CHAPTER 7

Economics

The great fortunes of the information age lie in the hands of companies that have established proprietary architectures that are used by a large installed base of locked-in customers.

— Carl Shapiro and Hal Varian

There are two things I am sure of after all these years: there is a growing societal need for high assurance software, and market forces are never going to provide it.

— Earl Boebert

If you try to buck the markets, then the markets will buck you.

— Margaret Thatcher

7.1 Introduction

The economics of information security has recently become a thriving and fast-moving discipline. We started to realise round about 2000 that many security system failures weren't due to technical errors so much as to wrong incentives: the classic case is where the people who guard a system are not the people who suffer when it fails. Indeed, security mechanisms are often designed quite deliberately to shift liability, which often leads to trouble.

Economics has always been important to engineering, at the raw level of cost accounting; a good engineer was one who could build a bridge safely with a thousand tons of concrete when everyone else used two thousand tons. But the perverse incentives that arise in complex systems with multiple owners make economic questions both more important and more subtle for the security engineer. Truly global-scale systems like the Internet arise from the actions of millions of independent principals with divergent interests; we hope that reasonable global outcomes will result from selfish local actions. In general, people won't do something unless they have an incentive to. Markets are often the best guide we have to what sort of mechanisms work, or fail. Markets also fail; the computer industry has been dogged by monopolies since its earliest days. The reasons for this are now understood, and their interaction with security is starting to be. When someone asks 'Why is Microsoft software insecure?' we can now give a principled answer rather than simply cursing Redmond as a form of bad weather.

The new field of security economics provides valuable insights not just into 'security' topics such as privacy, bugs, spam, and phishing, but into more general areas such as system dependability. For example, what's the optimal balance of effort by programmers and testers? (For the answer, see section 7.5.1 below.) It also enables us to analyse the policy problems that security technology increasingly throws up — on issues like digital rights management. Where protection mechanisms are used by the system designer to control the owner of a machine, rather than to protect her against outside enemies, questions of competition policy and consumer rights follow, which economics provides the language to discuss. There are also questions of the balance between public and private action. Network insecurity is somewhat like air pollution or congestion, in that people who connect insecure machines to the Internet do not bear the full consequences of their actions. So how much of the protection effort can (or should) be left to individuals, and how much should be borne by vendors, regulators or the police?

7.2 Classical Economics

Modern economics is an enormous field covering many different aspects of human behaviour. The parts of it that (so far) have found application in security are largely drawn from microeconomics and game theory. I'll discuss game theory in the next section; here, I'll give a helicopter tour of the most relevant ideas from microeconomics. The object of the exercise is not to provide a tutorial on economics, or even on information economics — for that, I recommend you read Carl Shapiro and Hal Varian's book 'Information Rules' [1159] — but to familiarise you with the essential terminology and ideas, so we can move on to discuss security economics.

The modern subject started in the 18th century when the industrial revolution and growing trade changed the world, and people wanted to understand why. In 1776, Adam Smith's classic *'The Wealth of Nations'* [1192] provided a first draft: he explained how rational self-interest in a free market economy leads to economic wellbeing. Specialisation leads to productivity gains at all levels from a small factory to international trade, and the self-interested striving of many individuals and firms drives progress, as people must produce something others value to survive in a competitive market. In his famous phrase, 'It is not from the benevolence of the butcher, the brewer, or the baker, that we can expect our dinner, but from their regard to their own interest'.

These ideas were refined by nineteenth-century economists; David Ricardo clarified and strengthened Smith's arguments in favour of free trade, Stanley Jevons, Léon Walras and Carl Menger built detailed models of supply and demand, and by the end of the century Alfred Marshall had combined models of supply and demand in markets for goods, labour and capital into an overarching 'classical' model in which, at equilibrium, all the excess profits would be competed away and the economy would be functioning efficiently. By 1948, Kenneth Arrow and Gérard Debreu had put this on a rigorous mathematical foundation by proving that markets give efficient outcomes, subject to certain conditions. Much of the interest in economics — especially to the computer industry, and to security folks in particular — comes from the circumstances in which these conditions aren't met, giving rise to monopolies and other problems.

7.2.1 Monopoly

A rapid way into the subject is to consider a simple textbook case of monopoly. Suppose we have a market for apartments in a university town, and the students have different incomes. We might have one rich student able to pay $4000 a month, maybe 300 people willing to pay $2000 a month, and (to give us round numbers) 1000 prepared to pay $1000 a month. That gives us the *demand curve* shown in Figure 7.1 below.

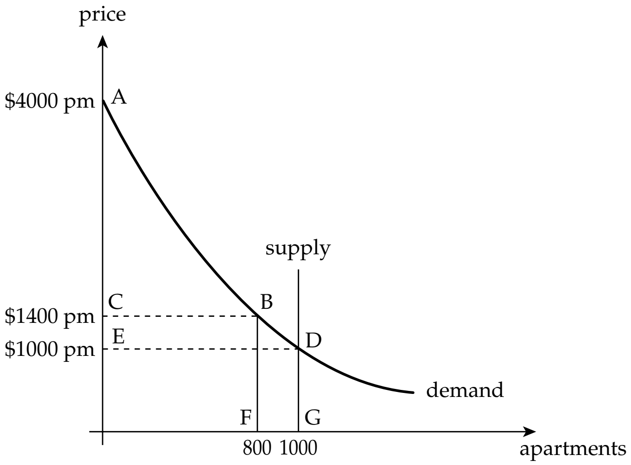

Figure 7.1: The market for apartments

So if there are 1000 apartments being let by many competing landlords, the market-clearing price will be at the intersection of the demand curve with the vertical supply curve, namely $1000. But suppose the market is rigged — say the landlords have set up a cartel, or the university makes its students rent through a tied agency. For simplicity let's assume a single monopolist landlord. He examines the demand curve, and notices that if he rents out only 800 apartments, he can get $1400 per month for each of them. Now 800 times $1400 is $1,120,000 per month, which is more than the million dollars a month he'll make from the market price at $1000. (Economists would say that his 'revenue box' is CBFO rather than EDGO.) So he sets an artificially high price, and 200 apartments remain empty.

This is clearly inefficient, and the Italian economist Vilfredo Pareto invented a neat way to formalise this. A *Pareto improvement* is any change that would make some people better off without making anyone else worse off, and an allocation is *Pareto efficient* if there isn't any Pareto improvement available. Here, the allocation is not efficient, as the monopolist could rent out one empty apartment to anyone at a lower price, making both him and them better off. Now Pareto efficiency is a rather weak criterion; both perfect communism (everyone gets the same income) and perfect dictatorship (the President gets all the income) are Pareto-efficient. In neither case can you make anyone better off without making someone worse off! Yet the simple monopoly described here is not efficient even in this very weak sense.

So what can the monopolist do? There is one possibility — if he can charge everyone a different price, then he can set each student's rent at exactly what they are prepared to pay. We call such a landlord a *discriminating monopolist*; he charges the rich student exactly $4000, and so on down to the 1000th student whom he charges exactly $1000. The same students get apartments as before, yet almost all of them are worse off. The rich student loses $3000, money that he was prepared to pay but previously didn't have to; economists refer to this money he saved as *surplus*. In effect, the discriminating monopolist manages to extract all the consumer surplus.

Merchants have tried to price-discriminate since antiquity. The carpet seller in Damascus who offers to 'make a very special price, just for you' is playing this game, as is Microsoft in offering seven different versions of Vista at different price points, and an airline in selling first, business and cattle class seats. The extent to which firms can do this depends on a number of factors, principally their market power and the amount of information they have. Market power is a measure of how close a merchant is to being a monopolist; under monopoly the merchant is a *price setter* while under perfect competition he simply has to accept whatever price the market establishes (he is a *price taker*). Technology tends to increase market power while reducing the cost of information about customers at the same time, and this combination is one of the main factors eroding privacy in the modern world.

7.2.2 Public Goods

A second type of market failure occurs when everyone gets the same quantity of some good, whether they want it or not. Classic examples are air quality, national defense and scientific research. Economists call these *public goods*, and the formal definition is that they are goods which are non-rivalrous (my using them doesn't mean there's less available for you) and non-excludable (there's no practical way to exclude people from consuming them). Uncoordinated markets are generally unable to provide public goods in socially optimal quantities.

Public goods may be supplied by governments directly, as in the case of national defense, or by using indirect mechanisms to coordinate markets. The classic example is laws on patents and copyrights to encourage people to produce inventions, literary works and musical compositions by giving them a temporary monopoly — by making the goods in question excludable for a limited period of time. Very often, public goods are provided by some mix of public and private action; scientific research is done in universities that get some public subsidy, earn some income from student fees, and get some research contracts from industry (where industry may get patents on the useful inventions while the underlying scientific research gets published for all to use). The mix can be controversial; the debate on global warming sets people who want direct government action in the form of a 'carbon tax' (which would be simple and easy to enforce) against others who want a 'cap and trade' system whereby firms and countries can trade licenses to emit carbon (which in a perfect world would cause emission reductions by the firms who could do so most cheaply, but which might well be more open to abuse and evasion).

The importance of this for us is that many aspects of security are public goods. I do not have an anti-aircraft gun on the roof of my house; air-defense threats come from a small number of actors, and are most efficiently dealt with by government action. So what about Internet security? Certainly there are strong externalities involved, and people who connect insecure machines to the Internet end up dumping costs on others, just like people who burn polluting coal fires. So what should we do about it? One might imagine a government tax on vulnerabilities, with rewards paid to researchers who discover them and larger fines imposed on the firms whose software contained them. Again, one of the early papers on security economics suggested a vulnerability cap-and-trade system; vendors who could not be bothered to make their software secure could buy permits from other vendors who were making the effort to tighten up their products [256]. (Both arrangements would be resisted by the free software community!) But is air pollution the right analogy — or air defense?

Threats such as viruses and spam used to come from a large number of small actors, but since about 2004 we've seen a lot of consolidation as malware writers and users have become commercial. By 2007, the number of serious spammers had dropped to the point that ISPs see significant fluctuations in overall spam volumes as the big spammers run particular campaigns — there is no law of large numbers operating any more [305]. This suggests a different and perhaps more centralised strategy. If our air-defense threat in 1987 was mainly the Russian airforce, and our cyber-defense threat in 2007 is mainly from a small number of Russian gangs, and they are imposing large costs on US and European Internet users and companies, then state action may be needed now as it was then. Instead of telling us to buy anti-virus software, our governments could be putting pressure on the Russians to round up and jail their cyber-gangsters. I'll discuss this in greater detail in Part III; for now, it should be clear that concepts such as 'monopoly' and 'public goods' are important to the security engineer — and indeed to everyone who works in IT. Just think of the two operating systems that dominate the world's desktops and server farms: Windows is a monopoly, while the common Unix systems (Linux and OpenBSD) are public goods maintained by volunteers. Why should this be so? Why are markets for information goods and services so odd?

7.3 Information Economics

One of the insights from the nineteenth-century economists Jevons and Menger is that the price of a good, at equilibrium, is the marginal cost of production. When coal cost nine shillings a ton in 1870, that didn't mean that every mine dug coal at this price, merely that the marginal producers — those who were only just managing to stay in business — could sell at that price. If the price went down, these mines would close; if it went up, other, even more marginal mines, would open. That's how supply responded to changes in demand.

7.3.1 The Price of Information

So in a competitive equilibrium, the price of information should be its marginal cost of production. But that is almost zero! This explains why there is so much information available for free in the Internet; zero is its fair price. If two or more suppliers compete to offer an operating system, or a map, or an encyclopaedia, that they can duplicate for no cost, then the incentive will be for them to keep on cutting their prices without limit. This is what happened with encyclopaedias; the Britannica used to cost $1,600 for 32 volumes; then Microsoft brought out Encarta for $49.95, forcing Britannica to produce a cheap CD edition; and now we have Wikipedia for free [1159]. One firm after another has had to move to a business model in which the goods are given away free, and the money comes from advertising or in some parallel market.

Linux companies give away an operating system, and make their money from support; many Linux developers give their time free to the project while at college, as their contribution strengthens their CV and helps them get a good job when they graduate.

Many other industries with high fixed costs and low marginal costs moved to an advertising or service model; think terrestrial TV. Others have moved in this direction: most newspapers made most of their money from advertising, and so have had little difficulty moving to free online editions plus paid paper editions, all putting the lucrative ads in front of eyeballs. Yet other industries, such as airlines and hotels, tended instead to become monopolists who try to dominate particular routes or areas and to charge different prices to different customers.

So what other characteristics of the information goods and services industries are particularly important?

1. There are often *network externalities*, whereby the value of a network grows more than linearly in the number of users. For example, the more people used fax machines in the 1980s, the more useful they became, until round about 1985 fax machines suddenly took off; every business needed one. Much the same happened with email in about 1995. Network effects also apply to services more generally: anyone wanting to auction some goods will usually go to the largest auction house, as it will attract more bidders. They also apply to software: firms develop software for Windows so they will have access to more users than they would if they developed for Linux or Mac, and users for their part prefer Windows because there's more software for it. (This is called a *two-sided market*.)
2. There is often technical lock-in stemming from interoperability. Once a software firm is committed to using Windows as a platform for its product, it can be expensive to change; for users, too, changing platforms can be expensive. They have to buy new software, convert files (if they can), and retrain themselves.

These features separately can lead to industries with dominant firms; together, they are even more likely to. If users simply want to be compatible with other users (and software vendors) then they will logically buy from the vendor they expect to win the biggest market share.

7.3.2 The Value of Lock-In

There is an interesting result, due to Shapiro and Varian: that the value of a software company is the total lock-in (due to both technical and network effects) of all its customers [1159]. To see how this might work, consider a firm with 100 staff each using Office, for which it has paid $500 per copy. It could save this $50,000 by moving to OpenOffice, so if the costs of installing this product, retraining its staff, converting files and so on — in other words the total switching costs — were less than $50,000, it would switch. But if the costs of switching were more than $50,000, then Microsoft would put up its prices.

Technical lock-in existed before, but the move to the digital economy has made it much more significant. If you own a Volvo car, you are locked in to Volvo spares and accessories; but when you're fed up with it you can always trade it in for a Mercedes. But if you own an Apple Mac, you'll have Mac software, a Mac printer, and quite possibly thousands of music tracks that you've ripped to iTunes. You'd also have to learn to use different commands and interfaces. Moving to Windows would be much more painful than just shelling out $700 for a new laptop. You'd have to retrain yourself; you'd have to throw away Office for Mac and buy Office for Windows. And if you'd bought a lot of music tracks from the iTunes music store, it could be more painful still (you'd probably decide to keep your iPod with your new Windows machine rather than moving to a Windows music player, even though the iPod works better with the Mac). This shows why lock-in can be so durable; although each piece of equipment — be it a Mac laptop, an iPod, or a printer — wears out, the lock-in persists in the complementary relationship between them. And this doesn't just apply to PC platforms, but to ISPs; commercial software systems such as databases; equipment such as telephone exchanges; and various online services.

This is why so much effort gets expended in standards wars and antitrust suits. It's also why so many security mechanisms now aim at controlling compatibility. In such cases, the likely hackers are not malicious outsiders, but the owners of the equipment, or new firms trying to challenge the incumbent by making compatible products. The issues are made more complex by the fact that innovation is often incremental, and products succeed when new firms find killer applications for them [607]. The PC, for example, was designed by IBM as a machine to run spreadsheets; if they had locked it down to this application alone, then a massive opportunity would have been lost. Indeed, the fact that the IBM PC was more open than the Apple Mac was a factor in its becoming the dominant desktop platform.

So the law in many countries gives companies a right to reverse-engineer their competitors' products for compatibility [1110]. More and more, security mechanisms are being used to try to circumvent that law: incumbents try to lock down their platforms using cryptography and tamper-resistance so that even if competitors have the legal right to try to reverse engineer them, they are not always going to succeed in practice. Many businesses are seeing brutal power struggles for control of the supply chain; for example, mobile phone makers' attempts to introduce sophisticated DRM into their handsets were frustrated by network operators determined to prevent the handset makers from establishing business relationships directly with their customers. These struggles set the scene in which more and more security products succeed or fail.

7.3.3 Asymmetric Information

Another of the ways in which markets can fail, beyond monopoly and public goods, is when some principals know more than others. The study of *asymmetric information* was kicked off by a famous paper in 1970 on the 'market for lemons' [19], for which George Akerlof won a Nobel prize. It presents the following simple yet profound insight: suppose that there are 100 used cars for sale in a town: 50 well-maintained cars worth $2000 each, and 50 'lemons' worth $1000. The sellers know which is which, but the buyers don't. What is the market price of a used car? You might think $1500; but at that price no good cars will be offered for sale. So the market price will be close to $1000. This is one reason poor security products predominate. When users can't tell good from bad, they might as well buy a cheap antivirus product for $10 as a better one for $20, and we may expect a race to the bottom on price.

A further distinction can be drawn between hidden information and hidden action. For example, Volvo has a reputation for building safe cars that survive accidents well, yet it is well known that Volvo drivers have more accidents. Is this because people who know they're bad drivers buy Volvos so they're less likely to get killed, or because people in Volvos drive faster? The first is the hidden-information case, also known as *adverse selection*, and the second is the hidden-action case, also known as *moral hazard*. Both effects are important in security, and both may combine in specific cases. (In the case of drivers, there seems to be a growing consensus that people adjust their driving behaviour to keep their risk exposure to the level with which they are comfortable. This also explains why mandatory seat-belt laws tend not to save lives overall, merely to move fatalities from vehicle occupants to pedestrians and cyclists [10].)

Asymmetric information explains many market failures in the real world, from low prices in used-car markets to the difficulty that older people have in getting insurance on reasonable terms (people who know they're sick will tend to buy more of it, making it uneconomic for the healthy). It tends to lead to surveillance or rationing.

7.4 Game Theory

There are really just two ways to get something you want if you can't find or make it yourself. You either make something useful and trade it; or you take what you need, by force, by the ballot box or whatever. Choices between cooperation and conflict are made every day at all sorts of levels, by both humans and animals.

The main tool we can use to study and analyse them is game theory, which I will define as 'the study of problems of cooperation and conflict among independent decision makers'. We're interested in games of strategy rather than games of chance, and we're less interested in games of perfect information (such as chess) than in games of imperfect information, which can be much more interesting. We try to get to the core of games by abstracting away much of the detail. For example, consider the school playground game of 'matching pennies': Alice and Bob toss coins and reveal them simultaneously, upon which Alice gets Bob's penny if they're different and Bob gets Alice's penny if they're the same. I'll write this as shown in Figure 7.2:

		Bob	
		H	T
Alice	H	−1,1	1,−1
	T	1,-1	−1,1

Figure 7.2: Matching pennies

Each entry in the table shows first Alice's outcome and then Bob's outcome. Thus if the coins fall (H,H) Alice loses a penny and Bob gains a penny. This is an example of a *zero-sum game*: Alice's gain is Bob's loss.

Often we can solve a game quickly by writing out a *payoff matrix* like this. Here's an example (Figure 7.3):

		Bob	
		Left	Right
Alice	Top	1,2	0,1
	Bottom	2,1	1,0

Figure 7.3: Dominant strategy equilibrium

In this game, no matter what Bob plays, Alice is better off playing 'Bottom'; and no matter what Alice plays, Bob is better off playing 'Left'. Each player has a *dominant strategy* — an optimal choice regardless of what the other does. So Alice's strategy should be a constant 'Bottom' and Bob's a constant 'Left'. (A *strategy* in game theory is just an algorithm that takes a game state and outputs a move.) We call this a *dominant strategy equilibrium*.

Another example is shown in Figure 7.4:

		Bob	
		Left	Right
Alice	Top	2,1	0,0
	Bottom	0,0	1,2

Figure 7.4: Nash equilibrium

Here each player's optimal strategy depends on what the other player does, or (perhaps more accurately) what they think the other player will do. We say that two strategies are in Nash equilibrium when Alice's choice is optimal given Bob's, and vice versa. Here there are two symmetric Nash equilibria, at top left and bottom right. You can think of them as being like local optima while a dominant strategy equilibrium is a global optimum.

7.4.1 The Prisoners' Dilemma

We're now ready to look at a famous problem posed in 1950 by John Nash, and for which he won the Nobel. It applies to many situations from international trade negotiations to free-riding in peer-to-peer file-sharing systems to cooperation between hunting animals, and Nash first studied it in the context of US and USSR defense spending; his employer, the Rand corporation, was paid to think about possible strategies in nuclear war. However, Nash presented it using the following simple example.

Two prisoners are arrested on suspicion of planning a bank robbery. The police interview them separately and tell each of them the following: ''If neither of you confesses you'll each get a year for carrying a concealed firearm without a permit. If one of you confesses, he'll go free and the other will get 6 years for conspiracy to rob. If both of you confess, you will each get three years.''

What should the prisoners do? Let's write the game out formally, as shown in Figure 7.5:

		Benjy	
		Confess	Deny
Alfie	Confess	−3,−3	0,−6
	Deny	−6,0	−1,−1

Figure 7.5: The prisoners' dilemma

When Alfie looks at this table, he will reason as follows: 'If Benjy's going to confess then I should too as then I get 3 years rather than 6; and if he's going to deny then I should still confess as that way I walk rather than doing a year'. Benjy will reason similarly. The two of them will each confess, and get three years each. This is not just a Nash equilibrium; it's a dominant strategy equilibrium. Each prisoner should logically confess regardless of what the other does.

But hang on, you say, if they had agreed to keep quiet then they'll get a year each, which is a better outcome for them! In fact the strategy (deny,deny) is Pareto efficient, while the dominant strategy equilibrium is not. (That's one reason it's useful to have concepts like 'Pareto efficient' and 'dominant strategy equilibrium' rather than just arguing over 'best'.)

So what's the solution? Well, so long as the game is going to be played once only, and this is the only game in town, there isn't a solution. Both prisoners will logically confess and get three years. We can only change this state of affairs if somehow we can change the game itself. There are many possibilities: there can be laws of various kinds from international treaties on trade to the gangster's *omertá*. In practice, a prisoner's dilemma game is changed by altering the rules or the context so as to turn it into another game where the equilibrium is more efficient.

7.4.2 Evolutionary Games

An important class of problems can be solved where the game is played repeatedly — if Alfie and Benjy are career criminals who expect to be dealing with each other again and again. Then of course there can be an incentive for them to cooperate. There are at least two ways of modelling this.

In the 1970s, Bob Axelrod started thinking about how people might play many rounds of prisoners' dilemma. He set up a series of competitions to which people could submit programs, and these programs played each other repeatedly in tournaments. He found that one of the best strategies overall was *tit-for-tat*, which is simply that you cooperate in round one, and at each subsequent round you do to your opponent what he or she did in the previous round [99]. It began to be realised that strategy evolution could explain a lot. For example, in the presence of noise, players tend to get locked into (defect, defect) whenever one player's cooperative behaviour is misread by the other as defection. So in this case it helps to 'forgive' the other player from time to time.

Simultaneously, a parallel approach was opened up by John Maynard Smith and George Price [848]. They considered what would happen if you had a mixed population of aggressive and docile individuals, 'hawks' and 'doves', with the behaviour that doves cooperate; hawks take food from doves; and hawks fight, with a risk of death. Suppose the value of the food at each interaction is V and the risk of death in a hawk fight per encounter is C. Then the payoff matrix looks like Figure 7.6:

	Hawk	Dove
Hawk	$\frac{V-C}{2}, \frac{V-C}{2}$	V,0
Dove	0, V	$\frac{V}{2}, \frac{V}{2}$

Figure 7.6: The hawk-dove game

Here, if $V > C$, the whole population will become hawk, as that's the dominant strategy, but if $C > V$ (fighting is too expensive) then there is an equilibrium where the probability p that a bird is a hawk sets the hawk payoff and the dove payoff equal, that is

$$p\frac{V-C}{2} + (1-p)V = (1-p)\frac{V}{2}$$

which is solved by $p = V/C$. In other words, you can have aggressive and docile individuals coexisting in a population, and the proportion of aggressive individuals will at equilibrium be a function of the costs of aggression; the more dangerous it is, the fewer such individuals there will be. Of course, the costs can change over time, and diversity is a good thing in evolutionary terms as a society with a minority of combative individuals may be at an advantage when war breaks out. Again, it takes generations for a society to move to equilibrium. Perhaps our current incidence of aggression is too high because it reflects conditions in the Dark Ages, or even on the African highveld 500,000 years ago[1].

This neat insight, along with Bob Axelrod's simulation methodology for tackling problems that don't have a neat algebraic solution, got many people from moral philosophers to students of animal behaviour interested in evolutionary game theory. They give deep insights into how cooperation evolved. It turns out that many primates have an inbuilt sense of fairness and punish individuals who are seen to be cheating — the instinct for vengeance is part of the mechanism to enforce sociality. Fairness can operate in a number of different ways at different levels. For example, the philosopher Brian Skyrms found that doves can get a better result against hawks if they can recognise each other and interact preferentially, giving a model for how social movements such as freemasons and maybe even some religions establish themselves [1188].

[1] A number of leading anthropologists believe that, until recent times, tribal warfare was endemic among human societies [777].

Of course, the basic idea behind tit-for-tat goes back a long way. The Old Testament has ‘An Eye for an eye’ and the New Testament ‘Do unto others as you’d have them do unto you’; the latter formulation is, of course, more fault-tolerant, and versions of it can be found in Aristotle, in Confucius and elsewhere. More recently, Thomas Hobbes used primitive prisoners’-dilemma-style arguments in the seventeenth century to justify the existence of a state without the Divine Right of Kings.

The applications of evolutionary game theory keep on growing. Since 9/11, for example, there has been interest in whether hawk-dove games explain the ability of fundamentalists to take over discourse in religions at a time of stress. From the economists’ viewpoint, evolutionary games explain why cartel-like behaviour can appear in industries even where there are no secret deals being done in smoke-filled rooms. For example, if there are three airlines operating a profitable route, and one lowers its prices to compete for volume, the others may well respond by cutting prices even more sharply to punish it and make the route unprofitable, in the hope that the discounts will be discontinued and everyone can go back to gouging the customer. And there are some interesting applications in security, too, which I’ll come to later.

7.5 The Economics of Security and Dependability

Economists used to be well aware of the interaction between economics and security; rich nations could afford big armies. But nowadays a web search on ‘economics’ and ‘security’ turns up relatively few articles. The main reason is that, after 1945, economists drifted apart from people working on strategic studies; nuclear weapons were thought to decouple national survival from economic power [839]. A secondary factor may have been that the USA confronted the USSR over security, but Japan and the EU over trade. It has been left to the information security world to re-establish the connection.

One of the observations that rekindled interest in security economics came from banking. In the USA, banks are generally liable for the costs of card fraud; when a customer disputes a transaction, the bank must either show she is trying to cheat it, or refund her money. In the UK, banks generally got away with claiming that their systems were ‘secure’, and telling customers who complained that they must be mistaken or lying. ‘Lucky bankers,’ you might think; yet UK banks spent more on security and suffered more fraud. This was probably a moral-hazard effect: UK bank staff knew that customer complaints would not be taken seriously, so they became lazy and careless, leading to an epidemic of fraud [33, 34].

Another was that people were not spending as much money on anti-virus software as the vendors might have hoped. Now a typical virus payload then was a service-denial attack on Microsoft; and while a rational consumer might spend $20 to stop a virus trashing her hard disk, she will be less likely to do so just to protect a wealthy corporation [1290]. There are many other examples, such as hospital systems bought by medical directors and administrators that look after their interests but don't protect patient privacy. The picture that started to emerge was of system security failing because the people guarding a system were not the people who suffered the costs of failure, or of particular types of failure. Sometimes, as we'll see, security mechanisms are used to dump risks on others, and if you are one of these others you'd be better off with an insecure system. Put differently, security is often not a scalar, but a power relationship; the principals who control what it means in a given system often use it to advance their own interests.

This was the initial insight. But once we started studying security economics seriously, we found that there's a lot more to it than that.

7.5.1 Weakest Link, or Sum of Efforts?

The late Jack Hirshleifer, the founder of conflict theory, told the story of Anarchia, an island whose flood defences were constructed by individual families who each maintained a section of the flood wall. The island's flood defence thus depended on the weakest link, that is, the laziest family. He compared this with a city whose defences against missile attack depend on the single best defensive shot [609]. Another example of best-shot is medieval warfare, where there was on occasion a single combat between the two armies' champions. Hal Varian extended this model to three cases of interest to the dependability of information systems — where performance depends on the minimum effort, the best effort, or the sum-of-efforts [1292]. This last case, the sum-of-efforts, is the modern model for warfare: we pay our taxes and the government hires soldiers. It's a lot more efficient than best-shot (where most people will free-ride behind the heroes), and that in turn is miles better than weakest-link (where everyone will free-ride behind the laziest).

Program correctness can depend on minimum effort (the most careless programmer introducing a vulnerability) while software vulnerability testing may depend on the sum of everyone's efforts. Security may also depend on the best effort — the actions taken by an individual champion such as a security architect. As more agents are added, systems become more reliable in the total-effort case but less reliable in the weakest-link case. What are the implications? Well, software companies should hire more software testers and fewer (but more competent) programmers.

7.5.2 Managing the Patching Cycle

There has been much debate about 'open source security', and more generally whether actively seeking and disclosing vulnerabilities is socially desirable.

It's a debate that has flared up again and again; as we saw in the preface, the Victorians agonised over whether it was socially responsible to publish books about lockpicking, and eventually concluded that it was [1257]. People have worried more recently about the online availability of (for example) the US Army Improvised Munitions Handbook) [1271]; is the risk of helping terrorists sufficient to justify online censorship?

Security economics provides both a theoretical and a quantitative framework for discussing some issues of this kind. I showed in 2002 that, under standard assumptions of reliability growth, open systems and proprietary systems are just as secure as each other; opening up a system helps the attackers and defenders equally [54]. Thus the open-security question will often be an empirical one, turning on the extent to which a given real system follows the standard model.

In 2004, Eric Rescorla argued that for software with many latent vulnerabilities, removing one bug makes little difference to the likelihood of an attacker finding another one later. Since exploits are often based on vulnerabilities inferred from patches, he argued against disclosure and frequent patching unless the same vulnerabilities are likely to be rediscovered [1071]. Ashish Arora and others responded with data showing that public disclosure made vendors respond with fixes more quickly; attacks increased to begin with, but reported vulnerabilities declined over time [88]. In 2006, Andy Ozment and Stuart Schechter found that the rate at which unique vulnerabilities were disclosed for the core OpenBSD operating system has decreased over a six-year period [998]. These results support the current system of responsible disclosure whereby people who discover vulnerabilities report them to CERT, which reports them on to vendors, and publicises them once patches are shipped.

This is by no means all that there is to say about the economics of dependability. There are tensions between vendors and their customers over the frequency and timing of patch release; issues with complementers; difficulties with metrics; companies such as iDefense and TippingPoint that buy and sell information on vulnerabilities; and even concerns that intelligence agencies with privileged access to bug reports use them for zero-day exploits against other countries' systems. I'll come back to all this in Part III.

7.5.3 Why Is Windows So Insecure?

The micromanagement of the patching cycle begs a deeper question: why are there so many bugs in the first place? In particular, why is Windows so insecure, despite Microsoft's dominant market position? It's possible to write much better software, and there are fields such as defense and healthcare where a serious effort is made to produce dependable systems. Why do we not see a comparable effort made with commodity platforms, especially since Microsoft has no real competitors?

To be honest, Microsoft's software security is improving. Windows 95 was dreadful, Windows 98 slightly better, and the improvement's continued through NT, XP and Vista. But the attackers are getting better too, and the protection in Vista isn't all for the user's benefit. As Peter Gutmann points out, enormous effort has gone into protecting premium video content, and almost no effort into protecting users' credit card numbers [570]. The same pattern has also been seen in other platform products, from the old IBM mainframe operating systems through telephone exchange switches to the Symbian operating system for mobile phones. Products are insecure at first, and although they improve over time, many of the new security features are for the vendor's benefit as much as the user's.

By now, you should not find this surprising. The combination of high fixed and low marginal costs, network effects and technical lock-in makes platform markets particularly likely to be dominated by single vendors, who stand to gain vast fortunes if they can win the race to dominate the market. In such a race, the notorious Microsoft philosophy of the 1990s — 'ship it Tuesday and get it right by version 3' — is perfectly rational behaviour. In such a race, the platform vendor must appeal at least as much to complementers — to the software companies who decide whether to write applications for its platform or for someone else's. Security gets in the way of applications, and it tends to be a lemons market anyway. So the rational vendor will enable (indeed encourage) all applications to run as root, until his position is secure. Then he will add more security — but there will still be a strong incentive to engineer it in such a way as to maximise customer lock-in, or to appeal to complementers in new markets such as digital media.

From the viewpoint of the consumer, markets with lock-in are often 'bargains then rip-offs'. You buy a nice new printer for $39.95, then find to your disgust after just a few months that you need two new printer cartridges for $19.95 each. You wonder whether you'd not be better off just buying a new printer. From the viewpoint of the application developer, markets with standards races based on lock-in look a bit like this. At first it's really easy to write code for them; later on, once you're committed, there are many more hoops to jump through. From the viewpoint of the poor consumer, they could be described as 'poor security, then security for someone else'.

Sometimes it can be worse than that. When racing to establish a dominant position, vendors are likely to engineer their products so that the cost of managing such security as there is falls on the user, rather than on the application developers. A classic example is SSL/TLS encryption. This was adopted in the mid-1990s as Microsoft and Netscape battled for dominance of the browser market. As I discussed in Chapter 2, SSL leaves it up to the user to assess the certificate offered by a web site and decide whether to trust it; and this has turned out to facilitate all kinds of phishing and other attacks. Yet dumping the compliance costs on the user made perfect sense at the time, and competing protocols such as SET, that would have required heavier investment by banks and merchants, were allowed to wither on the vine [357]. The world has ended up not just with a quite insecure system of Internet payments, but with widespread liability dumping that makes progress towards a better system difficult. Too much of the infrastructure has weakest-link rather than sum-of-efforts security.

7.5.4 Economics of Privacy

The big conundrum with privacy is that people say that they value privacy, yet act otherwise. If you stop people in the street and ask them their views, about a third say they are privacy fundamentalists and will never hand over their personal information to marketers or anyone else; about a third say they don't care; and about a third are in the middle, saying they'd take a pragmatic view of the risks and benefits of any disclosure. However, the behavior that people exhibit via their shopping behavior — both online and offline — is quite different; the great majority of people pay little heed to privacy, and will give away the most sensitive information for little benefit. Privacy-enhancing technologies have been offered for sale by various firms, yet most have failed in the marketplace. Why should this be?

Privacy is one aspect of information security that interested economists before 2000. In 1978, Richard Posner defined privacy in terms of secrecy [1035], and the following year extended this to seclusion [1036]. In 1980, Jack Hirshleifer published a seminal paper in which he argued that rather than being about withdrawing from society, privacy was a means of organising society, arising from evolved territorial behavior; internalised respect for property is what allows autonomy to persist in society. These privacy debates in the 1970s led in Europe to generic data-protection laws, while the USA limited itself to a few sector-specific laws such as HIPAA. Economists' appetite for work on privacy was further whetted recently by the Internet, the dotcom boom, and the exploding trade in personal information about online shoppers.

An early modern view of privacy can be found in a 1996 paper by Hal Varian who analysed privacy in terms of information markets [1289]. Consumers want to not be annoyed by irrelevant marketing calls while marketers do not want to waste effort. Yet both are frustrated, because of search costs, externalities and other factors. Varian suggested giving consumers rights in information about themselves, and letting them lease it to marketers with the proviso that it not be resold without permission.

The recent proliferation of complex, information-intensive business models demanded a broader approach. Andrew Odlyzko argued in 2003 that privacy erosion is a consequence of the desire to charge different prices for similar services [981]. Technology is simultaneously increasing both the incentives and the opportunities for price discrimination. Companies can mine online purchases and interactions for data revealing individuals' willingness to pay. From airline yield-management systems to complex and ever-changing software and telecommunications prices, differential pricing is economically efficient — but increasingly resented. Peter Swire argued that we should measure the costs of privacy intrusion more broadly [1237]. If a telesales operator calls 100 prospects, sells three of them insurance, and annoys 80, then the conventional analysis considers only the benefit to the three and to the insurer. However, persistent annoyance causes millions of people to go ex-directory, to not answer the phone during dinner, or to screen calls through an answering machine. The long-run societal harm can be considerable. Several empirical studies have backed this up by examining people's privacy valuations.

My own view on this is that it simply takes time for the public to assimilate the privacy risks. For thirty years or so, IT policy folks have been agonising about the death of privacy, but this remained a geek interest until recently. The significance is now starting to percolate down to sophisticated people like stock-market investors: Alessandro Acquisti and others have found that the stock price of companies reporting a security or privacy breach is likely to fall [8, 265]. It's only when tabloid newspapers and talk-radio shows give lots of coverage to stories of ordinary people who've suffered real harm as a result of 'identity theft' and phishing that the average voter will start to sit up and take notice. There are some early signs that this is starting to happen (for example in the growing number of requests that privacy experts like me get invited to appear on radio and TV shows). But another behavioural economist, George Loewnstein, points out that people are more sensitive to large changes in their circumstances rather than to small ones: they will get concerned if things suddenly get worse, but not if they get worse gradually. They also become habituated surprisingly easily to bad circumstances that they don't believe they can change.

It may be of particular interest that, in late 2007, the British government suffered spectacular embarrassment when it lost the electronic tax records on all the nation's children and their families — including bank account details — leading to a personal apology in Paliament from the Prime Minister and massive media coverage of subsequent privacy breaches. I'll discuss this in more detail in section 9.4; the privacy economist's interest will be in whether this changes public attitudes in any measurable way over time, and whether attitudes stay changed.

7.5.5 Economics of DRM

Rights-management technologies have also come in for economic scrutiny. Hal Varian pointed out in 2002 that DRM and similar mechanisms were also about tying, bundling and price discrimination; and that their unfettered use could damage competition [1291]. I wrote an FAQ on 'Trusted Computing' in 2003, followed by a research paper, in which I pointed out the potential for competitive abuse of rights management mechanisms; for example, by transferring control of user data from the owner of the machine on which it is stored to the creator of the file in which it is stored, the potential for lock-in is hugely increased [53]. Think of the example above, in which a law firm of 100 fee earners each has a PC on which they install Office for $500. The $50,000 they pay Microsoft is roughly equal to the total costs of switching to (say) OpenOffice, including training, converting files and so on. However, if control of the files moves to its thousands of customers, and the firm now has to contact each customer and request a digital certificate in order to migrate the file, then clearly the switching costs have increased — so you can expect the cost of Office to increase too, over time.

There are some interesting angles on the debate about rights management in music too. In 2004, Felix Oberholzer and Koleman Strumpf published a now-famous paper, in which they examined how music downloads and record sales were correlated [978]. They showed that downloads do not do significant harm to the music industry. Even in the most pessimistic interpretation, five thousand downloads are needed to displace a single album sale, while high-selling albums actually benefit from file sharing. This research was hotly disputed by music-industry spokesmen at the time, but has since been confirmed by Canadian government research that found a positive correlation between downloading and CD sales among peer-to-peer system users, and no correlation among the population as a whole [28].

In January 2005, Hal Varian made a controversial prediction [1293]: that stronger DRM would help system vendors more than the music industry, because the computer industry is more concentrated (with only three serious suppliers of DRM platforms — Microsoft, Sony, and the dominant firm, Apple). The content industry scoffed, but by the end of that year music publishers were protesting that Apple was getting too large a share of the cash from online music sales. As power in the supply chain moved from the music majors to the platform vendors, so power in the music industry appears to be shifting from the majors to the independents, just as airline deregulation favoured aircraft makers and low-cost airlines. This is a striking demonstration of the predictive power of economic analysis. By fighting a non-existent threat, the record industry helped the computer industry forge a weapon that may be its undoing.

7.6 Summary

Many systems fail because the incentives are wrong, rather than because of some technical design mistake. As a result, the security engineer needs to understand basic economics as well as the basics of crypto, protocols, access controls and psychology. Security economics is a rapidly growing research area that explains many of the things that we used to consider just 'bad weather', such as the insecurity of Windows. It constantly throws up fascinating new insights into all sorts of questions from how to optimise the patching cycle through whether people really care about privacy to what legislators might do about DRM.

Research Problems

So far, two areas of economics have been explored for their relevance to security, namely microeconomics and game theory. Behavioural economics (the boundary between economics and psychology) has also started to yield insights. But economics is a vast subject. What other ideas might it give us?

Further Reading

The best initial introduction to information economics is Shapiro and Varian's *'Information Rules'* which remains remarkably fresh and accurate for a book written ten years ago [1159]. I generally recommend that students read this first. For those who want to go on to do research in the subject, I then suggest Hal Varian's textbook *'Intermediate Microeconomics'* which covers the material from an academic viewpoint, with fewer case histories and more mathematics [1284].

The current research in security economics is published mostly at the Workshop on the Economics of Information Security (WEIS), which has been held annually since 2002; details of WEIS, and other relevant events, can be found at [58]. There is a current (2007) survey of the field, that I wrote with Tyler Moore, at [72]. There are two books of collected research papers [257, 548], and a popular account by Bruce Schneier [1129]; I also maintain an Economics and Security Resource Page at `http://www.cl.cam.ac.uk/ rja14/econsec.html`. Two other relevant papers, which are in press as this book goes to print, are a report I'm writing with Rainer Böhme, Richard Clayton and Tyler Moore on security economics in the European internal market [62], and an OECD report by Michel van Eeten and colleagues that reports extensive interviews with information security stakeholders about the incentives they face in practice [420].

A number of economists study related areas. I mentioned Jack Hirshleifer's conflict theory; a number of his papers are available in a book [610]. Another really important strand is the economics of crime, which was kick-started by Gary Becker [138], and has recently been popularised by Steve Levitt and

Stephen Dubner's 'Freakonomics' [787]. Much of this analyses volume crime by deprived young males, an issue to which I'll return in Chapter 11; but some scholars have also studied organised crime [392, 473]. As computer crime is increasingly driven by the profit motive rather than by ego and bragging rights, we can expect economic analyses to be ever more useful.