applications expose an organisation to new threats

New kinds of applications may be risky in two respects. First, the introduction of any new application may bring with it unknown or unforeseen threats. Second, there is an inherent risk in introducing computer systems into working cultures where people are unaware of, or unwilling to accommodate, the potential vulnerabilities of such systems. Two examples serve to illustrate the point.

The use of computer systems in conjunction with telephone-based marketing and sales campaigns is increasingly common in Europe. Many companies are using telemarketing techniques to sell high-value financial services — mortgages, insurance, and savings products — and to run major customer-contact services — holiday bookings, assessing responses to television advertising, and so on. These activities create two kinds of threats — the threat of non-availability, and the threat of theft of customer details. If either the telephone equipment or the computer fails, customer service will be degraded. If customer lists are stolen, a valuable asset is lost, and the confidentiality of customer contact is no longer guaranteed. Computerised customer lists stolen from a holiday company were used to identify houses that could be burgled while their owners were away.

The second example concerns the growing use of knowledge-based systems in commerce and industry. While a knowledge base may represent a valuable asset because of the information or competitive advantage it can give, it will also be important to ensure that it is not accidentally or deliberately corrupted. To lose the confidentiality or availability of the knowledge base is bad enough; to be making wrong decisions on the basis of corrupted information is even worse.

Increased dependence on others may jeopardise security

Businesses are becoming increasingly enmeshed in networks for service provision and data interchange. It will gradually become impossible to do business without these networks, and their reliability is therefore increasingly important. Systems managers must be aware of the threats that arise from being dependent on other organisations, whether these be electricity-supply utilities, public telecommunications operators (usually a PTT), facilities-management companies, managed data network companies, electronic data interchange partners, and so on.

Electricity-supply utilities: The reliability of electricity-supply utilities is normally very high. Each organisation must assess how long it could survive without its computer systems during a break in the electricity supply, and decide whether it is justifiable to install on-site emergency generators. The decision will depend on a variety of factors ranging from the time criticality of the business application to the likelihood of a prolonged industrial dispute occurring.

PTTs: Security measures instituted by PTTs against fire and other major damage are normally very good, but disasters can happen. For example, in Hinsdale, Chicago, in 1988, the voice-switching centre was burnt out and users were without service for several weeks. A similar event took place in Lyon-Sevigne in 1981, when a fire closed the switching centre for a week. At Hinsdale, the fire was so severe that it damaged the cables entering the exchange. This hampered recovery efforts because tens of thousands of wires and optical fibres had to be reconnected, and equipment had to be replaced. Not only were local services affected, but links passing through Hinsdale were wiped out.

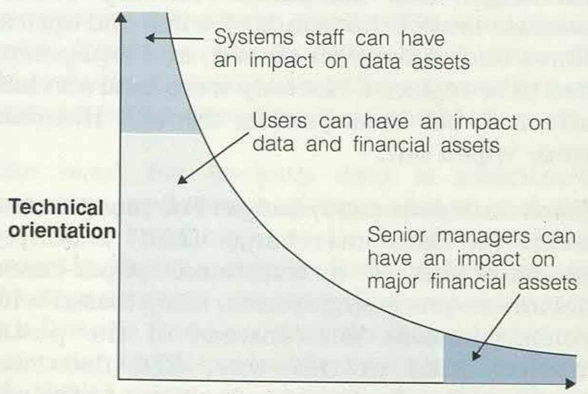
Electronic data interchange: For many users, electronic data interchange (EDI) is simply an extension of a traditional paper-based document-processing system, using leased telecommunications lines instead of the postal service. Used in this way, EDI does not represent a major threat to security. As the use of EDI grows and public data networks become involved, however, security-related concerns will increase. The automation of document-processing systems does create new problems. There have, for example, been reports of automatic re-ordering systems restocking warehouses when the intention was to run down the stocks. A wise precaution is to build checks into automatic-ordering systems to ensure that orders outside the regular pattern are referred to a supervisor. The speed and efficiency of EDI systems can turn a simple error into a potential disaster.

Deliberate actions account for a small but costly proportion of incidents

Although losses due to deliberate actions account for only about 4 per cent of incidents, according to APSAIRD, they account for some 50 per cent of financial losses. The difficulty of dealing with these incidents is that the majority of those who abuse computer systems are insiders who already have access to the systems as part of their legitimate activities. Estimates suggest that between 70 and 80 per cent of computer abuse comes from sources inside organisations.

Figure 4.2 shows that the kinds of abuse to be expected vary according to the different kinds of user. The figure shows that business and computer experts have a limited range of opportunities to abuse systems, but to potentially very significant effect. It also shows that, apart from hacking, the more common forms of abuse are committed by those who use systems every day as part of their normal work, and get to know their weaknesses. These people are thus in a position to commit fraud or acts of sabotage, or to misuse computer systems in a malicious way. Systems may, of course, also be misused by those outside the company, if they can gain access.

Figure 4.2 Skills and scope for deliberate action are related



Gaining access

Organisations that operate closed computer systems with no dial-in links or other connections to the public telephone or data networks, or other computer systems, run very little risk of being attacked by outsiders. The opportunity to run closed systems is, however, diminishing, as more and more companies are compelled to link into EDI networks, access databases, or run other networked applications.

There are two main approaches that an outsider wishing to attack a system might take. The first is to exploit the weaknesses of the network to gain access to the system and then to try out known or likely password combinations. The second is to eavesdrop to identify a password or other useful information and to use this to gain access via the network. Once access has been gained by either approach, it is then possible to plant viruses or Trojan horse programs, to corrupt files, to read confidential information, and so on. (The principles of a Trojan horse program are described in Figure 4.3.)

Gaining access via networks

Networks increase the number of people who can gain access to computer systems and extend the geographic area from which they can gain access. They carry information, including user identities, as sequences of electrical impulses.

Figure 4.3 A Trojan horse program is used as a means of inserting a virus into a computer system

A Trojan horse program is simply one that carries out an additional (usually undesirable) task in addition to its described purpose. For example, the recent 'Aids' virus scare involved a disc that contained a program that offered advice on the disease Aids. In addition to giving advice on Aids, it also inserted a virus program into the personal computer running the program. In this case, the Aids advice program served as a Trojan horse to carry the virus program into the personal computer.

All an intruder has to do is to reproduce these impulse sequences exactly and the network and the computer will be deceived. Networks fall into two main categories — local area networks and wide-area networks. Both are subject to abuse.

Local area networks (LANs) are now an everyday part of business life. They normally work by sharing a common communication mechanism among many users. They are frequently used to provide data services throughout large buildings, and when linked together by wide-area networks, across complete businesses. Essentially, all information, however confidential, is potentially available to every device connected to the LAN. As LANs grow, they gradually accumulate users and departments all over the business, and they become connected to mainframes and to the outside world.

LANs are often treated as an office automation facility and their security is seen as a local concern. The wider security implications should not be forgotten, however. The operation of a LAN, for example, is often monitored by a specialist data monitor, whose function is to detect faults on the network. This device must, by definition, have access to all the data being transmitted and to all the data stored on devices connected to the LAN. Someone with appropriate technical skill can use a data monitor to capture all the information, including passwords. Obviously, the technical skill required restricts the opportunity for abuse, but some user terminals can be placed in 'promiscuous mode', which means that they can capture and read all data on their section of the LAN.

LANs are also subject to all the normal problems of password administration and password abuse, particularly where any terminal on the LAN may be designated as the administration terminal. Clever users can also access print queues and other temporary storage areas to read data if the LAN operating system itself does not incorporate strong security procedures.

Other problems are related to the ease with which some of the equipment can be accessed. For instance, many LANs have one or more service providers (servers) connected to them. These act as file stores, communication gateways, or print servers. Although some LAN operating systems are reasonably secure, server hardware is frequently based on a conventional high-power personal computer without a keyboard. All that is required to access any file stored on the server is a copy of MS-DOS, and possibly a keyboard. Use of file encryption or special chipsets can help to limit this problem.

An organisation is also vulnerable to failure of either a server's software or hardware, because servers support the activities of all users connected to that part of the LAN. Communications servers will often not automatically restart after a power failure, and file servers frequently store data in a 'cache' memory before it is written to disc. Data stored in a cache memory can be lost during a temporary power failure, thus compromising the integrity of data files.

Wide-area networks (WANs) are normally operated over lines provided by public telecommunications operators (PTTs). PTTs do not guarantee the confidentiality of the information they carry. In practice, this is not a major concern for commercial users, for most of whom the security offered by normal PTT lines is adequate. Where confidentiality is critical, encryption techniques may be used. More important is the integrity and proper delivery of information. Data must be protected from loss, errors, or alteration.

One common means of ensuring the integrity of data transmitted over a network is to add a message-authentication code to the data, a practice that has been adopted for many years in the banking industry. It is necessary, however, to ensure that all relevant components of the message are covered by the code. Recently, in an attempted funds-transfer fraud, certain important parts of the message were altered without the message-authentication code being disturbed. Where extreme security is required, it may be necessary to conceal not only the contents of messages, but also the fact that messages exist. An example might be information leading up to the announcement of a take-over bid, a change in national interest rates, or a currency revaluation.

Access control and user authentication are other features of wide-area network security that need to be considered. Two members reported problems from modems that had been forgotten about or had been attached to the system by non-IT staff. Several members expressed concern about unknown connections, a problem that is more likely to occur in a large and mature network. One company reported that it had experienced problems from a modem hidden under the computer-room floor. It had been used by staff for unauthorised access from home.

Another weakness of networks is their switching equipment. Equipment such as multiplexors, PABXs, and packet switches often have special ports that are used for remote maintenance, and are frequently delivered with a standard password. It should be changed as soon as possible to avoid any risk of unauthorised access via these maintenance ports.

As computer systems have become more difficult for hackers to access, their attention has turned to other IT items:

- For many years now, telephone toll fraud has been a problem in the United States. In Europe, the limited switching facilities allowed by the PTTs, and by law, have made this less likely, but it is sometimes possible to dial into a corporate communications system from the public network and back out again, sometimes into another country. Strictly speaking, this offends PTT regulations in some countries — but it does happen. This facility allows corporate users to save money or to make calls more quickly, and outsiders who know about this facility can use it to make cheap calls at the company's expense (Figure 4.4 explains how this happened to one company in New York).

Figure 4.4 Insecure computer systems can be subverted for telephone-toll frauds

One company's corporate network allowed local users in New York to dial in to the central computer system (at local call rates) and to dial out again to another country. The aim was to allow corporate users to call offices in another country at local-call rates. However, once access was gained to a particular country, it was then possible to dial into that country's national telephone network and call any telephone number in the country. The existence of this facility was discovered by Puerto Ricans, who were selling the appropriate dialling sequences to their compatriots on the streets of New York. They were able to call home at local-call rates, with the rest of the call charges being billed to the company.

- Facsimile is one of the fastest growing communications media and is overtaking telex in many parts of the world. There are some concerns regarding the security of facsimile messages, however. These include the lack of confirmation of delivery, the logging of messages, and the nuisance of junk messages. The use of personal computers as facsimile machines and network 'fax servers' simplifies the task of introducing facsimile transmission into an office automation environment. It does, however, have its risks. With some facsimile packages, it is possible to cut and paste images, and even to store signatures.
- Some air conditioning units and building-management systems can now be accessed via modems for remote adjustment and maintenance purposes. Password protection ensures that only authorised users can adjust these systems, but (as we demonstrate in Chapter 5) password systems are not foolproof.

Eavesdropping

Wiretapping and the monitoring of radiation emitted from terminals are feasible but unlikely threats, and few of our interviewees took definite measures against them. We believe that this attitude is currently justified for all but the most highly sensitive or valuable data.

The very act of wiretapping exposes the attacker to the risk of being caught in the act, because he must gain access to the physical circuit. The most likely location for this is in or near the building housing the facilities under threat. Untidy and poorly documented cabling schemes make such attacks easier because unauthorised connections and apparatus will be harder to spot.

Eavesdropping on the radiation emitted from terminals and personal computers is possible, but is not as easy to carry out as early publicity implied. Cheap and portable equipment has not been able to capture useful data at more than short distances.

Sophisticated equipment can, however, pick up the radiation from screens and other data sources at significant distances. Modern screens and personal computers emit less radiation than earlier models, although some home computers still give off significant radio signals. Other sources of radio emission are local area networks, RS232 cables, and optical-fibre driver circuitry. Even optical-fibre cables can be tapped if access to the fibre can be gained. These techniques border on those used for gathering military intelligence, however, and are fairly expensive and difficult to carry out.

Fraud

In most cases of computer-related fraud, the computer is used simply as a tool. The computer systems usually work perfectly; it is the control systems surrounding them that fail. A few examples illustrate the point:

- In two companies that we interviewed, experienced clerks entered false invoices and received payment via bank accounts in other names. In one case, the fraud was exposed by an informant. In the other, it was discovered purely by accident, when the payment went astray and was followed up by another clerk. In both cases, the staff concerned were dismissed and only one prosecution followed.
- A company found that its accountant had defrauded the company of a sum of money in excess of $100,000 over a period of two years. Investigation revealed that he had carried out a similar fraud with his previous employer, and had been dismissed. As no references had been taken up, the opportunity to detect his previous crime was missed. The accountant was prosecuted on this occasion, and served a prison sentence.
- Other fraudulent activities have involved the theft of equipment, the fraudulent refunding of money to customers, and the fraudulent issuing of money-transfer instructions.

Sabotage

Sabotage has traditionally been associated with terrorism, and sometimes, with industrial action. Sabotage usually involves physical damage but it can also be directed at physically corrupting programs and data. Logic bombs, viruses, and worms fall into this class of threat (each of them is described briefly in Figure 4.5). A saboteur can also attack systems by overloading them with wasteful processing jobs that prevent legitimate users from running their systems. This has been done by using rogue programs that initiate extremely long and wasteful print runs, or by tying up data networks with illegitimate traffic that keeps access ports busy or moves massive volumes of data around.

The message is clear. The modern, intelligent saboteur has plenty of weapons available but requires computer skill and access in order to

Figure 4.5 The main threats to programs and data are logic bombs, viruses, and worms

Logic bombs are simply programs that damage data or restrict access to it after some triggering event — typically, the passage of time or the deletion of a named employee. Back-ups are not necessarily a useful countermeasure because a logic bomb could well have become incorporated into the back-up cycle. Logic bombs have been used by suppliers to ensure that leasing payments are renewed on time — whether this constitutes sabotage or sensible commercial protection is a matter for debate.

Virus programs are designed to find, and subsequently to attach themselves to, other programs. Virus programs are a particular threat to personal computers. As the virus continues to seek programs to infect, it spreads. To do damage, viruses usually contain some form of delayed-action logic bomb. The delay is important to the virus builder because it gives a virus time to infect other personal computers and thus to extend the 'infection'. Viruses can be a useful sabotage weapon, easy to plant and difficult to trace.

Worm programs operate by copying themselves into networked systems. They differ from viruses in that they can normally exist as standalone programs, even if they attempt to masquerade as other more benign programs. The danger is that worms are designed to exploit security 'wormholes' and are therefore liable to work around the normal security and access-control measures. IBM experienced a worm program in its electronic mail network that sent a Christmas-tree image to every terminal in IBM's network, thus jamming it. On this occasion, the source was an authorised network user; the worm simply got out of control. During November 1988, Robert Morris, a PhD student at Cornell University in the United States, created a very ingenious worm program that exploited a whole series of 'wormholes', including common and not so common passwords, to help propagate his 'worm'. The effect was to bring down Internet and to infect some 6,000 computers in the United States. It is estimated that it cost $100 million in lost time and the costs associated with clearing up the damage done by this worm. Morris was eventually fined $10,000 and ordered to do 400 hours of community work in 1990.

make them work. Worm programs are worthy of study by security specialists in order to identify wormholes and modes of attack. Armed with this knowledge, systems managers can build better defences.

Misuse

Misuse covers all non-malicious uses of computer systems, including use for private purposes, playing games, or running illicit businesses. None of the members we surveyed cited misuse as a major problem. With the availability of personal computers with powerful accounting and database packages and interesting games, there seems little need, or motivation, for users to misuse mainframe systems for these purposes.

We have seen, in this chapter, how varied the sources of potential threats can be, ranging from malicious acts of sabotage, to vulnerabilities created by inertia or a simple lack of awareness. A checklist of the most obvious threats to the various types of computers and networks is given in Figure 4.6. The measures that can be taken to counter such a range of threats are also, therefore, very wide-ranging. In Chapter 5, we suggest how systems managers should select countermeasures that are appropriate to help prevent particular kinds of potential breaches of security.

Figure 4.6 Each type of computer system is subject to a wide range of threats

The most common threats are listed below.

Mainframes and minicomputers
Inadequate back-up procedures
Lack of contingency plans and recovery procedures
Insufficient change-control procedures
Inadequate control of user access
No division of critical duties
Insufficient log-in (and other access) failure reporting
Forgetting to remove standard passwords
Insufficient use of suppliers' systems security expertise
Poor use of access-control software

Personal computers
Lack of back-up procedures
Lack of virus-detection measures
Insecure modems
Unauthorised copying of software
Uncontrolled use of 'shareware'

Local area networks
Inadequate protection of LAN servers
Inadequate protection of modems or personal computers connected to the LAN
Insufficient security in the bridge to a wide-area network
Insecure links to mainframes or minicomputers

Wide-area networks
Inadequate security measures for dial-in services
Insufficient protection for packet-switched services
Forgetting to remove standard passwords from switching equipment
Inadequate monitoring of network traffic
Poor documentation of legitimate access paths
Insecure access to cabinets, connection frames, and riser cables

Chapter 5
Preventing breaches of security

In the previous chapter, we described the many possible sources of security problems. We have shown how varied the threats can be and how important it is to be constantly on the lookout for potential problem areas. In this chapter, we concentrate on the main preventive measures against threats to systems security, other than those deriving from natural perils like fire and flood, which are covered comprehensively in contingency-planning manuals, and from theft and sabotage, which are usually of a physical nature.

As part of our research, we interviewed Steffen Wenery of the West German Chaos Club. The exploits of the Chaos Club members illustrate the strengths and potential weaknesses of computer systems:

- Chaos members claim to be able to come and go as they please within any networked computer system. This is certainly an exaggeration, but Chaos members do have a formidable reputation and great technical skill.
- Chaos members admit that they are unable to defeat strong, very well administered password controls and systems using authenticator techniques. This, at least, is reassuring.
- Chaos members and other hacker groups have access to considerable expertise. Some hackers are known to work with the types of machines and operating systems that they attack, as part of their normal work. Where they run up against a problem, they are able to work out a solution at leisure.

Steffen Wenery told us that the hacker culture is changing. In the early 1980s, most Chaos members were computer experts and enthusiasts — hackers, in the original sense of the word. Now, there are fewer true experts, and many more semi-skilled enthusiasts. A true hacker will regard himself as a guest in a computer system. As a guest, he will not do any damage, and will leave politely if asked to do so. This is doubtless an honourable intention, but Foundation members will most certainly be well advised to avoid having any uninvited guests in their systems.

The activities of Chaos members and other hacker groups are proof of the fact that any networked system is liable to attack by experts, but that, without a place to start — a password, or an obvious weakness in the system — Chaos and other similar groups are powerless.

Make passwords more secure

Passwords have been a security feature in computer systems for many years. A simple password system is fairly easy to design, and requires no special terminals or other hardware. It is therefore cheap, but is not a guarantee of security. Better methods are available, but for technical reasons, and for reasons of cost, passwords remain the principal means of authenticating users.

To make good use of the security available from password systems, systems managers must know and be able to detect the most common methods of defeating passwords and make vigorous efforts to preserve the security of passwords by making them hard to guess and preventing disclosure. The most common methods of defeating password systems are use of inside information, use of standard system passwords, and use of proper names and common words.

Discourage use of inside information

The simplest way for an insider to discover someone's password is simply to watch a colleague key in his password. In the United States, this is known as 'shoulder cruising'. There is very little that can be done to avoid this problem, other than isolating sensitive terminals by partitioning or careful location within an office. Obviously, users should be discouraged from writing down or displaying their passwords. In some companies, including Barclays Bank, one of the large UK clearing banks, divulging a password can be a disciplinary offence. Others, including American Express and Citibank, make the disclosure of a password a matter for dismissal.

Always change standard system passwords

The dangers of leaving standard system passwords unchanged have been amply explained elsewhere, and all members should ensure that this highly dangerous loophole is closed in all systems under their control. In his book, '*The Cuckoo's Egg*', Clifford Stoll describes how standard system passwords were used on several occasions by hackers to penetrate systems. Our own research has revealed a case where this loophole had been left open and a mischievous user had used it to shut down the computer system. No damage was done other than the loss of a few minutes of computer time, and an embarrassed systems administrator.

Avoid the use of proper names and common words

Using proper names and common words as passwords makes it easy for users to remember them, but such passwords are a security problem for several reasons. First, they are easy to guess, particularly when passwords and account names are the same. Second, they are susceptible to automated attack, either by an automated password-guessing program or by a cryptographic method.

Password-guessing programs can be used only if continuous access is available. One of the best ways of defeating this approach is therefore to check for two or three unsuccessful attempts to enter a password and to block access to the offending terminal or communications port for a short time, or to block it altogether until a systems administrator manually restores service. Most secure operating systems have facilities to do this automatically. Even this may not be sufficient if the intended intruder has access to many terminals or ports, because two or three guesses may be made on each one without an alarm being raised. However, sophisticated security packages will identify even this sort of activity. A log-in report highlighting failed password entries should always be sent to the systems administrator, so that such episodes can be investigated. Some organisations have arranged that a password log-in failure is not indicated at the offending terminal. This wastes the attacker's time, and gives the victim time to trace the attacker.

An obvious defence against password-guessing programs is to avoid using commonly used passwords — a partial list is shown in Figure 5.1. The usual advice is to avoid proper names, words with a sexual or obscene connotation, or words drawn from role-playing games, science fiction, and fantasy literature. However, even less commonly used passwords are not immune from attack by password-guessing programs. The Internet worm program (which was referred to in Figure 4.5) had a built-in list of more than 400 possible words, including cantor, ersatz, pizza, and sossina. Figure 5.2 gives a representative sample.

Some operating systems, notably Unix, store the file of permissible passwords in an encrypted form. The systems designers believed that doing this would make the password system more secure. However, there have been cases of hackers obtaining a copy of the encrypted file of passwords and the Unix encryption algorithm. By feeding a dictionary of English words through the algorithm and comparing the results with the encrypted file, they were able

Figure 5.1 Proper names and commonly used system passwords should not be used as user passwords

ALEX
BACKUP
DEC
DEFAULT
DEMO
DIGITAL
DOG
FIELD
GUEST
HELP
HIAWATHA
IBM
LAZARUS
NETWORK
MANAGER
OPERATOR
OXFORD
RJE
SERVICE
SYSTEM
TEST
USER
VAX
VMS

Figure 5.2 The Internet worm program contained a list of less obvious passwords

Some of those included were:

AEROBICS
AMORPHOUS
ANTHROPOGENIC
BACCHUS
BEOWULF
CAMPANILE
CAYUGA
CERULEAN
CREOSOTE
EIDERDOWN
FOOLPROOF
FUNGIBLE
GRYPHON
GUMPTION
IMBROGLIO
JIXIAN
JUGGLE
LEBESGUE
NEPENTHE
NYQUIST
OCELOT
PERSIMMON
PROTOZOA
TARRAGON

to identify many of the permissible passwords. Using this type of cryptographic method, West German hackers managed to identify hundreds of passwords on US research systems. The error made by users suffering this kind of attack was to allow the encrypted password file to be copied. There is normally no need for this file to be read other than by systems programs.

One defence against these types of attack is to exclude passwords that are likely to be contained in a dictionary. Even this may not be foolproof for international users, however, because words that do not appear in a French dictionary (for example) may well appear in a German or an American dictionary. A solution to this problem is to incorporate at least one numeral or non-alphabetic character in passwords. This will make the password very hard to guess and immune from an attack based on the use of a dictionary.

Passwords are, however, fundamentally insecure, because they depend on something the user knows. This can easily be passed on to another person, with or without the user's knowledge.

Use authenticator tokens or smartcards for basic security

As businesses become more dependent on information technology, and as pressure grows to link systems together, the security offered by passwords alone will prove insufficient and unmanageable. New tools are much more secure than passwords, because they make access depend on something the user has, or on something about the user. This means that the user is in possession of some form of 'key' to the system, and as soon as he is aware that the key has been lost, access can be denied to the user's account.

We believe that Foundation members should begin to move towards this type of 'token-based' security over the next five years, starting with those applications most at risk, and selecting the most appropriate token method. Authenticator tokens are most suitable for use with existing terminals and for remote access from normal personal computers and laptop computers. Smartcards are a more convenient longer-term measure, but will remain somewhat inconvenient to use until terminals and workstations with built-in card readers become more widely available.

Authenticator tokens

Authenticator tokens are usually hand-held, calculator-like devices that hold an encryption-type algorithm. Unlike other types of tokens, authenticators can be used with existing terminals and keyboards. To log-in to a host system, the user enters his personal identification number (PIN) via the terminal in the usual way, and the system issues a challenge in the form of a randomly generated number. The user enters the challenge into the authenticator, which processes it through its algorithm and displays the result for the user to type into the terminal. In the meantime, the host system has carried out the same calculation and checks that the response to its challenge is the expected answer. If it is, the host system may reasonably conclude that the user is in possession of a valid authenticator.

Software packages are available for most major host computers through the authenticator suppliers. The secret key used by the encryption algorithm must, of course, be protected. This can be done at the computer end by the normal computer software security mechanism, or where higher security is required, through the use of dedicated secure hardware containing the security algorithm and keys. The physical construction of the authenticator is normally such that attempts to dismantle it will destroy its copy of the secret key.

These devices have been available for some time, and represent a fairly mature technology.

Particular products include Sytek's PFX random password generator and Racal Guardata's Watchword RG500 product. This kind of device is convenient to use, does not require special terminals or protocols, and can be used with existing equipment. It can be used from any location — even a hotel room. A photograph of the Racal device is reproduced in Figure 5.3.

Smartcards

The development of smartcards (also known as chip cards) provides one of the most significant advances in user authentication and the provision of secure access to computer systems. They have been in common use in France and Japan for several years in varying degrees of sophistication. Some of the latest 'super smartcards' have built-in keypads and calculator displays. More usually, they contain either a semiconductor memory, or a microprocessor combined with some memory. In either form, the card can be used as a security device, the memory acting as a key, with any processing to control access to the key being carried out in the card or terminal. Users insert their smartcard into their terminal and enter a PIN to initiate the log-on procedure. On receiving the PIN, the host engages in a dialogue with the smartcard to verify that the card belongs to the holder of the PIN.

One of the most secure forms of dialogue is based on the concept of zero knowledge systems, which enable the host computer to be almost totally sure that the user is entitled to access the system, but without the 'secret' itself ever being transmitted from the smartcard to the host. Zero knowledge systems are described in more detail in Appendix A.

If smartcards are to be used as security keys, it is vital to prevent access to useful information stored in them. The use of semiconductor memory for storing information within the card assists in this respect. The information is usually stored in the form of small static electrical charges buried within the silicon chip. These contain the security key. Physical access to the chip is made more difficult by barriers such as metallic and epoxy resin encapsulations. These barriers can also contain mechanisms to destroy the electrical charges, should the barrier be broken. Probing the chip by light beams, X-rays, or electron microscopes will also destroy the buried electrical charges. Thus, the physical security of smartcard keys appears to be strong. A typical smartcard construction is shown in Figure 5.4.

The one weakness of using a smartcard as a general security measure is that users might leave it in their terminals or where others might find it. The card is not, in itself, any good to an unauthorised user, because a PIN is also required. The PIN itself is, however, vulnerable,

Figure 5.3 Authenticators are calculator-like devices that can be used with existing terminals

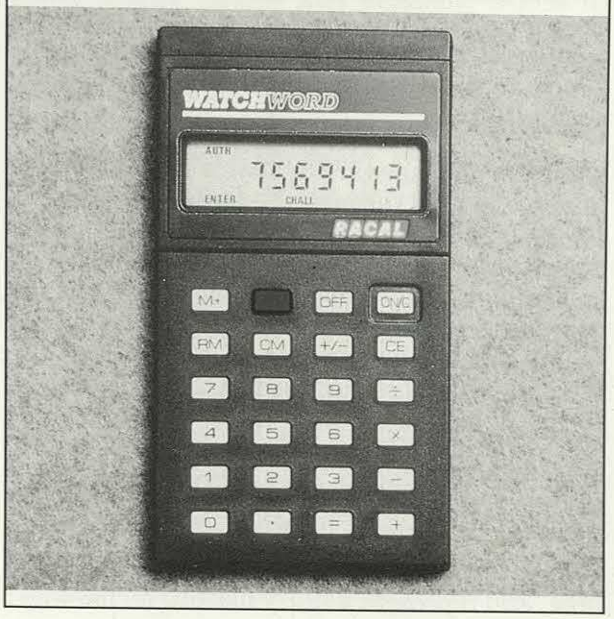
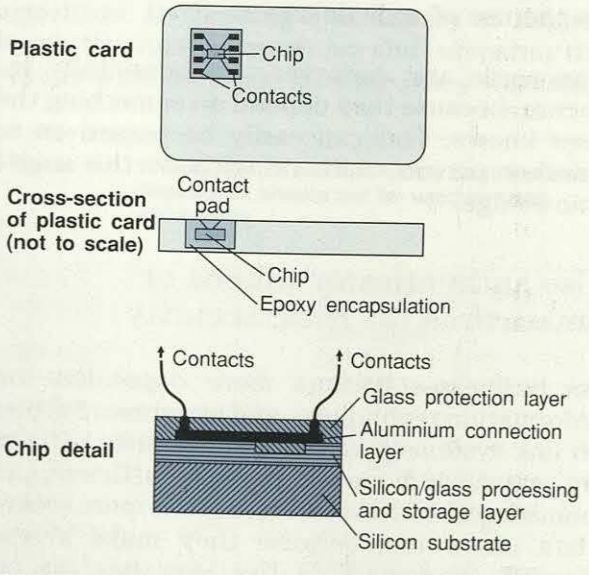


Figure 5.4 The physical security of smartcards is strong because of their construction



if the owner writes it down or is not wary of 'shoulder cruising'. One way of overcoming this is to create a bond between the card and its owner — for example, arranging for the same card to act as a key to the office, or providing a neckband or a waistband to keep it on.

The cost of smartcards, compared with alternative access-control cards, such as infra-red cards, magnetic cards, or laser cards, depends on the ratio of cards to reading terminals. However, the smartcard approach can be the cheapest, because of the simple technology needed to access the card — electrical contacts or induction loops and no moving parts. Smartcards will undoubtedly increase in popularity in the future.

Consider biometric methods for very-high-security systems

Biometric methods are based on measuring something about an individual, rather than something that he knows. Their objective is to make it extremely unlikely that a given stream of electrical impulses could have come from anything other than a valid biometric sensor and an authorised user. Many methods have been developed, fingerprint recognition being one of the oldest and most successful. Many of these techniques were originally developed for military use or for use in checking criminal records, and not all are acceptable for normal commercial use. To be acceptable, a means of identification must be socially acceptable, safe, not invade the user's privacy, and appear credible as a business method. Those measures generally considered acceptable for user authentication are shown in Figure 5.5. The first four are the most commonly used.

Signature-verification systems are among the most acceptable for business purposes. There are long legal precedents for accepting a signature as a binding authentication measure. Available signature-verification systems are of two main types — active pen, and active tablet. Active-pen systems use an instrumented pen to follow the changes in stylus acceleration and pressure created by writing a signature. Active-tablet systems require the signature to be written with a conventional pen on a special writing pad or tablet. Sensors within the tablet detect the motion of the pen, either by pressure, position, or in some cases, sound sensors.

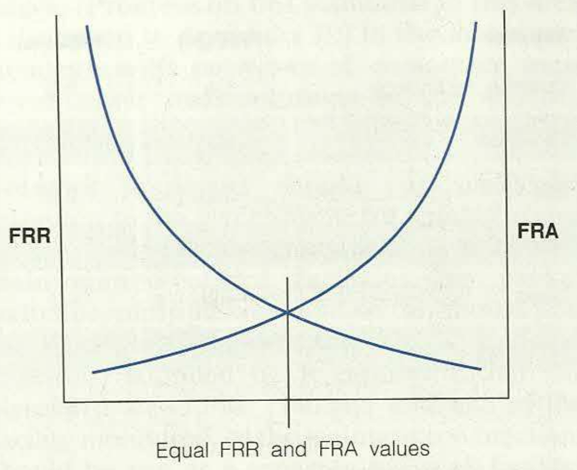
Figure 5.5 Biometric methods of user authentication measure something about an individual

Most commonly used methods
Signature
Facial shape
Fingerprint
Typing rhythm

Other methods
Voice pattern
Hand shape
Palm print
Foot print
Keyboard latency
Hand grasp
Wrist blood vessel
Eye blood vessel
Odour
Phrenological features

All biometric systems consist of some form of physical sensor coupled to some controlling logic, and a computer system designed to accept or reject patterns received from the sensor. The reliability of biometric-recognition methods is measured by the False Rate of Rejection (FRR), also known as the Type 1 error rate, or the 'insult rate', and the False Rate of Acceptance (FRA), also known as the Type 2 error rate, or the 'impostor rate'. Software within the biometric system can be adjusted so that the biometric 'lock' becomes stronger (a low FRA) or weaker (a low FRR). Figure 5.6 shows that

Figure 5.6 As the biometric system is adjusted for a low FRA (false rate of acceptance), the corresponding FRR (false rate of rejection) becomes higher



as the biometric system is adjusted for a low FRA, the corresponding FRR becomes higher. Generally, the two parameters cannot be controlled independently.

It is a difficult task for designers and users of biometric systems to optimise FRR and FRA values for a particular application. This depends on two main factors — the consequences of a false acceptance or rejection, and the relative shapes of the FRR and FRA curves. For example, an access-control system for a high-security military site will require an extremely low FRA value, and the personnel involved may be prepared to tolerate a high FRR in order to achieve this. On the other hand, for a banking application used by the public, too high an FRR may well result in loss of customers.

Different biometric technologies have differently shaped curves, and suppliers may have different criteria for quoting FRA and FRR values that they believe to be optimum. Figure 5.7 compares some FRR/FRA figures for commercially available signature-verification systems.

The primary disadvantages of most biometric methods are that they are slow — several seconds verification time is common — and they are expensive. There is also a limited amount of biometric data available for larger populations. Nobody yet knows how the FRR/FRA numbers will be affected for large populations, except in the case of fingerprints. Another problem is ensuring that any files of prestored information are secure. By definition, these will contain extremely valuable information for someone who wishes to gain unauthorised access to a system. Biometric methods will be actively developed, however, and more, and faster, methods are likely to appear during the early 1990s.

Figure 5.7 Different suppliers recommend different combinations of FRA and FRR values for signature-verification systems

Supplier	FRR (%)	FRA (%)
Confirma Technology	1.4	1.4
De la Rue	0.7	3
IBM	0.2	0.6
Inforite Corp	2	4
Quest	≈2	≈2
Signify Inc	0.2-1.4	—
TITN	3.5	3.5

(Source: JRP Consultants)

Apply access controls

As computer systems have been made available to more people both within and outside the organisation, so the need to control who can access the systems, and in which ways, has also grown. Such access control is concerned with much more than just checking that a valid password has been used. Different users will be entitled to access different subsets of applications and data. Some will be entitled to change the data; others will not. Some users will be permitted to access the system only via specific terminals or machines connected to a network. Commercially available software, known as access-control tools, can be used to help manage these problems.

Simply applying an access-control tool will, of course, achieve very little if it is not part of a wider security policy. In particular, access-control software requires detailed knowledge about access paths and access rights for individuals or groups of users. In a large-scale networked environment, these may be constantly changing, and in this situation, a security scheme based on access-monitoring and system-warning messages is the only practical way of controlling user access. To work at all, the concept of access control must be supported by management, and the varying levels of protection to be given to different types of information must be clearly defined. Otherwise, security will be so loose as to be ineffective, or so tight as to be unnecessarily restrictive.

In the IBM environment, the access-control tools most often used are Computer Associates' ACF2 and TOPSECRET, and IBM's RACF. RACF has recently been updated by IBM, and its ability to integrate with other IBM software has been improved. Computer Associates is beta testing versions of ACF2 and TOPSECRET designed to be used in networked systems containing a mixture of IBM and Digital hardware.

In the minicomputer area, there are many small suppliers of security software, characterised by technical expertise and a strong national focus, concentrating primarily on one manufacturer's range of minicomputers. Some of these products are designed to enhance and manage the basic security provisions of the operating systems, and others are intended to monitor or audit the security provisions in place. In both cases, the fundamental security provided by the operating system is unchanged. The objective is to improve the delivered security by eliminating hidden weaknesses and errors in the way the security provisions are used. The security-management approach is, of course, more expensive than the security-monitoring approach, but requires less skill and effort, which may be important in a larger installation. The disadvantage is that the procedures embodied in the tool must always be followed, and the reporting mechanisms must be trusted not to conceal any important information.

The problem with most of these proprietary systems is that the security offered works only within the hardware and software environment of one mainframe or minicomputer supplier. There are developments in hand, particularly in the Unix and OSI arenas, to improve this situation, but progress is slow, because good security goes to the heart of network and operating-system design. One attempt to address this problem is Project Athena.

Project Athena is a large network (up to 10,000 workstations) being developed at MIT. The project, which began in 1983, and is funded principally by Digital and IBM, addresses two main issues — the management of a large, distributed heterogeneous network, and the control of security within that network. Athena provides a single log-in procedure for a Unix environment, with a special emphasis on security.

Security within Athena is handled by a central authentication machine called Kerberos. Kerberos is based on a trusted authentication scheme where a user's identity, access rights, and privileges are held within an authentication database. Timestamping is used to prevent the replay of transactions, and encryption is used to prevent eavesdropping. Because Athena uses a variety of intelligent workstations that can easily be corrupted by clever users, the workstations are not trusted. Athena downloads trusted systems software at each log-in, and deliberately cleans up all network connections and temporary storage areas at log-out to prevent attempts at 'data scavenging' — the process of reconstructing the work just completed from data left lying about in the memory of the workstation or in temporary disc files.

Although Athena is in active use at MIT and will continue to be developed, the concept of a central authentication machine appears unsuitable for large-scale commercial development, partly because the system cannot, apparently, be scaled up to a size suitable for a very large commercial network, and partly because of the dependence on a single authentication device.

As we have seen, no single unifying means of providing secure access control is available today. Organisations must therefore address each of the potential threats separately. We discuss below the measures that Foundation members might take to control access to networks, databases, operating systems, and personal computers.

Network-access controls

To prevent unauthorised access to computer systems, the networks that give access to those systems must be controlled. Until quite recently, however, networks have been implemented primarily as data transport mechanisms; work on standards to enable public networks to control access to the network is only in its early stages. (Progress on OSI standards in this area is discussed in Appendix B.) In the meantime, members with networks of computers must devise their own solutions to the security problem.

Network managers should pay particular attention to the mechanisms for gaining access to the network for management purposes. Maintenance-control facilities for packet switches, multiplexors, and so on, should not, for example, be accessible via a standard password supplied by a vendor. Either the password should be changed and use of the facility monitored, or the maintenance functions should be run as a separate network facility, with no means of accessing application data files.

Loss of control over security is potentially one of the most serious problems of networked systems. Once computers are networked, the least secure part of that network defines the level of security that applies throughout the whole network. This means that even quite secure mainframes can be compromised by an insecure minicomputer or workstation. The following precautions should be taken:

— Where dial-in access is used, ensure that suitable dial-back modems are in use, so that the system can verify that access is being made from an authorised telephone number. In some countries, a dial-up connection is not broken until the caller hangs up. If the caller holds the line, a dial-back modem can be 'fooled' into believing that it has verified that the call came from an authorised number, even when it did not. The use of dual-line dial-back modems (which ensure that the modem calls back on a different telephone line) overcomes this problem.

— Where dial-back is not practical, consider the use of modem controllers that require an authentication code before access to any computer port is given. Ensure that all unused authentication codes and ports are blocked.

— Avoid contiguous dial-in telephone numbers, and select numbers on different exchanges from the company telephone number.

— Consider the potential threats very carefully before allowing modem access via a PABX that allows direct inward dialling to individual extensions. First, such a switchboard narrows the search for hackers trying to find a modem. Second, it is easy to arrange for calls to be diverted within the PABX, and thus subvert the dial-in security. Diversion is also possible on some modern public telephone exchanges. The use of 'follow me' services provided by the PTT can be used to subvert dial-back modems. Ensure that these facilities are not available to modem lines.

— Keep accurate records of all connections.

— Control the security of packet-switched networks. Where these are connected to the public network, ensure that any standard Packet Assembler Disassembler (PAD) passwords are changed. Hackers have been known to use the standard passwords to gain access to a PAD so that they can amend the control tables in a way that then allows them to gain access to a system connected to the PAD. Consider monitoring the ports that are connected to a packet-switched network. They, too, have been used by hackers to access systems, particularly where failed log-in attempts are not registered. This omission means that the system is vulnerable to an automated password-guessing attack.

Database-access controls

Internal database controls define who can access and amend specific items of information, and usually this arrangement works adequately to preserve data integrity. Three matters deserve particular consideration, however:

— The password-verification system for logging on to the computer is often separate from the password-verification system for the database. This means that no check is made to ensure that a user seeking to access a database is the same user who has just logged on to the computer. It is critical that the two password systems are kept in step; this is a tedious administrative process into which errors can easily be incorporated. Product upgrades that address this problem are becoming available.

— The introduction of strict database access controls may have a negative impact on performance, and in such cases, there is a danger that security will be a low priority. Where security is compromised in this way, the security policy committee should be informed.

— Databases are usually accessed through application programs that take care of the security arrangements. Use of other database-access tools, particularly utilities that can alter database records and fields without making an audit log, must also be carefully controlled if database security is not to be subverted.

In the business environment, database security is usually more concerned with data integrity than data confidentiality, but in particular circumstances, data confidentiality could be an issue. Potential weaknesses in the area of confidentiality are data inference and data aggregation:

- *Data inference* is the ability to draw a conclusion from available data where the inferred conclusion is deliberately hidden. An example might be a database that allows salary figures (without names) and skills (with names) to be accessed. By comparing known salaries — those of senior directors and one's own salary, for example — with the skills list, it might be possible to identify that the skills are listed in the same order as the salaries. If they are, the relationship between salaries, names, and skills can be inferred.
- *Data aggregation* is concerned with drawing a conclusion, which would otherwise be hidden, by examining several apparently unrelated data items. For example, by accessing information on skills (including languages), travel dates, and names, it might be possible to determine that a particular type of project is about to begin in a certain country.

Research work aimed at protecting secret information in relational databases is being carried out by Teresa Lunt at SRI International, Menlo Park, California, and commercial database developers (including Oracle Corporation) are known to be involved in research in this area. Eventually, the results of this work may become incorporated in commercial products. Until then, these problems can be addressed only at the application level by ensuring that sensitive linkages cannot be made.

Operating-system security

Today's access-control security systems are closely interlinked with the computer's operating systems. In the main, the security features work well and additional packages are available to enhance security when required. Suppliers of operating systems and access-control software are not, however, convinced that user organisations fully understand or use all the security features that they provide, whereas users claim that the tools are difficult to understand, difficult to use, and limited in their application. We recommend that user organisations should take the following actions to improve operating-system security:

- Ensure that staff understand the security facilities available by sending them on courses and subsequently contacting suppliers to resolve any problems or misunderstandings.
- Monitor the security of the software systems in use. Set in place standards for new application designs. Seek advice from supplier specialists about the most obvious weaknesses in their systems, and ensure that, as far as possible, these are eliminated. Look for common defects such as open access points, buffers and files left in a vulnerable state, and so on.
- Ensure that software standards and guidelines are integrated with the organisation's security policy and plans — for example, an eventual move to smartcards or authenticators.
- Be aware that government bodies are setting up standards for evaluating software (and other systems) security. The best known is the US Orange Book (and Red Book) series, but efforts are being made to establish European standards for commercial computer security. In the United Kingdom, the Department of Trade and Industry has published a draft series of 'Green Books'. This is not as far advanced as the Orange Book series but UK members should watch for developments. In West Germany, the Zentralstelle für Sicherheit in der Informationstechnik (ZSI), a government body, has also produced a Green Book, setting out standards and evaluation criteria for commercial and industrial computer systems. This is not as explicit as the US Orange Book, but German members should track further developments in the standards and guidelines. There are also initiatives to establish certification bodies for security mechanisms based on 'colour book' standards.

No operating system is perfectly secure, and defects will come to light from time to time. Obviously, suppliers are reluctant to publicise the defects, both for commercial and for security reasons. Systems managers should maintain close liaison with supplier experts and recognise that while they may be reluctant to spell out the details of a defect, they may well be prepared to advise that a particular 'fix' can be applied.

Personal computer security

The best known problem that personal computers suffer from is viruses. Many members have taken the precaution of issuing all their users with a pamphlet warning of the dangers of viruses and offering advice on how to avoid introducing a virus into their personal computers. In the main, this has been very successful. Not everyone, however, has been lucky. One company had a personal computer infected when a new recruit, who had not been issued with the warning pamphlet in time, ran a disc infected with the 'Aids' virus. Foundation members should quote the recent publicity about hackers and viruses to encourage personal computer users to take security seriously.

Where users store sensitive data on personal computers, removable hard disc units should be used so that they can be locked away when not in use. Where access has to be restricted, the use of proprietary access-control and file-encryption systems should be considered, ensuring that any such system is relevant to the kind of threat that is being guarded against. Simple access locks or password schemes will guard against the casual data 'groper'; stronger measures may be necessary to defeat the determined data thief. Encryption systems requiring an additional circuit card in the personal computer are, for example, very secure, and strong encryption schemes are impossible to crack, even for those who designed them. The problem is that hardware-based systems such as these use up one of the spare slots in the workstation. This may be a problem when encryption is required in certain types of applications (financial-dealing or CAD systems, for example) that already use most of the spare slots.

Encryption methods for personal computer security have other disadvantages:

- The encryption process slows down access times, which can be quite serious for spreadsheet applications (still one of the most common uses for personal computers).
- The introduction of a single error in an encrypted file can destroy a substantial amount of data. A single bit error could easily invalidate 64 characters.
- If a user loses the encryption key, the information is effectively lost.

To avoid security problems with radio emissions from personal computers, users can apply TEMPEST techniques. These are a series of techniques concerned with controlling the amount of radio frequency electromagnetic signals entering and emanating from electronic equipment. Originally developed for military and government use, they are now available in a commercial form and are suitable for use in sensitive applications, where eavesdropping, in particular, is likely to be a threat. An alternative is to use devices that emit a scrambled radio signal at the same frequency as the workstation's display-scanning rate. This will swamp any emitted signals.

These techniques are not cheap, however, and should be considered only in extreme circumstances. For the occasional workstation user who is concerned about radio emissions, the most practical solution is to surround the workstation with similar devices, each running a program that varies the screen pattern frequently — a game program, perhaps. Some experts advise standing the workstation on a wooden table, well away from metal objects.

Use encryption as a security tool

Cryptographic techniques enable systems designers to provide message and data confidentiality, or integrity, or both. Applied to data, encryption is the orderly transformation of one bit stream into another, such that the output bears no apparent relationship to the input. Encryption algorithms are used to 'lock up' data with a mathematical cipher that requires massive amounts of computer resources and time to unlock without the correct key (which may well be different from that used to lock up the data). These processes of encipherment and decipherment are so complex and so computationally intensive that special hardware and/or software is required to handle them.

Used correctly, encryption is the most secure means of protecting confidential data against unauthorised disclosure. It is a more effective means of protection than the access-control mechanisms described earlier in this chapter, which are relatively simple for hackers and other intruders to defeat.

The strength of an encryption code is governed by the laws of mathematical intractability, and by the complexity and granularity of the overall cryptographic system design. Modern cryptographic systems are constructed in such a way that it is not possible, even with the knowledge of the precise techniques used, to trace back the transformations made so as to mount an attack on the application using the particular cryptographic system. However, the all-important issues are how the keys used for encryption are chosen, to whom they are made known, how they are made known, and how they are used. The problems of distributing and changing keys are discussed below.

Choosing between hardware and software encryption methods

Encryption is generally performed using either a program running within an organisation's computer (software encryption) or a special-purpose electronic device (hardware encryption). The results are usually identical, but various factors need to be considered in choosing one or the other method.

Dedicated hardware devices are typically some ten to a hundred times faster than their software equivalents. Speed is the main advantage of hardware solutions, and speed is often an essential requirement. Hardware devices are also generally easier to protect physically, but they are expensive to build and they lack flexibility. The relative ease with which software encryption may be upgraded is advantageous where the pace of change within secure systems is a complicating factor.

The importance of managing keys

Encrypted data remains safe only as long as the keys used for encryption remain safe. The generation, distribution, storage, and regular changing of cryptographic keys must therefore be managed in an efficient and secure fashion.

Usually, two levels of keys have to be considered — data-encrypting keys for the protection of data, and key-encrypting keys for the protection of keys during transmission or while held in storage. Since keys become more vulnerable the longer they are in use, data-encrypting keys are changed frequently. It is generally advisable to do this once per session, and because of this, the data-encrypting key is often also known as the session key.

Kinds of encryption

For the commercial user, two groups of cryptographic systems are available, based either on industry or *de facto* standard algorithms, or on proprietary or purpose-built algorithms. Both can be used satisfactorily and with the assurance that they are 'registered' by ISO. For reasons of national security, the member bodies of ISO have agreed not to standardise cryptographic algorithms as such, but instead, to set up a registration authority to be operated (in Europe) by the UK National Computing Centre. The register maintained there will contain information supplied to it by algorithm builders regarding the characteristics and availability (for example, import/export restrictions) of the registered item. The register will not, however, provide any qualitative information enabling users to assess the relative strengths and weaknesses of the listed algorithms.

There are two well established 'standard' algorithms available today — the DES or Data Encryption Standard originally developed by IBM, and the RSA algorithm, named after its inventors, Rivest, Shamir, and Adleman. Although the methods by which these algorithms work are public knowledge, their use, particularly as hardware devices, is controlled by government agencies, and by patent and product-licensing restrictions.

Private algorithms have a role to play where the strength and cost of DES or RSA are not required or not justified, or where international import/export restrictions make their use difficult. Private algorithms are not necessarily stronger or weaker than DES or RSA; they simply have a different role to play. One advantage of a private algorithm is that it is frequently easier to 'tune' it to meet particular requirements.

From a technical stand-point, there are two basic kinds of encryption mechanism — symmetric and asymmetric systems:

— Symmetric, or private key algorithms use the same key to lock and unlock data. The best known symmetric algorithm is DES, which is widely used particularly in the banking industry. For a symmetric algorithm to work securely, there must be a highly secure means of distributing the keys. Much has been written on key distribution and many methods are available, but it is a complex and difficult task to do well, especially where hundreds of dispersed sites are involved.

— Asymmetric, or public key algorithms use one key to lock the data and another key to unlock the data. Thus, the user can make his encrypting key widely known (that is, public) but will keep his decrypting key secret; anyone can therefore encrypt a message to send to him, but only he will be able to decrypt its contents. Public key systems are particularly useful in overcoming some of the key-distribution problems referred to earlier, and for applications where a large number of non-trusted users are connected into the network.

In certain applications, it is appropriate to use a combination of private key and public key systems, gaining the best from each. (The distinction between public and private keys is fully described in Report 51, *Threats to Computer Systems*.)

Security of encryption

Designing strong encryption algorithms is a job for experts, and for the vast majority of commercial purposes, the existing algorithms that have stood the test of time will provide more than adequate security. No successful attempt to decipher information encrypted using DES has so far been proven, after almost 20 years of use by many financial institutions. There are more than 70,000 billion possible DES keys, and there is no known way of cracking DES other than by trying all possible keys until the right one is found. Technically, this is known as exhaustive key searching, and it has been estimated that DES would require more than 1,000 years of computing on a Cray-2 supercomputer, or in excess of 10 million years on an IBM-AT, to complete this task.

Nevertheless, DES has been subject to criticism and has some perceived technical weaknesses. DES depends heavily for its strength on the 56-bit key used to encrypt the information, and it has been argued that this key length is too short. In addition, some bit patterns are known to result in weak or semi-weak keys, easy for a crypto-analyst to deduce. However, these keys are well known and can readily be eliminated. Likewise, the strength of DES can be further improved by double and triple application, a simple but effective way of overcoming possible key length problems.

RSA keys are typically 512 bits long and provide strong security, particularly in more open network situations. RSA is often used to encrypt DES keys prior to their transfer across a network. Unlike DES, the strength of RSA is critically dependent upon the key length chosen — the greater the length, the more secure the algorithm. In order to crack RSA, huge numbers must be factorised — for a 512-bit key, this would typically take around 90,000 years using a Cray-2 supercomputer.

Other uses of cryptographic techniques

One of the most common uses of cryptographic techniques is for message authentication. Authentication can, of course, be achieved without recourse to the use of cryptography, but it does provide significant advantages in many cases where data integrity is essential and confidentiality is desirable. Authentication is the means whereby the receiver of a message can validate its source and all or part of its contents, while at the same time being assured that the connection is not with someone attempting a masquerade. Such an authentication method involves a key-establishment stage as well as the authentication process itself, and protocols using both private and public key systems have been developed for this purpose.

For many purposes, it is not necessary to encrypt the entire message, but it is essential to ensure that the message is not altered after it leaves the sender. Financial transactions are an example of this kind of message. Well known methods of ensuring the authenticity of messages are:

— Appending a message authentication code (MAC) or cryptographic 'checksum' to the message.

— Using digital signatures, which transform the message, or a condensed version of it, into a cryptographically derived code or 'signature' unique to that message.

Any accidental or deliberate corruption of the message will destroy the unique correspondence between the message contents and the appended MAC or digital signature.

To guarantee that an electronic message is authentic will generally require that a digital signature be assigned to that message. Otherwise, the message could easily be altered, corrupted, or forged, and could not be relied upon for commercial transactions. If the source or the timing of the message is contested, the digital signature can be used to confirm the authenticity of the message. The use of digital signatures therefore facilitates open trade and the expansion of EDI and similar services.

Where the parties involved are not previously known to each other, or for some reason do not trust each other, the use of digital signatures can be coupled with a notarisation service to help resolve disputes. The notarisation service acts as an electronic notary or independent witness that can be called upon to prove that not only was a given message sent, but that it was also received and acknowledged as being received. Thus, disputes involving the denial of sending or receiving messages become futile because the notarisation service can be called upon to prove the matter beyond dispute.

Build in integrity checks

The integrity aspects of security must be designed into systems at the start. Built-in checks and balances will help to ensure that errors, as well as frauds, are automatically prevented.

Use reconciliation in files and programs

A degree of integrity is assured by introducing routines that check the internal consistency of data. This is a normal design technique that, at least, ensures that 'the books balance'. Closely allied to this technique is the concept of 'the well formed transaction', discussed by David Wilson, an authority on systems security, who addressed members at the International Foundation Conference in Cannes in 1989. The well formed transaction is an accounting principle that should be embodied in application programs and audit trails. In essence, it, too, ensures that the books balance. For every debit entry, there is a corresponding credit entry, and it is never possible to access one entry without affecting the other. The well formed transaction can be ensured through good program design and good access controls.

Use inspections for systems integrity

The corruption of existing program code can be largely prevented by strong physical protection and by software digital signatures. There is, however, little that can be done to prevent a determined programmer from introducing a Trojan horse or a logic bomb into a program at the design or coding stage. Inspections and walkthroughs should find most of the problems at the design stage, but as with any manual method, there is always a chance that some will escape detection.

In theory, it should be possible automatically to compare the design specification with the delivered code. Some work has been done in this area, particularly for safety-critical applications. Verilog, a specialist French software house, offers automated design and simulation aids intended to verify that industrial process-control programs under test have only those executable paths that the design says they should have. In time, perhaps, this kind of aid will be available to commercial-application programming teams working on mainframes. Until then, most organisations must rely on employing trustworthy people and formulating good back-up and contingency plans.

Separate critical duties

Although they are an essential feature of systems security, technical countermeasures will do little to prevent dishonesty. David Wilson and his colleague, David Clark, have pointed out that although many computer security schemes address confidentiality and availability, few fully address the integrity of information. Information integrity is greatly enhanced by the well formed transaction, discussed earlier, and the separation of duties.

The separation of duties is an organisational countermeasure that is under the direct control of management. It must be remembered that computer applications are merely systems for manipulating numbers and symbols. Computers have no means of checking that the numbers and symbols match reality. Application designers have to rely on the integrity of individuals, entering data that correctly represents reality. Separating the duties of key individuals is one way of ensuring that the computer cannot be deceived by a single individual. Of course, collusion between individuals can subvert the separation-of-duties principle, but other measures, such as the periodic rotation of duties among individuals, will restrict opportunities for collusion.

In one company we spoke to, the separation-of-duties principle had been subverted inadvertently owing to a shortage of staff. Staff had two passwords and were able to perform each of the separated actions under the appropriate password. Passwords were shared, so that holidays, sickness, and so on, did not interrupt operations. No fraud was committed. The staff involved were conscious of the risk and kept silent to avoid advertising the threat to security. Only by looking at the complete business system in the light of the Clark/Wilson concepts of the well formed transaction and the separation of duties can this kind of error be avoided.

The separation-of-duties principle must be maintained, regardless of local difficulties, or broader organisational problems. Where a shortage of staff or the nature of business structures make local separation of duties difficult, it may be necessary to use information technology to separate the duties geographically. As business structures become flatter, and as responsibilities are devolved further down the organisation, the principle must continue to be adhered to. It is fundamental to the preservation of information integrity.

Preventing breaches of systems security is, as we have seen, a complex task for the systems manager. However stringent the security procedures that are in place, there is always a possibility that security will, at some time, be breached. It is therefore imperative that systems managers are also aware of methods of detecting such breaches and of limiting the damaging effects of any that do occur. These concerns are the subject of Chapter 6.

Chapter 6

Detecting and limiting the effects of breaches of security

Security is about the prevention and detection of threats that might damage a business. Prevention was discussed in Chapter 5, and as we saw there, it is impossible to ensure total security. One systems manager told us that the only way to make a computer totally secure would be to ''unplug it, case it in concrete, and sink it in a deep ocean trench''. In other words, a totally secure computer would be unusable. Detecting breaches of security is therefore also an essential part of a security policy. If the trail of events is followed quickly enough after something has happened, action can be taken to limit the damage that might be caused. In this chapter, we describe the methods and techniques for detecting breaches of security early in their development, and some of the measures for limiting the effect of the breaches that do occur.

Closing the loop to make detection automatic

A fundamental tool in the design of any type of system (including computer systems) is the notion of feeding back the results of an action to the source of that action, and comparing the desired result with the actual result. Of course, feeding back to all users the security implications of every transaction that they initiate is plainly impractical. First, the volume of data would, in most cases, be overwhelming. Second, most users would require enormous amounts of time, patience, and motivation to analyse the data. Third, users would need to have detailed knowledge of the way the system operates. The feedback concept cannot therefore be used at the detailed level, but it can be used where exception reports relating to security are available.

One of the most useful techniques is to feed back to users some of the analyses of log-in records. However, this must be done in such a way that users will see it as a help in identifying problems rather than as covert spying. In one organisation where such a policy was in operation, problems were spotted in the training process, when users were unsure how the password system operated, and did not know how to respond when a password was changed and then forgotten. Such a policy has also served to identify users regularly trying to exceed their access rights.

Members who have used this technique feed back log-in and other security monitoring results to business management, highlighting deviations from normal practice, and providing guidelines on action to be taken. The reports are sent to business managers because they normally have a good understanding of the day-to-day operations connected with the actions that have initiated the monitoring reports. The reports should therefore include information such as dates and times of use, files accessed and updated, and so on, so that business managers can spot, or account for, any anomalies.

Developing sensitive monitoring systems

For security-monitoring systems to be of value, they must be sensitive enough to detect the occasional attempt by a hacker to gain access to a system, a user's efforts to extend his knowledge of the system, and the first awakenings of a virus program. One company's security advisor claimed, ''if your monitoring system does not detect anything, it is not sensitive enough''. An analogy can be drawn with tuning into a weak radio station — the volume needs to be turned up in order to find the station; it can subsequently be turned down for normal listening. The systems security manager does not have access to a simple 'volume control'; the tools and the availability of information are restricted and imperfect. However, three practical steps can be taken — monitor access (both to buildings and offices and to computer systems), use vaccines to detect and remove viruses, and institute controls in the form of regular data reconciliation and plausibility checks.

Monitor access

The first practical step is to monitor physical access to office and other areas, so that the security monitoring systems know who is present, when they arrive and leave, and where they are at any given time. Information about who is physically present should be related to software access controls, and the two records should always match. Since most systems security problems derive from internal users, this helps to control a major source of security breaches. Several Foundation members, for example, provide their security staff with printouts from the physical access logging systems. They can be compared with terminal access logs and file access logs to identify staff who are roaming from one department to another, outside normal working hours.

Another approach to improving the detection of security breaches is to build alarm indications into applications software. Barclays Bank and others have done this to detect ATM and credit-card misuse by building information about likely fraudulent transaction sequences into the application program. The Barclays system analyses credit card transactions for abnormal transaction types and looks for rapid sequences of cash withdrawals. Other systems detect the use of ATM machines for such events as cash withdrawals on the same card from widely dispersed locations, or an unusual pattern of withdrawals. Some ATMs are in constant use 24 hours a day; others are rarely used late at night. If a normally unused dispenser is used for several transactions late at night, it can be shut down until further checks are made.

Essentially, this kind of security represents an electronic 'trip-wire' set off by some predetermined event. Some members have incorporated these types of security checks into systems and arranged that user identities and keystrokes are recorded for subsequent analysis. Although this approach may prevent or limit the possibility of fraud, it may also cause inconvenience to legitimate users. Great care is necessary in designing systems that include these types of checks, but where valuable assets are at risk, it is a worthwhile approach.

Use vaccines to detect and remove viruses

Prevention is better than cure, and ideally, viruses would never get into computers in the first place. Even the best run installations, however, can suffer from a virus attack, and Foundation members should know how to detect and cure a virus. Until recently, viruses have been regarded as a serious nuisance, but little more. Most viruses were easily detected, relatively benign in their effect, and easy to cure. The situation is now more serious, as the effects of viruses are becoming more severe and less easy to cure, as they begin to affect local and wide-area networks, and as publicity begins to undermine confidence.

Virus detection depends on the fact that, to become active, a virus must attach itself to, or alter, an existing program. The virus, however, will not be activated until the 'host' program is run. A virus can reside on a floppy disc or a hard disc and do no harm whatsoever, so long as the executable component of the virus is never activated. It is crucial that companies know what to do and what not to do should a personal computer become infected with a virus, both to eliminate the infection and to avoid the massive loss of confidence that an infection can bring. There have been reports of companies discarding perfectly good personal computers simply because no-one in the systems department could get rid of a virus. If a local area network were to become infected, the results of such a policy would be expensive and extremely disruptive.

Detection methods depend on three main techniques: looking for virus 'signatures', identifying changes in the size of an existing program, and using encryption to protect program files:

- The simplest technique involves running a program that looks for virus 'signatures' — the specific bit patterns belonging to a virus program. These detection programs are fast to run, but are highly specific to a given virus, and are unreliable if an adapted version of an existing virus is created. As new viruses appear, new signature-detection programs are required. The cost of these is usually small, however, and minimal compared with the cost that might be incurred if a virus remains undetected.

- The size-change detection technique uses a program that identifies the additional memory requirement that some viruses add to existing programs. This technique is also fast, but is not reliable in detecting all virus types. It is possible, although difficult, for a virus to compress a program file such that the length remains unchanged. Alternatively, a virus can alter a small part of the infected program without changing its length, and keep the main body of the virus in a separate, hidden, file.
- The strongest virus-detection technique is to use hashing or encryption methods to generate a unique identifying number from the contents of a program file that is known to be free of viruses. The technique works by recalculating the number and comparing it with the 'trusted' number. Any change will indicate that the program file has been tampered with. Even if the encryption method is known, it should be impossible to make any significant change to the file without detection. This approach is slow to use, however, and depends on securely protecting the trusted number. Sometimes, the slowness of the approach means that program files are checked only daily or at random intervals. This obviously represents a compromise between security and ease of use.

Removing a virus involves isolating it and then either undoing the damage to program and data files, or replacing them from back-up files. Infected personal computers must immediately be isolated from any host computers and networks to avoid possible spread of the virus or re-infection. Infected local area networks must be checked, workstation by workstation, and at any servers connected to the network. All disc files must be examined for infection. Users must be aware of the importance of doing this and must be persuaded to submit *all* their discs for checking. It is also essential to ensure that no program copies stored on the back-up files have become infected. A virus-free library copy should always be kept separate from day-to-day activity. Where virus detection tools locate virus fragments among data files, the fragments should also be removed.

All this work is expensive and time-consuming. It is essential that those responsible for handling virus problems can either deal with the problem quickly and effectively themselves, or can call on experts who can help. The best way for systems staff to learn how to remove viruses is to receive training from specialists in this field.

Do data reconciliation and plausibility checks

In Chapter 5, we described how the principle of the 'well formed transaction' ensures that all data within a system is consistent and that it cannot be altered without a record of the alteration appearing. In a perfect system, this approach would preserve the integrity of data. However, systems are not perfect, and errors do occur.

One approach to detecting flaws in data integrity is to run error-checking routines continuously. These are designed to provide additional support to the well formed transaction principle. This approach works by ensuring the presence of supporting and balancing data elements and, where possible, doing checking calculations.

An example of the value of this approach comes from a financial-services organisation. Subtle faults were identified in some new applications programs, which were introducing errors into an online database. The early identification of this problem prevented the errors corrupting the entire database. Another example comes from a utility organisation that has incorporated plausibility checks into its billing procedures. This kind of organisation has had checks in place for many years to avoid sending out bills that are obviously wrong — with massively high, zero, or negative amounts. These checks have been developed and made more sophisticated by the use of moving-average and other statistical techniques to identify bills that are significantly different from the past pattern. These are sent to an 'exception' file and examined manually before being printed and sent to customers. The objective is to preserve the reputation of the utility, but the principle can be applied to other forms of integrity check.

Using existing tools to best effect

Today's access-control and computer-management systems can produce a complete audit trail of who has used what systems, when they were used, and what they were used for, but it is frequently impractical to store and analyse all of the information produced. A large banking organisation reported that it would need to process and store around one gigabyte of data every day if it were to adopt this approach. A well managed exception-reporting system is a preferable alternative, but there are problems with this approach:

- The tools must be used to protect *all* systems resources, not just those that are managed directly by the systems department.
- The facilities available with most access-control tools, such as the facility to grade the sensitivity of information, and the facility to restrict the days and times during which access is permitted, or locations from which access can be made, are not fully used. Identifying all the systems resources requiring protection and assessing the relative importance of information requires top management support, and restricting access means that those managing security must be delegated with the necessary authority.
- Those running the systems do not always know who accesses every resource, when they access it, and why.

For business reasons, it may be desirable to allow free access to most of the information. A totally open-access regime does, however, compromise security. A practical approach is to start with an open-access regime, monitor it, and gradually restrict access in the light of the monitoring information. During this phase, however, many access-control tools can be set to give warnings to the security controller rather than to block access. Once the access regime is working effectively, warning messages may gradually be replaced with blocking controls. In a dynamic business environment, management must constantly be aware of the need to maintain a delicate balance between restricting access and ease of access. Legitimate users should be unaware of the presence of an access-control regime. If the security measures hinder legitimate users, they have failed to achieve one of their prime objectives.

Tracking advances in detection tools

The problem with monitoring computer systems for possible breaches of security is that large volumes of irrelevant data can be produced, which is time-consuming and expensive to handle manually. Better automatic monitoring and detection tools are needed, and some advances are being made in applying expert-system techniques in this area. Two such projects are the Intrusion Detection Expert System (IDES), being developed by Teresa Lunt and her team at SRI, and the Wisdom & Sense expert system, also developed in the United States.

IDES, which is described in Figure 6.1, is designed to detect hackers, internal penetrators (masqueraders and clandestine users), and users who try to exceed their access rights. IDES works by 'learning' the behaviour of each user and detects significant shifts in behaviour. Work

Figure 6.1 The IDES system is designed to detect threats automatically

IDES (Intrusion Detection Expert System) is being developed at SRI International's Computer Science Laboratory in Menlo Park, California, funded by the US Navy's Space and Naval Warfare Systems Command. The system runs on two Sun workstations — a 3/260 with a 560 Mbyte disc to process the data, and a manager's enquiry terminal using a 3/60 — and is based on Oracle's database management system and SQL. No *a priori* rules are built into the system; instead, it 'learns' the behaviour of each user and detects significant changes in behaviour. It works by monitoring such factors as log-in time and location, the amount of connect time, CPU time, input/output usage, and any protection violations. Log-in time is divided into three parts — day, evening, and night/weekends/public holidays. The violations reported on include directory modifications and password errors.

This approach is unreliable, however, if it is introduced when people are already abusing the system, because IDES will simply learn their bad habits. Furthermore, because IDES maintains profiles of user behaviour averaged over 50 days, it could be defeated by a clever user who slowly varies his usage profile over a long period. Whether system abusers have that much patience is open to doubt. Work is proceeding to address these weaknesses.

reported so far suggests a false rate of acceptance (FRA) of about 1 per cent and a false rate of rejection (FRR) of some 5 per cent. This is comparable with biometric measures, but of course, IDES detects offenders only after they have gained access to the system.

Wisdom & Sense seeks to allow the systems security manager to 'manage by exception'; only those events that are not allowed are reported. It works in two parts — the Wisdom portion is used to develop a rule base, and the Sense portion is used to detect and report anomalies. The system, which is described in more detail in Figure 6.2, is on beta trial at several US sites and at the US National Computer Security Center, where in July 1990, a Tiger Team intended to attempt to intrude on a system running Wisdom & Sense. One of the problems with this detection system is the weighting of the rules. Under some circumstances, certain rules can become highly rated and dominate the analysis. Another is the false-alarm rate. Developers have attempted to reduce the false-alarm rate by comparing low event probabilities with the probability that such an event could be random, and eliminating unnecessary alarms.

Security tools based on expert systems are at an early stage of development, but it is worth tracking their development and looking out for products that incorporate such techniques. When selecting automated tools for detecting breaches of security, Foundation members should pay particular attention to the ease with which the tools can be adapted to cope with changing business requirements. Access requirements are continually changing, and the detection tools will have to be adjusted accordingly. Those described above appear to need a considerable amount of tuning before becoming sensitive and reliable.

Limiting the effects of breaches of security

Despite all the preventive and detection measures that may be taken, there is always a chance that a security breach will occur. It is therefore important that systems managers limit the damage that can occur, plan for back-up, identify when and where to get help of the right kind, and ensure that any insurance cover is adequate.

Damage limitation

The concept of physical fire walls is well understood in building design and has been extended to the design of computer centres. This aspect of damage limitation is well described elsewhere and is beyond the scope of this report. The fire-wall concept can, however, be extended to 'logical fire walls' that protect data assets. These include access controls, the encryption of critical files, and the division of critical duties,

Figure 6.2 Wisdom & Sense uses expert-system techniques to detect possible threats

Wisdom & Sense is an expert system developed in the United States by Hank Vaccaro and Gunar Liepins, at Los Alamos National Laboratory. It contains several rule bases generated by human input and from an analysis of past systems activity. These rule bases are used concurrently, each one looking for a particular range of possible threats:

- The physical rule base is concerned with where transactions come from.
- The policy rule base contains site-specific rules.
- The administrative rule base contains access permissions and so on.
- The intruder rule base contains information about how to detect intruders.
- The historical rule base is built from past experience.

Security audit records are processed to form individual 'rule trees' that are highly specific to individual users and to groups of users. Every few weeks, the collected audit records are processed to form a large 'rule forest' of highly specific rules. Patterns that occur frequently are graded on a scale from high to low, to reflect the quality of the rule.

All users of the system have a database entry that holds a score developed from the historical information in the rule bases. As a user inputs further transactions, the details of the transactions are written to an audit log and processed against the rule bases and against past experience with that user to develop a new score. Scores that exceed predetermined thresholds are notified to the systems manager as possible breaches of security.

The system runs on an IBM RT 6151-125 running AIX with a floating point accelerator card. This system can process about 20 audit records per second. Preprocessors can allow several different kinds of computer audit records to be processed by the same system against their own rule bases.

all of which have been discussed earlier in this report. Other logical fire walls include running high-risk applications on separate machines, rotating critical duties, and ensuring that back-ups are not always made by the same person, nor all accessible by one person.

Back-up planning

There are five stages involved in planning and implementing the recovery of an information system. Many organisations that have developed a back-up contingency plan have considered the first three, which are concerned with defining and testing the plan; the last two, which are concerned with recovering from an actual disaster and returning to normal working, are less often explicitly considered. The five stages are illustrated in Figure 6.3 and described below:

Stage 1: The first stage is to decide whether or not to support all systems in a back-up situation. Where systems are closely interlinked, it may be more effective to leave them that way and plan to back them all up. Otherwise, those that it is most critical to recover should be identified.

Stage 2: The second stage is to produce a back-up plan. Many organisations have prepared back-up plans for recovering immediately after a disaster; fewer have plans for using back-up facilities for an extended period and then returning to normal working.

Stage 3: Stage 3 is concerned with testing the immediate back-up process. Where mutual back-up arrangements have been made with another organisation, the partner will almost certainly have to go into a back-up support situation too, because it is unlikely that it will have enough capacity to run all applications. It may therefore be necessary to run extra shifts, to abandon some applications, or to accept slower response times. This needs to be planned for as well. It is important to test the plan regularly, since back-up plans rarely work properly first time.

Stage 4: Stage 4 is concerned with recovering from a disaster, such as that shown in Figure 6.4, which means that back-up facilities will be required for an extended period. In such a case, the planning and testing undertaken at Stages 1, 2, and 3 should ensure adequate short-term recovery. The problem is that Stage 3 cannot last for ever — back-up contracts are typically for a few weeks only. A longer-term back-up plan is also needed. Typical solutions are some form of portable cabin located in a car park, spare office space that could be turned into a temporary computer room, or reliance on speedy removal of damaged equipment and renovation of damaged facilities. The choice must be made during Stage 2. Computer systems (even mainframes) may be replaced fairly quickly, assuming that a replacement is available. In the area of telecommunications, however, racks of modems, front-end pro-

Figure 6.3 There are five stages involved in recovery planning

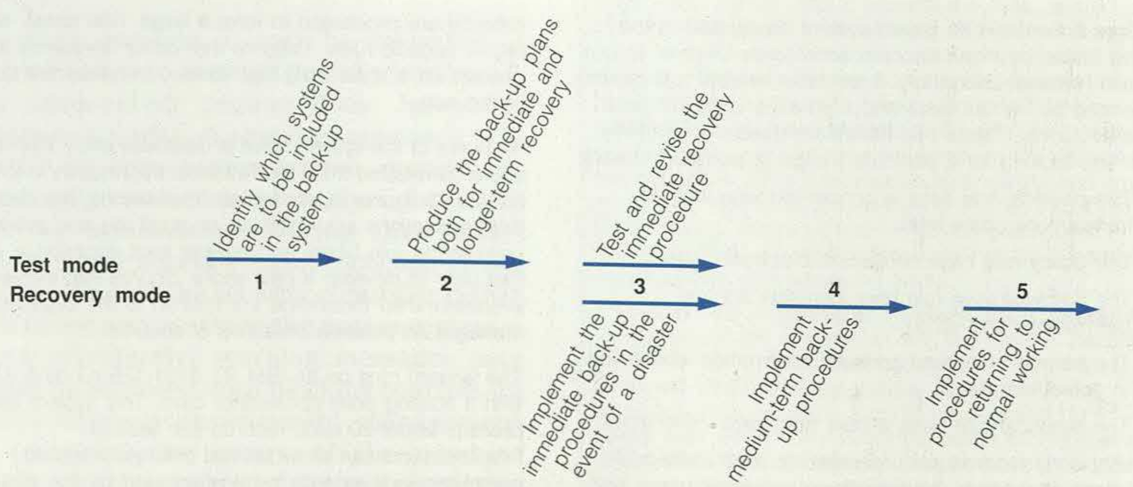

Identify which systems are to be included in the back-up system
Produce the back-up plans both for immediate and longer-term recovery
Test and revise the immediate recovery procedure
Test mode
Recovery mode
1
2
3
4
5
Implement the immediate back-up procedures in the event of a disaster
Implement medium-term back-up procedures
Implement procedures for returning to normal working

Figure 6.4 A major disaster will mean that back-up facilities will be required for an extended period

C E Wooding

cessors, multiplexors, and so on, come from a variety of suppliers and are more difficult to replace quickly. Efforts should therefore be made to locate computer equipment separately from telecommunications equipment, and perhaps, to locate the telecommunications equipment in two separate rooms to reduce the risk of total loss.

Stage 5: Stage 5 is concerned with returning from the temporary facilities established in Stage 4 to a normal working environment. Physically, this is quite straightforward — it is simply a question of ensuring that the measures taken in Stage 4 do not interfere too seriously with those in Stage 5. The more difficult problem is bringing together systems that have been separated during the back-up process. Major gaps in data files may, for instance, severely affect forecasting and accounting systems. Systems managers will have to strike a balance between, on the one hand, the amount of labour required to return to normal working and the business risk involved, and the likelihood of the event ever occurring, and on the other, the cost of a complete back-up system.

Collection of evidence

To investigate security breaches properly, the investigator must have experience with similar problems so that he is in a position to recognise the likely nature of the breach and to know when further detailed expertise is necessary. Systems managers do not generally have this breadth of experience, particularly if the investigation is likely to result in disciplinary action or a prosecution. In such circumstances, it may be necessary for the investigator and technical experts to appear at a disciplinary hearing or in a court of law. Not only must the expertise and integrity of the investigator and experts be beyond reproach, but they must be able and willing to present evidence and to be cross-examined.

To collect evidence that is adequate and acceptable in a court of law requires detailed legal and technical expertise. In particular, the investigator may require the technical expert to have very extensive experience of the particular computer system and the software involved. This is to ensure that all evidence is collected in an indisputably error-free manner, and that, where necessary, those suspected of breaching security are not alerted. This requirement narrows the field considerably and it may be wiser to let the investigator(s) advise on the choice of technical experts.

Use of experts

There is no compulsory licensing scheme for computer security experts or even for security experts in general. There are experts on encryption, database security, the security of specific operating systems, physical security, and personnel security. The field is so broad that no individual will be skilled in all aspects of security. In practice, therefore, Foundation members should select security experts with caution, choosing reputable people with experience in the relevant industry sector or area of special concern.

In some industry sectors, where security is of a particular concern — banking and finance, for example — there is a well developed security culture, and well known specialists serve that sector. Outside such sectors, the best way of contacting a security expert is through auditors, consultants, suppliers, and trusted industry contacts.

Insurance cover

Insurance is an appropriate means of reducing the impact of a breach of security where the probability of an event is low and the resultant claim is likely to be high. It is also appropriate when the effectiveness of other countermeasures is uncertain or when the countermeasures are more expensive than the insurance premiums.

The insurance of a computer system, in particular, is a specialised business and is subject to its own risks. Computer managers are seldom experts in insurance, general insurance brokers are not computer experts, and neither group is likely to have extensive experience of claims, or even of the losses that lead to claims. It is important that both groups understand the nature and the limitations of the insurance offered. Standard computer policies provide cover for four main risks — physical damage, interruption of business, employee fraud, and third-party fraud, although in the latter case, the insurer may require the identity of the defrauder to be established. This may prove impractical where large networks are involved.

Particular attention should be paid to the question of insurance cover for:

- Reinstatement of data (possibly from paper records).
- Accidental or malicious erasure or corruption of data.
- Existing records being incompatible with new hardware installed as part of the recovery process.
- Software costs (new site licences and so on).
- Long periods of business interruption.
- Failure of utility services.
- Rental of replacement equipment.

As we explained in Chapter 2, the risks to be insured against should be identified by the risk analysis exercise that underpins the systems security policy.

Report conclusion

Security is essential to protect the confidentiality, integrity, and availability of computer systems and information. As businesses become increasingly dependent on their information systems, it becomes more and more critical to protect them from breaches of security. We have seen that while no system can ever be totally secure, there is plenty of scope for most systems managers to improve security within their organisations. There is, however, a limit to what systems managers alone can do. A wider perspective is required, and this must be driven from the top of the organisation. A policy is required to blend all aspects of corporate security into a coherent whole.

Within the framework of a corporate security policy, systems security must be managed like any other aspect of systems activity. A sensible systems security policy depends on developing and constantly improving the reliability of methods of identifying sources of threats, of applying countermeasures to prevent breaches of security, and of detecting, and minimising the impact of, those that do occur. This is a long-term and continuing process; political, business, legal, and cultural changes mean that systems security procedures can never be considered permanent. Systems security must be recognised as an essential and integral element of good systems management.

Appendix A
Zero knowledge systems

One of the most interesting developments in access-control systems and encryption over the last few years has been the development of zero knowledge systems (ZKS), also known (more correctly) as Zero Transfer Systems.

Conventional encryption systems suffer from two particular disadvantages. First, if the same encryption method and key is used for a long time, it becomes easier for someone monitoring communications to decode the information. This is one of the reasons that encryption keys are usually changed regularly. The act of changing keys can, however, cause major administrative difficulties, and it may be practically impossible if very large numbers of users are involved. Imagine, for example, the difficulty of a bank's having to change the PINs of all cash dispenser users.

Conventional access-control systems have to rely on a certain degree of trust. Users of a computer system, for example, have to trust those running it not to steal their passwords with the intention of masquerading as valid users. Frequently, the implied level of trust poses a small risk, but where high-value assets are concerned, or where total trust in those running the computer and communications systems would be inappropriate, a better scheme is required.

ZKS seek to address these disadvantages in two ways — first, by never giving away enough information to enable the encryption scheme to be decoded — hence the term ZKS — and second, by randomly varying the point from which the access-verification process begins.

ZKS applied to access control are frequently based on smartcards, where the necessary processing power and secure keys can be held embedded within the smartcard's chip. Applications proposed for ZKS smartcards have included electronic passports and subscription services, such as information services and pay-TV, as well as high-security access and entry-control systems.

Several variations on ZKS are available. The description below is of the Fiat-Shamir ZKS protocol for a smartcard-based access-control card.

A trusted card-recording centre — possibly a supplier — takes a predetermined user identifier and records it on the card, together with a small number of encrypted codes (secrets) derived from the user identifier. Typically, some 20 or so encrypted secrets will be recorded on the card in such a way that they can never be read from outside the card, but can be read by the card's processing logic.

The card is then issued to the user. How the user identity is registered with the trusted centre or service supplier depends on the application; whether the identifier is 'in clear' or encrypted is not important to this discussion. The trusted centre passes certain mathematical details to the computer centre or service provider to which the user will require access, but these details in no way reveal the contents of the user's card. Indeed, they could be made public knowledge without affecting security.

In use, the user (A) inserts the card into a reader, which is in communication with the computer centre or service provider (B). The reader at A sends the user identifier to B, together with an encrypted random number, which determines the random starting point for the verification. B uses the identity supplied by A, together with the details provided by the card supplier, to derive a series of numbers related to the secret numbers held within A's card. B sends back to A a random number, which defines which of the secrets embedded in A's card B wishes to have included in the next step. A then performs a calculation involving the random number that A first produced, the random number provided by B, and the appropriate secret number encoded within A's card. A passes the results of this calculation back to B. Note that nothing about the individual secrets has left the card, because the secrets have been encoded with a random number that is most unlikely ever to be repeated.

Using the mathematical details provided by the card issuer, B is able to confirm that the encoded random number received from A, and the results of A's calculations, combine to match the numbers derived from the user identity.

Even the security offered by this approach may be insufficient, however. It is just conceivable that A is attempting to mimic a valid smartcard and has managed to guess or has replayed the first calculation response correctly. To avoid this, B can request A to generate a new random starting number, and repeat the process as many times as necessary. Typically, the request/response sequence would be repeated 20 or so times for a high-security application. It is extremely unlikely that A could guess a large number correctly 20 times in a row.

The scheme, as described, does not protect A from B's replaying a given access to itself at some date in the future. B would, however, have to explain how the exact random number sequences came to be repeated, a most unlikely event unless B is acting fraudulently. A is, however, protected from B's repeating a transaction sequence to a third party. This is because B, not knowing the secrets contained in A's card, could not guarantee to respond correctly to multiple random responses from a third party. A is further protected by the fact that fraudulent attempts by B or others will have to have used A's identity. A could then be warned of such fraudulent access attempts.

ZKS are an interesting and emerging area of security. Commercial interests arising from the large potential market for smartcards and the benefits to be gained from patenting the various approaches to achieving ZKS security will ensure intense development in this area.

Appendix B

Development of international standards for systems security

The provision of strong security goes to the heart of systems design. It is appropriate therefore that systems security is an increasingly important topic among the international standards-making bodies. However, in view of the all-pervasive nature of security, responsibility for the many aspects of security standards rests with many committees within the standards organisations. Their work is described below.

The International Standards Organisation (ISO), in conjunction with the International Electrotechnical Commission (IEC), has set up a Joint Technical Committee (JTC-1) to address a wide range of IT standards requirements, including security. There are seven subcommittees (SCs) within ISO/IEC JTC-1 dealing with security-related matters, as shown in Figure A.1. Of these, SC27 has specific responsibility for IT security techniques. Several other ISO and IEC technical committees also have an interest in, or specific responsibility for, security standards and they too are shown in Figure A.1.

The CCITT (International Telephone & Telegraph Consultative Committee) is also heavily involved with security-standards work. CCITT is structured into study groups, one of which (SGVII) is responsible for security standards. CCITT's primary concern has been related to security in wide-area networks and messaging systems, but with the advent of public-messaging and EDI services, its work has become closely linked to that of ISO/IEC.

Within Europe, the European Computer Manufacturers Association (ECMA) has mirrored ISO/IEC and CCITT developments, although ECMA has continued to evolve its own security standards via its technical committee structure, shown in Figure A.2, overleaf. The European Telecommunications Institute (ETSI) and the European Workshop for Open Systems (EWOS) are also active in the security-standards field, and their work is also summarised in Figure A.2.

Also in Europe, the activities of CEN/CENELEC (the standards-making body of the European Commission) have been extended during 1989

Figure A.1 There are seven security-related subcommittees within ISO/IEC JTC-1

SC6	OS lower layers
Subject:	Security at OSI layers three and four
SC17	Identification cards
Subject:	Smartcard security
SC18	Text and office systems
Subjects:	Secure message handling, distributed office automation security, ODA (open document architecture) security
SC21	OSI architecture, management, and upper layers
Subjects:	OSI security architecture and open systems frameworks, databases, management and directories, FTAM and TP security, upper layer security, ODP security
SC22	Languages
Subjects:	Posix security
SC27	IT security techniques
Subjects:	Cryptographic and non-cryptographic techniques/mechanisms, and supporting security-related functions, including authentication, integrity, non-repudiation, modes of operation, access control, and registration of algorithms.
SC14	Representation of data elements
Subject:	EDI security

Several other ISO and IEC technical committees also have an interest in or responsibility for security:

- ISO/TC68 (Banking) is responsible for message authentication, key management, PIN management, and other financial-transaction security matters.
- ISO/TC46 has an interest in information security for library systems.
- ISO/TC154 has an interest in EDI security, ISO/TC184 an interest in security related to industrial automation, and IEC/TC65 an interest in security in Safety Related Control Systems.

and 1990 to address security-standards issues. The immediate objective of CEN/CENELEC is to focus effort on the harmonisation of ISO/IEC, CCITT, and ECMA work, in order to meet the Commission's demand for specific European standards. In this context, an Ad Hoc Group on Security has been set up to recommend how to establish a comprehensive set of European security standards based on available international standards and European pre-standards (ENVs), where no appropriate standard already exists. The work carried out by ETSI and EWOS will be incorporated into the CEN/CENELEC programme, as shown in Figure A.3.

Figure A.2 ECMA and ETSI have several technical committees working on security standards

ECMA

TC22	Security of database systems
TC29	ODA security
TC32	ISDN security, OSI lower layer security, security framework, security protocols, data elements and services, and so forth.

ETSI

TC/TE	Smartcard terminals
TC/GSM	Mobile services

In addition, there is a joint ETSI/EWOS group working on an X.400 MHS security profile.

Figure A.3 CEN/CENELEC is recommending how to incorporate the work of the various standards bodies into a coordinated programme

	Type of standards work						
Standards body	Architectures	Frameworks	Models	Service and protocol extensions	Techniques/mechanisms	Applications	Management
ISO/IEC JTC-1							
SC6			✓	✓			
SC17				✓	✓		
SC18			✓			✓	
SC21	✓	✓	✓	✓		✓	✓
SC22					✓		
SC27					✓		
ISO TC 68		✓		✓	✓	✓	
CCITT SGVII	✓	✓	✓	✓	✓	✓	✓
ECMA		✓		✓		✓	✓
ETSI				✓	✓	✓	
ETSI/EWOS						✓	

✓ Responsibility for standards

Security-related standards work is also being carried out within European research projects such as RACE, ESPRIT, COST, TEDIS, and EUREKA, and the various national standards bodies, such as AFNOR, BSI, and DIN. Industry-specific security standards are also being developed by organisations such as SWIFT (for banking).

Other relevant work is being carried in the United States by the American National Standards Institute (ANSI), the American Bankers Association (ABA), the National Institute for Standards and Technology (NIST), and the Institute of Electrical & Electronic Engineers (IEEE). The first three are particularly concerned with the maintenance of DES and its uses, with message authentication and key management, and with EDI security. NIST is also active in the OSI security arena. The IEEE is concerned with the security of local area networks, and with Posix security.

In summary, a great deal of work is being done to create systems security standards. Much of this is an essential first step to the wider use of open networking and the more extensive exploitation of integrated services. Overall, there appear to be four main thrusts to all this work:

— CCITT is seeking to provide standards for secure X.400 (message-handling) services and X.500 (directory) services, to allow the PTTs to provide data-messaging facilities and electronic data interchange (EDI) services.

— ISO is concentrating on vertical protocol frameworks that enable application services to be implemented. This follows the success that earlier protocol frameworks had in making open-standard local area networks viable in the commercial marketplace.

— In the longer term, the evolution of OSI will lead to 'open distributed processing (ODP)'. It is recognised that this is still some way off, and that security forms only part of the much larger task of implementing usable distributed processing across disparate multi-vendor hardware and software environments.

— Complementing the technical developments in standards work, there is work going on within the European Community (Directorate General XIII) and within CEN/CENELEC to determine the need for protective legislation that will encompass such issues as IT security and EDI.

Bibliography

Journals

2600 magazine: the hacker quarterly. Available from: PO Box 752, Middle Island, NY 11953, USA.

Computer fraud & security bulletin. Oxford: Elsevier.

The computer law and security report: the bi-monthly report on computer security and the law governing information technology and computer use. Portsmouth: Solent Legal Exchange.

Computers & security. Oxford: Elsevier.

EDPACS: the EDP audit, control and security newsletter. Boston, Massachusetts: Auerbach.

Information security monitor. London: IBC Technical Services.

Books

Beker, H, and Piper, F. *Cipher systems: the protection of communications*. London: Northwood, 1982.

The complete computer virus handbook. London: Pitman.

Davies, D W, and Price, W L. *Security for computer networks: an introduction to data security in teleprocessing and electronic funds transfer*. 2nd edition. Chichester: Wiley, 1989.

Hruska, J, and Jackson, K. *The PC security guide*. 2nd edition. Oxford: Elsevier, 1990.

Meyer, C H, and Matyas, S M. *Cryptography: a new dimension in computer data security: a guide for the design and implementation of secure systems*. Chichester: Wiley, 1982.

Robert, D W. *Computer security*. London: Blenheim Online, 1990.

Smith, M R. *Commonsense computer security: your practical guide to preventing accidental and deliberate electronic data loss*. London: McGraw-Hill, 1989.

Stoll, C. *The cuckoo's egg: tracking a spy through a maze of computer espionage*. London: Bodley Head, 1990.

United Kingdom. Law Commission. *Computer misuse*. (Working Paper No 110.) London: HMSO, 1988.

The Butler Cox Foundation

The Butler Cox Foundation is a service for senior managers responsible for information management in major enterprises. It provides insight and guidance to help them to manage information systems and technology more effectively for the benefit of their organisations.

The Foundation carries out a programme of syndicated research that focuses on the business implications of information systems, and on the management of the information systems function, rather than on the technology itself. It distributes a range of publications to its members that includes Research Reports, Management Summaries, Directors' Briefings, and Position Papers. It also arranges events at which members can meet and exchange views, such as conferences, management briefings, research reviews, study tours, and specialist forums.

Membership of the Foundation

The Foundation is the world's leading programme of its type. The majority of subscribers are large organisations seeking to exploit to the full the most recent developments in information technology. The membership is international, with more than 400 organisations from over 20 countries, drawn from all sectors of commerce, industry, and government. This gives the Foundation a unique capability to identify and communicate 'best practice' between industry sectors, between countries, and between IT suppliers and users.

Benefits of membership

The list of members establishes the Foundation as the largest and most prestigious 'club' for systems managers anywhere in the world. Members have commented on the following benefits:

— The publications are terse, thought-provoking, informative, and easy to read. They deliver a lot of message in a minimum of precious reading time.

— The events combine access to the world's leading thinkers and practitioners with the opportunity to meet and exchange views with professional counterparts from different industries and countries.

— The Foundation represents a network of systems practitioners, with the power to connect individuals with common concerns.

Combined with the manager's own creativity and business knowledge, Foundation membership contributes to managerial success.

Recent Research Reports

57 Using System Development Methods
58 Senior Management IT Education
59 Electronic Data Interchange
60 Expert Systems in Business
61 Competitive-Edge Applications: Myths and Reality
62 Communications Infrastructure for Buildings
63 The Future of the Personal Workstation
64 Managing the Evolution of Corporate Databases
65 Network Management
66 Marketing the Systems Department
67 Computer-Aided Software Engineering (CASE)
68 Mobile Communications
69 Software Strategy
70 Electronic Document Management
71 Staffing the Systems Function
72 Managing Multivendor Environments
73 Emerging Technologies: Annual Review for Managers
74 The Future of System Development Tools
75 Getting Value from Information Technology
76 Systems Security

Recent Position Papers and Directors' Briefings

Information Technology and Realpolitik
The Changing Information Industry: An Investment Banker's View
A Progress Report on New Technologies
Hypertext
1992: An Avoidable Crisis
Managing Information Systems in a Decentralised Business
Pan-European Communications: Threats and Opportunities
Information Centres in the 1990s
Open Systems

Forthcoming Research Reports

New Telecommunications Services
The Role of IT in Transforming the Organisation
Electronic Marketplaces
Managing the Distribution of IT
The Future of Electronic Mail

Butler Cox

The Butler Cox Foundation is one of the services provided by the Butler Cox Group. Butler Cox is an independent international consulting company specialising in areas relating to information technology. Its services include management consulting, applied research, and education.

Butler Cox plc
Butler Cox House, 12 Bloomsbury Square,
London WC1A 2LL, England
☎ (071) 831 0101, Telex 8813717 BUTCOX G
Fax (071) 831 6250

Belgium and the Netherlands
Butler Cox Benelux bv
Prins Hendriklaan 52,
1075 BE Amsterdam, The Netherlands
☎ (020) 75 51 11, Fax (020) 75 53 31

France
Butler Cox SARL
Tour Akzo, 164 Rue Ambroise Croizat,
93204 St Denis-Cédex 1, France
☎ (1) 48.20.61.64, Télécopieur (1) 48.20.72.58

Germany (FR), Austria, and Switzerland
Butler Cox GmbH
Richard-Wagner-Str. 13, 8000 München 2, West German
☎ (089) 5 23 40 01, Fax (089) 5 23 35 15

Australia and New Zealand
Mr J Cooper
Butler Cox Foundation
Level 10, 70 Pitt Street, Sydney, NSW 2000, Australia
☎ (02) 223 6922, Fax (02) 223 6997

Finland
TT-Innovation Oy
Meritullinkatu 33, SF-00170 Helsinki, Finland
☎ (90) 135 1533, Fax (90) 135 2985

Ireland
SD Consulting
72 Merrion Square, Dublin 2, Ireland
☎ (01) 766088/762501, Telex 31077 EI,
Fax (01) 767945

Italy
RSO Futura Srl
Via Leopardi 1, 20123 Milano, Italy
☎ (02) 720 00 583, Fax (02) 806 800

Scandinavia
Butler Cox Foundation Scandinavia AB
Jungfrudansen 21, Box 4040, 171 04 Solna, Sweden
☎ (08) 730 03 00, Fax (08) 730 15 67

Spain and Portugal
T Network SA
Núñez Morgado 3-6°b, 28036 Madrid, Spain
☎ (91) 733 9866, Fax (91) 733 9910