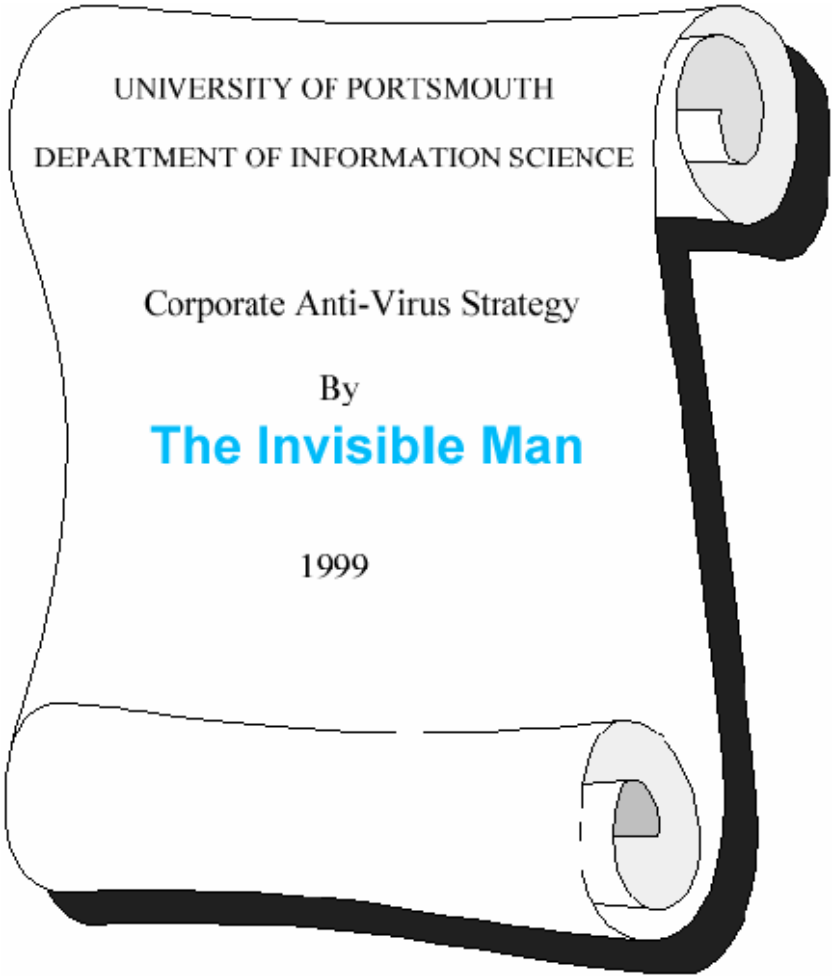

Final year project undertaken in partial fulfilment of the requirements for the BSc (Honours) Degree in Computer Science.

ABSTRACT

Computer viruses have become a major security threat to our work and home PC's. Their complexity and quantity have increased significantly in recent years, and anti-virus companies battle to deal with the solutions that detect and remove them.

The aim of the project is to improve awareness of the problem, and solutions. The nature of the work undertaken to meet this aim was a study of the viral code methods, and to enable a corporation to detect and remove these viruses thoroughly.

This involved writing suitable policies, training plans for staff and the selection of specialist software to perform these tasks correctly. The first section explains the types of viruses, the software tools to remove them and guidelines for management and staff that wish to create their own anti-virus strategy.

The second section of the project involved creating a generic virus scanner for word macro viruses, the most widespread type of virus problem to date. By exploiting a common aspect in these viruses, it is possible detect and remove most word viruses both old and new. This could be marketable as a useful anti-virus tool to help combat an increasing security risk.

In conclusion: This project was successful within the scope of the problem, and a steep learning curve was achieved to understand the issues involved. This will continue to be an area of great interest for future work and employment within the field.

Project Specification **Disclaimer**

Contents

1.0 Introduction

In 1986, two programmers Basit and Amjad realised that the first sector of a floppy disk contained executable code that ran when you booted from drive A:. They experimented by replacing the code with their own memory resident program that copied itself to other disks when accessed. They named their program a virus, causing much confusion at the time because people believed you might catch a biological virus from your computer. This was the first boot sector virus, but has now become extinct 'in the field' as it only infects 360KB floppy disks. The only effects of the virus were to copy itself and label the volume of each disk as "(c) Brain". The University of Delaware discovered many infected disks in 1987, and were able to remove the virus by searching for the Brain label and replacing the boot sector with a program.

Within the same year, another programmer called Ralf Burger discovered that you can write a program that copies itself inside executable COM files. This made the repair process more difficult then simply replacing a boot sector. The virus was a demonstration called 'VirDem' with explicit messages giving the game away to prevent wide-scale infection. He showed his findings to the controversial German hacking group the 'Chaos Computer Club' in December of that year. It interested so many people, that he was asked to write a book 'Computer Viruses – A High Tech Disease'. He showed that it was possible to write viruses in other computer languages Pascal, Batch file language and even Basic.

In Vienna, Burger acquired a virus called 'Charlie' (shortly to be called 'Vienna'), and was the first virus to be disassembled by a programmer named Bernt Fix.
He published the source code for VirDem and the Vienna virus that appeared shortly after the conference, and the book became very popular as a result. This generated new ideas from its readers to improve the virus, and many variants of Vienna were written as a direct consequence of his book.

By the end of 1987, over a dozen viruses existed and spread at a prolific rate. There were no measures to prevent these viruses so anti-virus companies started to appear in order to combat a growing problem. It has become a battle that cannot be won from either side. The virus author will eventually have their virus detected and eradicated. The virus researcher will always have new viruses to disassemble and prevent. In 1999 there are over 22,000 different viruses to defeat. This makes the virus scanner a very large and difficult program to write, due to the limited human resources that are capable of analysing the new influx of virus code. A recent merge between Mcafee and Dr Solomon's gives the product more expertise, so this may become a common trend for the future to prevent the amount of new viruses appearing in the world.

2.1 What is a computer virus?

> "A virus is a program that copies itself."
> (P.1 Alan Solomon, 1997)

It requires a programmer to purposefully create this program, in the hope that it will copy to many computer systems and spread to unsuspecting users.
These people are called "virus authors".

All a virus has to do is copy itself to be called a virus and some viruses do nothing more than replicate. Most of these have other effects too, like playing music, displaying messages and at the worst case, destroying your valuable data. It will do anything the virus author is able to program correctly. Some virus authors are less skilful that they would like to be, causing unexpected software results such as a crash. This may involve loss of work, even if this wasn't intentional.

Why are viruses written?

There are many possible reasons behind this, a virus author may want to create one as a personal challenge, revenge their company for being dismissed or an anarchist who enjoys the misfortunes of others, especially ones caused by the perpetrator.

The virus programmers clearly like showing off with messages left in their code, and call themselves by a codename so they have hidden recognition in the world. It may give them a sense of power, controlling the destruction of a network for example.

Viruses are not always released by the original author as the case of the Jerusalem virus, which was stolen by a friend of the author. It was clearly not ready for release, as the author had begun to implement changes as to how it infected another file. As a result, the virus was bugged and didn't work as well as it should have done.

How are viruses written?

Theoretically, you can use any programming language that has suitable file handling capabilities. Higher level languages such as Pascal, Ada and C produce large object code because they make use of libraries of functions, of which not all are used.
This is more likely to be discovered as the virus code will run slowly and take up a noticeable amount of disk space. For this reason most viruses are written in assembly language, producing low level code so they run quickly and remain inconspicuous.

Assembly language is difficult to use, thus more error prone for the author. Mistakes in the virus code can stop part of the virus from working properly as designed. This could be the replicating code infecting a file more than once until the program is too large to fit into memory or the effects (payload) do not trigger off properly.
Some viruses were so badly written, they did not infect a file or disk at all, these are known as "intended viruses" which isn't a true virus and doesn't generate a risk to us.

The source code for thousands of viruses are published on the Internet and BBS's (Bulletin Board Systems) and are available for download by the general public. There are tutorials available that teach you how to write viral code, and suggest books and on-line resources to aid your virus programming. This had undoubtedly sparked the virus creation revolution, and there are now more than 20,000 viruses in existence.

Virus creation kits make life very easy without coding a virus. You pick the options that you want your virus to have, and the program will generate the code for you. This gives a wide range of related viruses but are usually detectable as a new virus, if the anti-virus program compares their characteristic similarities.

New threats with Microsoft Word, Excel and Access data files have arisen because the word processor, spreadsheet and database programs incorporate a powerful macro language designed to automate, standardise, and control your work created with them. Unfortunately the language is powerful enough to handle files, giving the virus author the opportunity to write a virus macro that copies itself to other documents that you open or create. They can also contain a standard virus, so have been used as an easy way to infect a company from the outside due to the Microsoft Office packages gaining widespread popularity.

Where do you catch a computer virus?

Viruses hide in executable code. That means that any software you load, any floppy disk you leave in the machine runs the risk of copying a virus. Many people seem to believe that only copies of software and pirated software can harbour a virus, but this is not the case.

In 1992, Borland released a virus with their C++ programming development software (an original shrink-wrapped silver CD) and shipped thousands of copies worldwide. This is probably due to lack of anti-virus checks or the updates for the anti-virus software being out of date. Many PC magazines CD's have made the same mistake in recent years, so you cannot trust software even from reputable sources.

In addition, software from the Internet and BBS systems are not to be trusted, but often have to be used. It's the software and strategy that prevents problems here. The possible solutions to these problems are explained in the anti-virus policy section.

How do viruses work?

The majority of viruses work at low-level (precise & explicit machine instructions) so they are able to manipulate the operating system. This allows the code to be small in size and quick to execute. They hook onto "Interrupts" and "Functions" to perform their tasks, so that normal operations such as keyboard use, dates and times and disk access can be triggering the virus code to either copy or activate it's payload.

The more successful viruses hide in the background and do not announce their presence for a long time. This gives the virus time to copy further and perhaps cause subtle damage. If a virus destroys its host too quickly, measures will be taken to find and eradicate the virus.

2.2 Types of computer virus & their mechanisms

Types of virus

Individual viruses each have their own way of operating, either to avoid detection or written in the only way a programmer knows how. It is possible to categorise the way they work, because there are a limited number of ways to infect a PC successfully. The possible categories are listed below:

Companion

Because .COM files are executed before .EXE files, the virus creates a .COM file with the same name and runs first. These are one of the easiest to detect and remove, just deleting the repeated .COM filename. Some more advanced versions exist which hide the existence of the .COM file. These require a clean boot to see the hidden files. Generally not very widespread because the infection is obvious and limited.

Appending

These will add their own code inside an executable file, thus is much harder to remove without an anti-virus tool. Some viruses add themselves to a file in an unpredictable way, making sure the file needs to be replaced and cannot be repaired by a scanner.

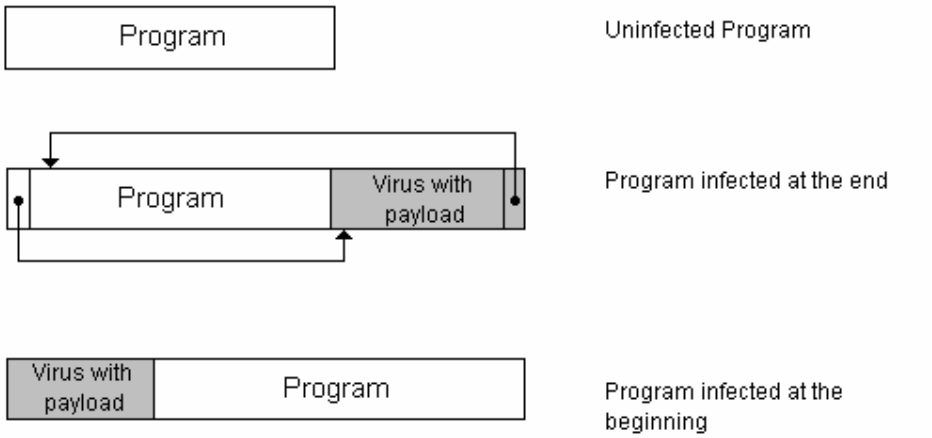

Fig 2.1 File viruses are infected by jumping to code sections or re-writing the whole file. There are many different ways a file virus can infect a file.

Overwriting

Posing little threat because it will overwrite the beginning of every file it infects. It is therefore possible to detect its presence due to files no longer working as they should. The shortest virus falls into this category at a miniature 25 bytes of viral code.

Directory

One of the worst types of viruses for cleanup because they manipulate directory entries and file allocation tables to remain hidden from view. This gives incompatibility with 32-bit disk access in Windows, cross-linked files and data loss.

TSR (Terminate-Stay-Resident)

These viruses take a portion of memory so that it can run in the background for the duration of the machine's use. It allows the virus to copy to every file and / or boot sector it loads or encounters. This gives the virus a much higher infection rate.

Non-TSR

Copy to other files at runtime, searching for a non-infected file to invade. They may infect once or more, depending on the infection rate counter but they are very slow to infect, as the disk becomes full of infected files. The virus has to work harder to find a clean file, but the scanner will not find the virus in memory.

Boot Sector / Partition Sector Virus

A common type of virus because it does not rely on files to run or copy. The majority of computer users own floppy disks that they store and share their work with. Every disk contains executable code at the beginning of the disk (whether bootable or not), which can be replaced or modified by a virus. When an infected disk is left in a machine during bootup, the virus copies itself to the hard disk boot sectors / partition table and infects every writeable floppy disk during the computers use.

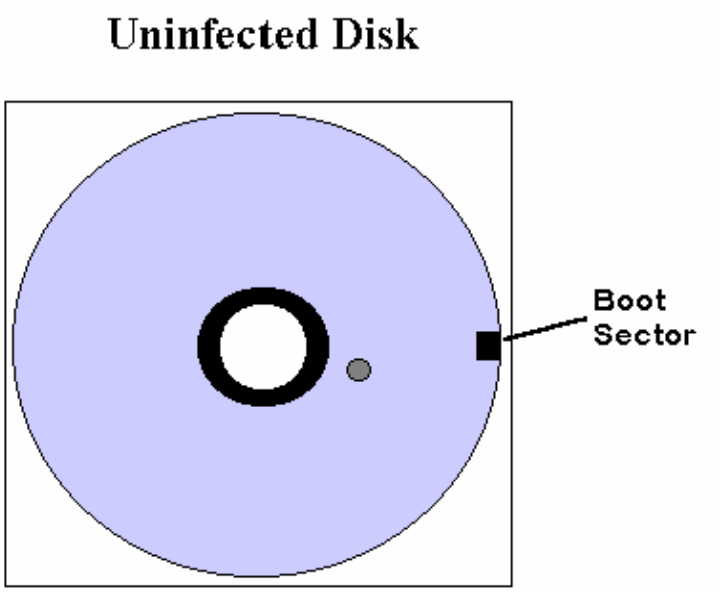

Fig 2.2 A floppy disk boot sector is present on every disk. It's simply a small executable program that is vulnerable to alterations by a virus.

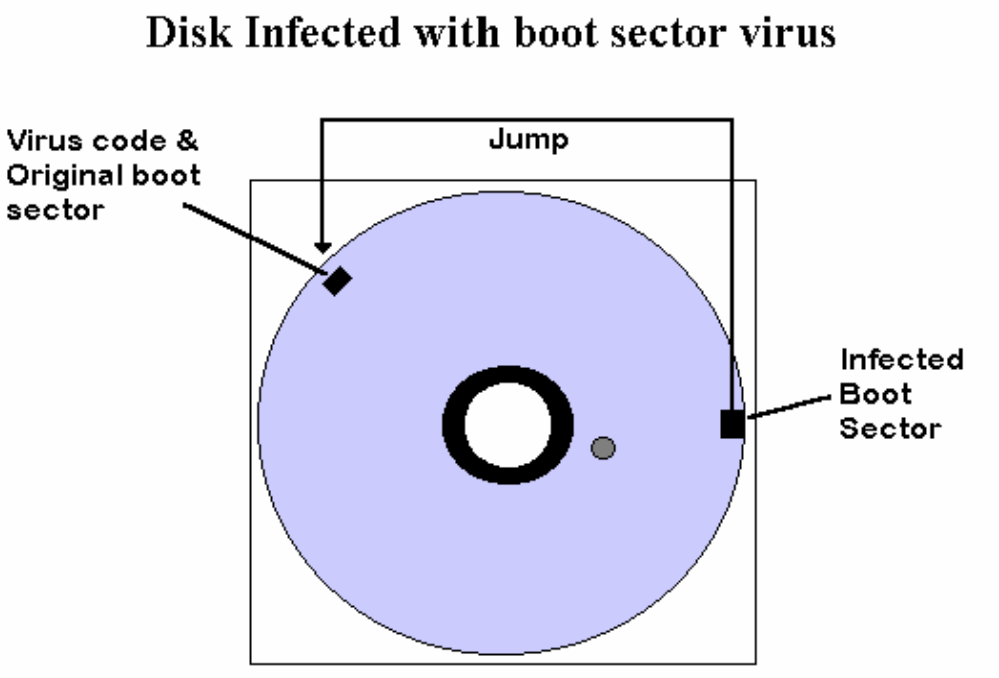

Fig 2.3 The infected boot sector contains the virus with a jump to the original code. Some viruses just replace the whole code without jumping.

Macro Viruses

Possibly the biggest threat to security because most people in an organisation share documents more often than programs. Documents are shared by floppy disk, network and e-mail attachments, so they can copy very quickly and cause headaches for the support staff that try to remove the virus. Examples of macro viruses are evident in Microsoft Word 6 and later, Lotus Ami Pro, Microsoft Excel 5 and later, Microsoft Access to name but a few.

The interesting thing about this type is that it's the first "platform independent" virus. Organisations with Apple Macintosh computers running word are able to give a PC user the virus just by sharing the same document. The language used to write the virus belongs to the application, not the machine code that most viruses work with. Because of the widespread use of word processors, and that companies usually use the same package with a site-license, this type of infection spreads very quickly and takes a considerable amount of time to eradicate from the workplace.

This project aims to deal with Word Macro Viruses (dot, doc) , but the engine could easily be adopted to cope with Excel (xls) viruses, just by changing the search string and file types. Access database (mdb) viruses are more difficult to remove, but further research will be sought to study this type of virus and allow Gw-Scan to disable it.

PowerPoint is the presentation software application that belongs to Microsoft Office. There has been speculation that a virus will appear to infect Presentation (ppt) files, and recently this has been proven. It could prevent a potential security problem if the scanner was to locate the possible names of automacros from these files too, as once a security vunerability is proven, it will almost certainly be repeated by another virus author. The programmer will look into this area once the original product specification of GW-Scan is satisfactory and complete.

Virii Mechanisms

Fast Infector

A technique designed by the Dark Avenger virus author from Sofia, Bulgaria. His idea was that every file copied both the source and destination files are infected. This allows the virus to copy over networked machines at an incredible rate. It includes the ability to infect all files opened or attributes changed as well as infecting upon execution. The impact of this meant that virus scanners that could not detect the virus actually copied the virus to every executable file on the hard disk!

Slow Infector

These will only infect on an infrequent basis or for example on the eighth program run. The idea behind this is that you would find it more difficult to monitor infection slowdown or change.

Polymorphic

There is much confusion about this category of virus, some people believe that the virus changes into something else every time it infects another file. This is a curious and widely misunderstood way of explaining what it actually does. The encryption and decryption keys produced differ on each infection. The result is that no two bytes in the virus appear the same, or in the same place. The virus code is still exactly the same, but simply chooses a new algorithm to mask itself with. The reason for doing this is to evade the virus scanner, which looks for a sequence of bytes in a file and determines whether or not there is a virus in the file.

Occasionally, new mechanisms are discovered that force the anti-virus companies to re-write their anti-virus software. This was the case in 1993 with Polymorphic viruses. Three major anti-virus companies Sophos, X-tree and Untouchable could not keep up with the difficulty and volume of these new types of virus. It was the first sign that virus authors were beginning to win the battle by the frustration of detecting them.

Scanners now had to be more sophisticated as to how they detected this type of virus, and Dr Solomon produced the “Generic Decryption Engine” to remove these random layers of encryption.

Stealth

There are levels of stealth that go to different extremes to remain hidden from view. Simple stealth tactics aim to hide date/time changes, filesizes and attribute changes made when the virus infects another file. Higher levels of stealth aim to conceal the virus from anti-virus software and are known to be “fully stealthed”. The virus can only perform these tactics when it’s loaded in memory. It is vital that computers are booted from a clean, write-protected DOS disk before using a virus scanner, so that the changes made by the virus are in full view of the scanner, without the virus attempting to conceal the changes when it is running in memory (TSR).

CMOS Tricks

The CMOS (Complimentary Metal-Oxide Semiconductor) is a chip on the computer motherboard that stores the date, time, hard disk and setup information retrieved when the computer is first switched on. A virus cannot store itself in this area, but can change the values that the PC uses to boot up.

The Goldbug virus is an example of how the CMOS can be used to make a virus cleanup more difficult. It modifies the area of the CMOS that tells the PC if it has a floppy drive, and sets the type of floppy drive to none. This results in the hard disk booting before the floppy drive can, loading the virus from hard disk, then determining if there is a floppy disk in the drive and booting from it if there is.

This prevents you from booting clean, because the virus will always be in memory unless you turn off the computer, go into setup and change the floppy information and boot from a disk straight away. Some machines need to have the battery taken out that allows the CMOS to keep its settings after a power off, making cleanup very tricky.

BIOS FLASH

Since the introduction of flash upgradable PC's, we have been able to update the software that is written into our computers motherboard, the central bridge that connects the main part of a computer together.

In July 1999, the CIH virus introduced another deadly weapon to include with a virus. It overwrites the BIOS program the computer needs to operate, and needs to be replaced with a new motherboard before the computer will work. This is expensive.

MODEM MAYHEM

The modem allows communication of computers using a normal telephone line. This gives the computer an ability to dial telephone numbers to anywhere in the world.

The programmer of the Armageddon virus abused this facility by including a very expensive payload to his virus. Between the early hours of the morning, the virus would get the modem to dial the Greek speaking clock, then disconnect after 3 or 4 hours had passed. This can mount up to hundreds of pounds in phone call charges per week, based at an international rate.

2.3 What isn't a computer virus

Trojan Pretends to be something that it isn't or does something unwanted in addition to the original purpose. It's a program that may have ill effects but does not copy itself. It relies on the user to distribute the program themselves, and do not travel otherwise. An example of this is the "Aids Information" Trojan distributed to thousands of people to learn about the human virus, only to have their hard disks formatted when using the software.

Joke May look like a message "Formatting Drive C: – 1% complete" to fool you it is destroying your data. They cause no harm but to shock the user. As a result, these are commonly detected under most leading virus scanners.

Intended Virus When a virus that was so badly written does not even replicate itself. These contain a fatal programming design flaw and cannot be a threat to the computer unless they are later modified with the mistakes corrected.

False Alarms A warning from anti-virus software telling you there is a virus when there isn't. This may be due to badly coded scanners, the likeness of a virus infected file compared to an innocent file because the search string that the anti-virus scanner uses matches the same sequence of bytes in the file. These cause panic and users have been known to re-format and loose their work when they really didn't have to and were not under viral attack. In the early stages of scanner development, anti-virus packages would find the search strings within other brands of anti-virus product. They now encrypt their search strings to prevent this type of false alarm.

Droppers Programs that when executed will copy a virus to a system or floppy disk. Although they contain a virus, the program does not copy itself.

Worms Propagate only through networked machines hopping and replicating from one node to the next, and maybe network to network. They can cause damage if designed to, but usually slowdown networks to a halt like when the Morris worm attacked the Internet in the U.S.A.

Hoaxes Caused by people sending e-mail about a virus threat that doesn't exist. Inexperienced people believe that viruses can spread via e-mail text messages because they think it works like a word virus. This causes many people to warn others with the same message and a chainmail effect is produced. To obtain a list of hoaxes you can visit: http://urbanlegends.miningco.com/library/weekly/aa030798.htm

2.4 Payloads

A payload is an unwanted effect of the virus program. Examples are messages, crashes, music, file deletion and data corruption. Here is an example of a virus payload from the "Ambulance" virus. In MS-DOS, a picture of an ambulance with a noisy siren and a flashing light moves across the screen from left to right, wiping out the screen's text. Apart from these effects it is harmless and trivial to repair:

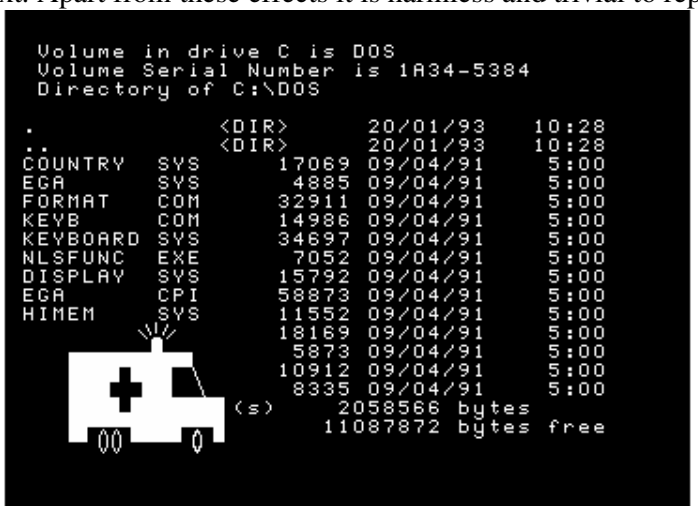

Fig 2.4 The Ambulance virus displays a moving graphic with sounds

The type of payload determines the category of damage that the virus can cause. It is important to know which category the virus falls under, so you know exactly what to deal with in order to recover from it. Other viruses are more sinister, and harder to recover from. It is therefore vital that you are able to recover everything from a daily backup copy. Here is another example of a payload from Casino (moderate damage)

Fig 2.5 The casino virus plays a game with the user, if the user wins the virus returns the FAT data back to the hard disk, otherwise a message is displayed and the computer locks up, losing vital data.

Categories of payload damage

Trivial Effects may be messages or music, but you only need to remove the virus to solve the problem. May take 3 minutes per computer when clean booting the anti-virus from a floppy disk.

Minor Files being deleted when executed like the Friday 13th virus are an example. You will need to replace these deleted executables with clean backup copies, and scan for the virus to prevent a re-occurrence.

Moderate When the hard disk data is scrambled, the FAT deleted or the boot / partition sectors destroyed or the hard disk overwritten. It is moderate because you are aware it has happened and it may loose you a day's work provided you backup every day, and a few hours to reformat and completely restore from the backups.

Major When a virus overwrites random sectors on the hard disk periodically or at random with a text string, such as "Eddie lives… somewhere in time" (Dark Avenger).
It may affect backups because the damage is subtle, and may occur for many weeks before it has been discovered. Restoring from a recent backup may still involve data corruption, but searches for "Eddie lives" will tell you the areas that were hit.

Severe As above, but does not contain a recognisable text string. It may contain spaces or randomly garbled data, in which case you cannot tell which areas of the disk were hit. Backups are affected but with no way of proving correct until they are used.

Unlimited Some viruses succeed in stealing administrator logins and passwords which are passed onto a third party. The third party may login to the system and cause as much or as little damage as they wish to. Data may be altered for fraudulent motives.

3.0 Anti-Virus Strategy

3.1 Anti-virus software

The tools used are important, and you will find that there are many free evaluation copies for you to test before your purchase. There are many different aspects of the software that should be considered before a purchase. This can make the process of choosing the right tool quite confusing. As a result of this confusion, a chart of relevant features could help a potential buyer to decide which one to obtain:

	PC-Cillin	Norton	VET	Solomon	Mcafee
Available Platforms	DOS, Win3.x, Win95, WinNT	DOS, Win3.x, Win95, WinNT	DOS, Win3.x, Win95, WinNT	DOS, Win3.x, Win95, WinNT	OS/2, DOS, Win3.x, Win95, WinNT
Pattern Updates	Free lifetime updates	Free lifetime updates	Free lifetime updates	Quarterly / Monthly	Free lifetime updates
Update frequency	Free web updates	Liveupdate from web	Web based update	Disks posted monthly / quarterly	Free web updates
Attachments	Scanmail + Smartmonitor	Autoprotect	Resident protection	Virusguard	Webscan X
HTTP/FTP Intercept	Smartmonitor	Autoprotect	Resident protection	Virusguard	Webscan X
Macro Viruses	Macrotrap	Autoprotect	Resident protection	Virusguard	Code Matrix
Unknown Viruses	Smart monitor	Striker Technology	Resident protection	Findvirus	Code Trace
Office 97	Yes	Yes	Yes	Yes	Yes
Viral Information	Yes	Yes + Web	Yes	Yes + Book & Web	Yes + Web
Emergency Disk	Clean disk	Scan during install	Scan during install	Magic Bullet	Emergency Disk
URL	http://www.antivirus.com/	www.symantec.com	www.vet.com.au	www.drsolomon.com	www.mcafee.com
Price	£45	£50	£60	£80	£40
Rating	**	****	**	*****	*****

Of course there will be many other options to consider that will affect the buyers reason to purchase. Cost is unfortunately for many the only deciding purchase factor, as people presume that they will all detect with equal ability. Sadly this is not the case. The products need to be tested professionally to acquire their detection ratio.

Each company concerned should draw their own table of features, because some aspects of the software design will be more important to them than others. Some of which are:

- User interface (Clumsy interfaces often aggravate the user, some look too bland)
- Ratio of detection / false alarms – The reason to buy the product in the first place!
- Speed of operation
- Customisability including password protection on some features
- Compatibility with other software & network used
- Documentation of features and ease of understanding the manual
- Level of technical competence to configure and use
- Concise information about viruses and their effects
- Quality & Speed of Technical Support
- Speed of response to deal with new viruses & send an antidote
- Network update facility

The problem with virus scanners

As there are now hundreds of new viruses appearing every month, the option of scanning for every virus becomes slower to complete and takes more disk space. Anti-virus programs that used to be shipped on floppy disk are now sent with a CD, Dr Solomon's is one of the few scanners that gives you an all-in one boot disk called "The Magic Bullet" currently used by students by loan request from University of Portsmouth. As the scanner is becoming larger, it will not be long before two disks are necessary for all search strings to be included.

The main problem with the scanner is that users will not use the software on a regular basis. It disrupts their normal working practice because of the time it takes to scan all of the executable files. The virus scanners that start each time a PC is loaded take more time to boot, and users find ways to bypass the scanner to get on with their work. The solution needs to force virus protection upon the users with the least amount of interaction from the user.

TSR Scanners

These work in the background and scan every executable file and word document as they are accessed. They are much more effective than a scanner because the user simply carries on as normal without having to do anything extra. Files will get scanned more frequently than they need be, but this is a small price to pay for real-time scanning and user compliance.

The only requirement is that a virus is reported if found, but many scanners automatically inform the administrator if configured this way and networked.
The most important part of this type of tool is a standard installation guide that is followed by support staff when initially installing the software. Choices like who to send warning messages to, conditions for file scanning e.g. "scan for viruses on all file writes" will protect against known word document viruses as e-mail attachments. Updates take the old settings into account, and reuse them for a newer version.

It was found that Dr Solomon's Anti-Virus Toolkit, Management edition was the best solution to install, configure and update all workstations and servers from one machine.

The recent acquisition of Dr Solomon by Network Associates will help Mcafee produce their new product VirusScan 4 using the Solomon engine. This will be the product to buy in the future, but for now Dr Solomon's is the best performer.

3.2 Choosing the right level of protection

This is the question that must be asked before the install and configure takes place. It is best to assess what protection level applies best to your workplace. If the hardware is slow, on-demand scanners are best configured to run at regular intervals. For faster hardware, the TSR scanner will run quick enough not to notice it's presence and gives you real-time protection. Some CD-writers dislike other processes running, so testing the product with such machines is necessary to keep the machine fulfilling its purpose. For most cases, the TSR scanner is the best approach, and should be the most widely adopted.

For special machines like the sheep-dip computer, a checksumming Bootdisk to detect changes to executables should be used and updated when the scanner is updated. If the sheep-dip ever caught a new virus, the effects would be catastrophic for the clean-up process – you would not be able to trust any disk in the workplace, regardless of all-clear stickers.

3.3 Anti-Virus Policies

The vast increase of viral threat to the workplace presents us with a problem, how can you prevent or reduce the risk of infection? An effective anti-virus policy needs to be constructed around four key elements: Rules, procedures, education and tools. These will vary depending on the nature and size of the organisation, but the outline described below shows how to define these elements into a concrete operation.

Rules

The rules must have suitable working practice methods and to be complete regarding every possible risk of a virus entering the company. For example, the rule that all media must be checked needs a sheep-dip computer, a virus checking staff member(s), permanent marker & labels to date and initialise and authorise the use of the media by a chosen category of sticker.

Records can be kept to see who is and who isn't getting their media checked, and a history log of previous infections to asses where the virus keeps re-appearing. These procedures would also benefit greatly from software auditing and asset control databases to limit the amount of software that needs to be used, and which versions currently reside on their hard disks. This prevents running into trouble with software licences and gives the support staff the ability to re-install at the worst case.

Since backups are a major part of the rule, data should be saved to a network drive and backed up centrally to ensure that all company data is safe from attack. The

backup tapes, optical storage and off-line storage should be frequently produced, and not replaced with backup tapes used after one week. Many viruses slowly corrupt your data, which can go unnoticed for a long time. If you only store a week's backup, and you re-use the tapes from last week, you cannot prevent a virus that caused damage two weeks ago. It is very important to store many months of backups onto fresh tapes, but work out a set time limit the backups are stored for before they are overwritten.

Procedures

This is how the rules are going to be implemented by a combination of staff skill, equipment, software and everyday maintenance to protect the data of a network. One of the most important procedures is how the data should be backed up and recovered in the event of a disaster. The rules will state who is responsible for the backups, the procedures will tell that person how to accomplish this step-by-step. If the procedures fail due to a new type of virus, they will need to be changed to accommodate the differences in attack arisen by the newest security threats.

Procedures need to be clearly written and accessible to the IT support staff. Periodic knowledge testing of the staff regarding these procedures is a good way of checking the clarity and understanding of what is required.

Education

Training allows the staff to understand the problem, and how to deal with a virus outbreak. Seminars such as the Virus Bulletin conference exist to inform the delegated anti-virus employee about the threats, products and management issues. It can also allow this person to train others in their company if they contain the interpersonal skills necessary to convey the information to them.

Dr Solomon's have been very successful in the past with their virus workshops. The attendants participate in the removal of live viruses, giving them a chance to practice with the real thing. This kind of simulation provides confidence to the staff, so that when a critical situation arises, it can be dealt with calmly and correctly.

These are typically expensive, starting at £500 to £2000 per person, so it may be worth considering a training policy to deal with these issues. A staff member will be required to keep the issues and solutions up-to-date, and to publish the learning materials distributed to the relevant people.

If the users understand why they need to get their disks virus checked before use, there is a higher likeliness that users will stick to the rules and procedures.

These are a few pictures from the 1997 Virus Bulletin conference:

Paul Ducklin gives the introduction to the 1997 conference in Germany.
All major anti-virus companies and security consultants hold these conferences.

The virus bulletin developer's exhibition allows anti-virus companies to demonstrate the features of their software, to attract the corporate market to their sales.

Tools

The software and hardware used to detect and remove viruses. See section 3.1 on Anti-virus software for a guide on which anti-virus package to use. A room of clean machines is required when checking all media in a company after a virus is discovered. Using the right software can limit the damage to just one machine, because of real-time scanning with a TSR scanner. If other measures are being used, all media will have to be scanned systematically to prevent the virus coming back.

Suitable stickers, marker pens, desks and power outlets need to be available to the staff who carry out the clean-up, so they may use many machines and categorise the media into usernames, locations and infection status (clean or infected).
Using a management edition of the anti-virus package can update all computers attached to the network, and updating the whole site against the latest viruses. Unfortunately, many companies lack the initiative to configure their software this way, forcing them to update each machine locally. As a result, the job doesn't get done due to the scale of machines, and known viruses get introduced very quickly. This causes a major problem of virus clean-up and possible re-infection.

Anti-Virus Policy guidelines

Media Checking

This is the concept of producing a standard set of rules that govern the use of any PC media: new software, pre-formatted blank floppy disks, copying disks, documents and the use of the Internet for software apps, drivers, updates and utilities. Any of these sources may introduce a virus if not prevented by anti-virus software and policy.
It is therefore vital that all media is checked at least once. In the case of CD's they only need to be checked once with the latest scanner, and all writeable media to be checked if it has been used on another machine (i.e. home or colleagues machine).
A TSR scanner and regular backups can only limit Internet access damage.

Staff requirements

Depending on the size of the company, you will need to employ a set amount of people to cope with routine tasks: Daily backups, media checking, installing anti-virus software and the updates. If no backup devices are in use, the purchase and installation of these, along with backup procedures are required. The choice of virus product can simplify the updates, so this will need planning in advance.

The process of corporate anti-virus is dependent on company size. The worst case will be a major company, where many helpers are required to make the system work. This section now demonstrates how it is possible to defend against most viruses, provided the procedures are followed.

3.4 Using a Policy when disaster strikes

An anti-virus policy is certainly tested when there is a real problem to deal with. The first time the policy is followed there will usually be sections to add that make the whole operation of virus recovery a simpler process for the future. The following case study illustrates how important it is to be prepared in advance, which minimises panic and uncertainty when dealing with the task.

Example Case Study: Infection occurs at Acme Computer Systems Ltd

A virus outbreak occurs at a mid-sized company on an unknown amount of computers and disk based media. The first step is to know which type of virus you are up against. A scanner will give you the name of the virus, which type of infection occurs, it's effects and how to remove it without data loss. Other points to consider are minimising the disruption of the business whilst a full-scale check is carried out. This is not always possible, but contingency plans should be prepared in advance.

You need to check all media in the building for viruses to prevent a re-occurrence, but this can be limited to just a few computers, and an assortment of disks. A good strategy allows you to find the original source of infection, provided file auditing is practised at the company. This gives a log of when files are changed and may give messages to auditors about new virus attacks, and the first infection.

The servers should be disconnected from the rest of the network and be checked first, along with workstations and media checks afterwards. You can limit the amount of CD's to check by only checking silver CD's once and giving them a permanent OK sticker to save time in the future, it should still be monitored for the threat of a new virus though. Moving valuable data to writeable CD's can prevent write access to the majority of programs and data, including complete system backups. They only need to be virus scanned another time if a session is added to the CD, creating a multi-session CD with new files that could harbour a virus. When the CD is closed it has had all sessions written to the disc without the ability to alter it. It may then be given a permanent OK sticker to prevent unnecessary scans of it in the future.

The people within the company must know about the virus, and to restore their goodwill it is best to scan everyone in the building. Staff may be worried if they have the virus or not, but you have to be careful that you express yourself in a manner that protects the user, and doesn't incriminate them. They should only be punished for not following the rules if you can prove they didn't check their media. Allowing users to check their home PC's at work is a viable option to eradicate the virus, and ensures that the PC is checked properly with the latest versions.

If there are thousands of disks and CD's to check it is worth considering a floppy disk auto-loader, and CD-ROM auto-changer drives to speed up the scanning process. Other equipment would consist of an uninfected PC (or several) to control autoloaders, floppy disk drives and CD's. Stationary would involve thin marker pens and anti-virus labels to show virus check clearance. It is best to pick a room with suitable furniture to allow the use of computers and the adequate seating of the users. Power points and adapters / extensions are useful to provide power to more computers. Of course, the whole room of computer hard disks should be checked for viruses before the cleanup operation continues.

All infected disks must be correctly labelled with the user name, perhaps a person code, date checked and a warning to show the disk has a virus and needs scanning. This should be easy to remove so that when it is cleaned, the sticker can be replaced for an 'all clear' status, with user name, code, date and an OK.

There should be a technician present to install and configure PC hardware and software and to be on call for problems. There should be a few assistants to work with the chief disk checker, and a supervisor to manage the situation. This will apply to less than 100 employees within a company, but the amount of 5 people should double if there are twice the amount of computer user staff. The more people that can do the job with understanding, the quicker the operation becomes, but this would depend on the threat and scale of invasion, and resources at your disposal. It is therefore very wise to dedicate a room to virus emergency & recovery, but more often than not it will be returned to the support staff.

Stickers as identification are vital to prevent confusion over what has been checked and what needs to be dealt with. It allows project managers to predict cost estimates and time taken to complete the operation so that forecasts may be produced to vision when the anti-virus checks will be complete.

Initials must be written on the sticker, and dated so that reference to a media-checking database can be kept. A data input clerk could be employed to keep track of all viral infections, their origin, name, details about the virus, user infected, date, time etc. This will be useful to see if infection re-occurs and allow for future scanning with the user's home PC and home disks after tackling the work machine.

The truth with anti-virus strategy becomes clearer, the better strategy takes more time to initially create and possibly maintain. This involves additional man-hours required by virus experts to work out new ways of combating the problem and keep up to date with the technology. As discussed, the anti-virus sector is a small part of the overall security scheme. If you wanted to work a sound strategy, the best advice is to couple it with other security aspects like software auditing and network firewall protection. Up to the date protection and new virus outbreak alerts by fax are a good early warning system that something is amiss.

With these measures in place, it would be successful if the virus were eradicated, otherwise compulsory constant disk checks should be made for users disks. This is how you can deal with a problem of multiple ‘virus introducers’ which is more than one unwitting member of staff introducing the virus into the workplace.

4.0 Producing an Anti-virus tool

4.1 Feasibility Study

To complete a satisfactory solution that would benefit the anti-virus community, you become aware of the limitations to provide a useful product. Many packages already exist that give a better detection / cure ratio than a single person can ever hope to compete with. This is due to the sheer amount of people that work for the anti-virus companies, the disassembly of thousands of complicated viruses per year, and the top programmers employed to bring it all together.

The only way to compete with existing anti-virus scanner tools are to perhaps channel the scope of the program to one particular type of virus, and think of a new way of detection / cure that does not rely on constant virus study and ever changing updates. The most common virus threat today is the Word Macro virus, and they are very easy to study and understand as they are written in a language variant of visual basic.

Scanners already incorporate the detection and cure of these, but each new virus has to be hard coded for a search string match, and because there are hundreds of these virii and the rate at which they copy, scanners are no longer a viable solution to provide protection against them. The ease of creating a new virus by varying old code and the increase of macro virus construction kits also proves that evading the updated scanner is an easy task, especially with a macro virus construction kit.

The choice of operating system is an important decision, as it will determine which platform the macro scanner will work under. Since the nature of the program implies that an end user of the program will be using word for windows, it should be more intuitive and user friendly to code the scanner for the windows environment.

If the program code is to be created as a word macro template, it runs the risk of becoming infected itself! The prospect of macro names that are encrypted within the document, forces you to understand the structure of word files to decrypt the names, or write the program within a template to open documents from word. This would require a method to prevent the scanner template becoming infected and thorough testing to make sure it will work with a large collection of word macro Virii.

On further study of infected documents, it becomes clear that every different type of word virus has a fundamental flaw that announces its presence. When you look at the file with a disk-editing tool, you can see the ASCII format of the document, including the names of every macro used. Password protected documents scramble the text that the document contains, but the macro names still appear in plaintext. This allows a program to be written that searches for these names within any document or document template, and tell us if any dangerous macros exist within it.

This project will attempt to produce a generic detection program. This will warn the user if the document contains specific macros within the file that are most likely to be a macro virus. In order to do this, the program will analyse the names of the macro within a .DOC file for names such as:

AutoOpen, AutoClose, AutoExit, FileSaveAs, ToolsMacro (Etc)
These can vary depending on the language version of word, so the design should incorporate the user being able to change these search-strings by hand.
These are the macro functions that are altered or appended to so that a virus can manipulate the most commonly used menu commands and actions to copy itself. Upon studying the code within many different macro viruses, it has become apparent that these macro names are visible inside the document, and left unencrypted.

Because of these potential problems with a template scanner and due to the simplicity of avoiding structural understanding of the data files, the program will be coded with Visual Basic 5. This will allow the program interface to be designed quickly and give the program a wide choice of pre-written file and string procedures required to produce a scanner / remover.

Manual detection & removal of a word virus

Using two software tools: Norton Text-Search and Norton Diskeditor

At this stage you realise that it is possible to code a program to detect and remove a virus, because it can be accomplished using two software tools from Norton Utilities (a popular assortment of handy utilities to help maintain and use the PC.)

The first tool is Norton text search, activated by typing TS and then answering specific questions about the type of files to check and the search string to find. Specifying the type of file as .DOC and the search string as "AutoOpen", it is possible to find matches in documents infected by many types of word virus, although some will use other macro names to perform their infections. It is very handy because it searches the whole disk for documents, rather than just loading one as a parameter. This proves that a scanner can be coded, but targeted directly at this type of use.

The second tool is Norton diskedit, activated by typing diskedit. You can load the document found by text search into the diskedit window for direct analysis of the word document. Then using tools menu, then 'find' to search for the name of macro i.e. AutoOpen, AutoClose etc you can find the exact byte location of the first character used in the macro name. To disable these macros you simply change the first character. For example, changing from AutoOpen to xutoopen will prevent word from loading the macro automatically. This needs to be done for every automacro, to stop word from using them on normal operation of opening and exiting the word program. Stealth word viruses can be removed too by changing the macroname from ToolsMacro to xoolsMacro. If it's not renamed it will prevent you seeing the macro dialog box within word.

They will try and prevent access to the visual basic editor, and the macro organiser components of word, as they also allow you to view and remove offending macros. Without renaming these, it may not be possible to remove the more complicated viruses from word. Some stealth viruses will produce word basic errors if these macros are renamed, which leaves you to copy the document and paste it into a new one to remove disabled macros, and to change the structure of the file back to a document. A word virus has to convert a document to a template before it can copy, but it will still use the 'doc' extension and not the standard 'dot' abbreviation.

The word macro dialog box

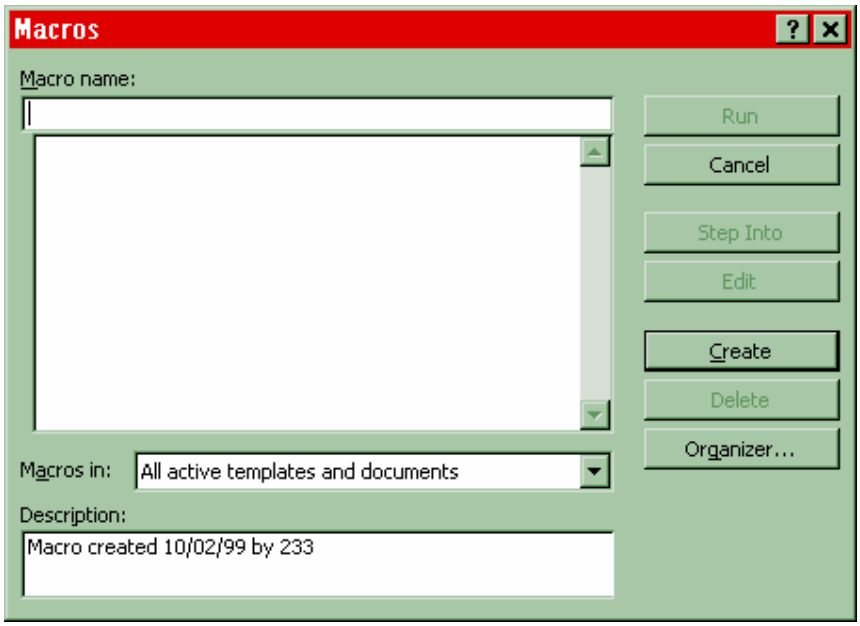

Fig 4.1 The macro dialog box allows you to view loaded macros with the ability to individually select and remove them.

This is the part of word that allows the user to see if macros are loaded, and gives you the ability to remove any offending macros. Of course, virus's author would rather hide this from you to make detection and cure more difficult. This is achieved by adding the macrocode within a macro called ToolsMacro, which does nothing:

```
Private Sub ToolsMacro()

End Sub
```

By renaming this macro to xoolsmacro, word will still load this code as a macro, but will not run when you select the macro option from the tools menu. This will allow for the complete removal of the virus from the word document. When macros are disabled, they still load into word with each infected document, taking time and diskspace. The second stage of virus cleanup would be to remove every macro via the macro menu which are not created or required by the user, then save the document to the same file, overwriting the file and thus removing the disabled macros.

Not all macro viruses attempt to conceal themselves this way, many of the macros used by the virus are visible from within word, and the programming code can be viewed and eliminated. Word viruses that prevent the code being seen or removed are known to be stealth viruses, but can still be detected and disabled by the GW-Scan virus scanner.

Here are some screenshots of actual virus macros visible within word's macro list:

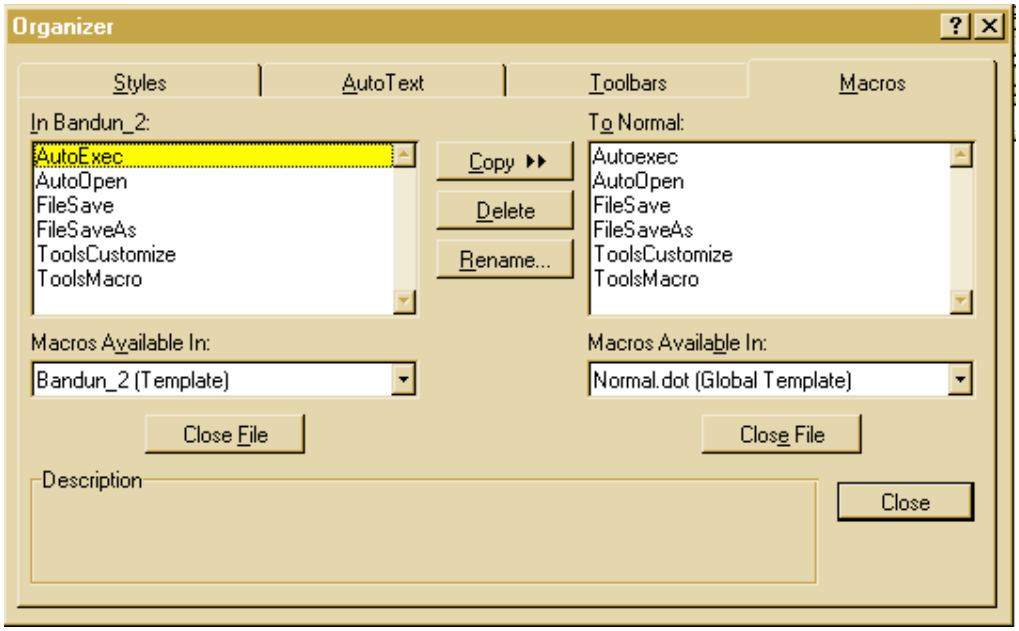

Fig 4.2 The Bandung virus macros viewed within the 'macro organiser'

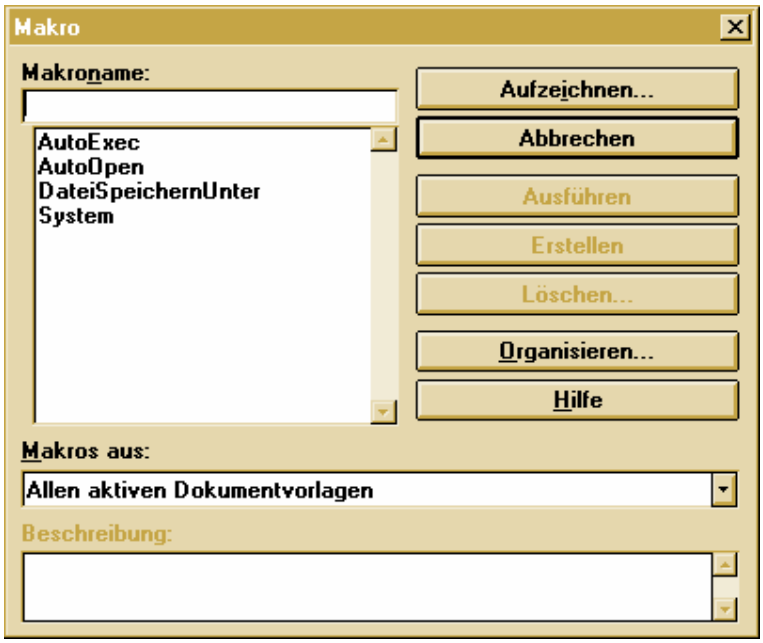

Fig 4.3 A German version of Word 6 allows the Boom Virus to be viewed from the 'Tools | Macro' option

4.2 Designing & coding of the program

There are many considerations to the design that allow the product to flourish. The interface must be intuitive and flexible enough to allow for customised scanning. Visual Basic is a nice choice of language because it allows you to get the look of the program before coding. The design must allow for lower graphics resolutions, and plenty of error handling to prevent the user from crashing the program.
Visual Basic provides a resolution finder, showing how each form will look at different resolutions, and the error handling is more than sufficient.

The initial stage was to piece together the modules to be re-used and how they communicate with the overall program. The step-by-step method of manually removing a document virus had to be constructed into a series of smaller tasks.
This allowed the interface appearance to proceed, automating file selection and string searches where appropriate. This is the initial list produced:

Basic features: File selection boxes with add/remove file buttons
Tree traversal routine to find all subdirectories and files on a drive.
File text search routine (Scanner)
File text replacement routine (Cleaner)

Extras: Main template replacement (Overwrite normal.dot)
File exclusion (User's macros that use search strings etc)
Help file
Splashscreen
Delay messages
Search String Editor

These modules were not written all at once. It was pieced together once the skeleton program worked, but this list helped the organisation of the interface and modules.

Data structures used throughout the program are list boxes, just like arrays. They allow data to be seen during the development of the program and hidden from view when the program has been tested. It also contains the regular string and integer types, but there was emphasis on using as few variables as possible to ease maintenance.

Algorithms for the search had to be replaced because it was too slow. The product used to read in a file byte-by-byte until a method was found that read the whole file in at once. This made a great difference between an unusable product and a useful one. An existing algorithm for traversing the directory tree was used for efficiency, it was exactly what was required and had a pre-tested track record. This was obtained from the VB programmer's code library CD.

Usability of the product is quite simple. There are minimal controls required to solve the problem without confusing the end user. The interface has been designed in a way that prevents you from causing file I/O errors (such as greyed out buttons) and the messages make the status of the program clear. The splashscreen can be removed by a click, or disappears on a 3-second delay.

Once the filenames to scan have been selected by the user, the program searches for a list of macronames that could be viruses. If these are found, it will search for another set of macros that help to remove it when the user presses the clean button.

The main problem with this type of detection is that of false alarm. The product has no way of distinguishing between an innocent file and an infected one apart from the macronames within the document. It cannot be expected that macros should not be used, because it is now a functional requirement that some documents carry out an automated process. The next stage of design is a file exclusion routine. This is a list of files that the scanner should not scan, because it contains user macros with the same name as some of the search strings used.

Search string editing allows for foreign macro virus protection if the correct names of the macros are inserted into the search text file. Upper and lower case conventions could be used when inputting via the keyboard, so these must be converted to uppercase before a comparison of search strings is made by the scanner.

Other design considerations are tidiness of the product. This includes colour schemes, layout and language used. The target audience has been considered to a beginner level, assuming knowledge of windows applications including word. The forms have been locked from change when the application is run. This prevents the user from altering the size of the form, changing text in the boxes and making the interface untidy. Screen positioning for the forms had to take different resolutions into account, so that boxes do not appear off the screen.

Simplicity of the program will determine it's use in the field, and attract a larger user audience willing to try it out and keep using it. The amount of options has been kept to a minimum to increase the speed of learning without a daunted user. The information presented is left uncluttered and readable, considering the spacing between the different controls used in the screens.

GNU Public Licence – The future development of this product would benefit by releasing the source and object code on the Internet for free. Other programmers would not be able to resell the software by law and they can make improvements as members of the general public. This is an open source policy, and would prevent the sale of the product. It is the best method to improve GW-Scan, and this will be the approach used for future production of the program.

Distribution of the program is important to gain an audience. The best method is to design a web page for the Internet that allows product description, downloading and links to other related sites of interest for anti-virus activists. The Winzip self-extractor has been used to install the product to the hard drive, and ensure that the correct file locations are adhered to preventing file I/O errors. The web page should be added to a search engine so it may be found, with relevant keywords that people may use to find such a product.

Design relies on common sense and redesign. If something doesn't look or perform well, it has to be re-considered and done again. Asking other users to test the product will ensure that the product is simplified, error resistant and user friendly. A trusted at on-site anti-virus expert at a large company, performed this role.

4.3 Testing

There were two types of testing required for this product.

The first stage of testing

This is testing the internal program workings of each and every module. They were tested with a series of dummy documents containing the macro names that were being searched for. Other modules required different tests, like the tree traversal routine to find all documents on a drive.

Files consisting of the macro virus search words were written into a document using word, saved into a file and then detected, cleaned and tested in word to see if the structure of the document was still intact.

Accessing the floppy drive without a disk, trying to clean read-only documents tested error handling, in addition to using the product over read-only network drives. Once these all proved successful, the product had more chance of a stable operation.

These tests were performed at each stage of writing the program, an ongoing process before the next module of code was written. It prevents many bugs entering the code leaving a nightmare debugging session at the end of development.

The second stage of testing

Testing the product's ability relies on the download of real word viruses and infecting a Windows 95 machine with a copy of Word 6, 95 and 97. A spare computer was used for this task, infecting it with one virus at a time. The scanner is run to find the virus, then to clean it if possible.

Before testing begins, a testing strategy will produce a fair investigation of the product ability. It has to be consistent for each and every virus examined, so that they all have the same conditions for the test. Here is the strategy used:

- Infecting word with the virus to test it's infection ability
- Removing the virus from NORMAL.DOT with the overwrite option in GW-Scan (Needs to recognise version of template and overwrite with the correct version)
- Detecting each file separately, and cleaning with the scanner (This will rename the macros effectively disabling them)
- Trying to remove the macros from the tools|macro option in word
- Checking the NORMAL.DOT template for changes
- If there are changes, the virus has escaped a clean-up
- If there are no changes, the scanner has disabled it

The viruses used were downloaded at random (their effects and complexity were unknown). The last virus on the test list was created especially for the project and contains many stealth and polymorphic modules from a virus kit. This produced a fair test to include a virus that is very complicated and difficult to remove.

4.4 Test results

A List of working Word Macro viruses tested with GW-Scan:

Macro Virus Name	Detected	Disabled	Comments
Carrier-b	✓	✓	Removed from word
Atom-c	✓	✓	Cannot remove from word
Bandung	✓	✓	Removed from word
Boom	✓	✓	Cannot remove from word
Clock-A	✓	✓	Cannot remove from word
Date	✓	✓	Cannot remove from word
Date-d	✓	✓	Removed from word
Divina	✓	✓	Cannot remove from word
Divina-d	✓	✓	Removed from word
Doggie-A	✓	✓	Removed from word
Dream	✓	✓	Removed from word
Edds-A	✓	✓	Removed from word
Gipsy-A	✓	✓	Removed from word
Imposter	✓	✓	Removed from word
Irish-A	✓	✓	Removed from word
Kid-a	✓	✓	Removed from word
Killdll-a	✓	✓	Removed from word
Mdma-a	✓	✓	Removed from word
Mdma-ao	✓	✓	Removed from word
Mdma-wazc	✓	✓	Removed from word
Metamorph	✓	✓	Removed from word
Nemesis	✓	✓	Cannot remove from word
Nightshade	✓	✓	Removed from word
Nopvirus	✓	✓	Removed from word
OPIM	✓	✓	Removed from word
Phantom-a	✓	✓	Cannot remove from word
Pheeew	✓	✓	Removed from word
Polite	✓	✓	Removed from word
Satan-666	✓	✓	Removed from word
Sparkle	✓	✓	Removed from word
Splash	✓	✓	Removed from word
Twno	✓	✓	Removed from word
Unpad	✓	✓	Removed from word
Unpad_d	✓	✓	Removed from word
Vampire	✓	✓	Removed from word
Zmk-b	✓	✓	Removed from word
Joke98	✓	✓	Removed from word
Rainbow2	✓	✓	Cannot remove from word
Nuclear	✓	✓	Cannot remove from word
Prank 98	✓	✓	Removed from word
DMV	✓	✓	Removed form word
Concept	✓	✓	Removed from word
BigBertha	✓	✓	Removed form word

Non-Threats

APRMS	not a virus (A Polymorph routine)
BDC	German Word Only (No threat in UK)

These results show a great success in the detection of word viruses, and an excellent clean-up ratio with the older types of word virus. The newer viruses are trickier to remove, but they were all detectable with GW-Scan.

By saying the macro virus is disabled, it does not mean that the virus is removed. Word97 will still warn the user that the document could contain macros because every Macro virus document is converted into a template. To remove these warnings, a further step of copying all text into a new document has to be carried out.
It does mean that the document will not be able to copy any more, and spread its infection.

The only virus that proved difficult to remove was created by a virus construction kit and the virus is named "BigBertha". This used the macro virus construction kit to produce a polymorphic, stealthed word virus that deleted all text in a document on the 13th day of the month. When cleaned, it did not allow the manual removal of macros within word. Instead, a work-around was achieved by copying all of the document text and pasting it into a new document. The original Normal.Dot template also had to be replaced by GW-Scan, but after these steps the virus was completely eradicated. It did not copy after disablement however, so did not pose a threat once GW-Scan had dealt with it. The source code for the virus is included before the source of Gw-Scan.

Replacing NORMAL.DOT prevented word-basic errors when you loaded word.
This is one of the options of GW-Scan to prevent these start-up error incidents.

A method to remove any of these viruses once disabled by the scanner is to add a toolbar to word, enabling you to go into Word Basic and delete the code for each macro. The option is under Tools, Customize, Visual Basic check-box (then OK), Select the design mode on the VB toolbar, then click on the view VB code button.

It is best to make sure this toolbar exists within your word installation early-on, because if you catch a modern word virus, the stealth options will prevent you from displaying this toolbar with a macro called ToolsCustomize

Button **A** enters you into design mode, button **B** can then be used to view macrocode.

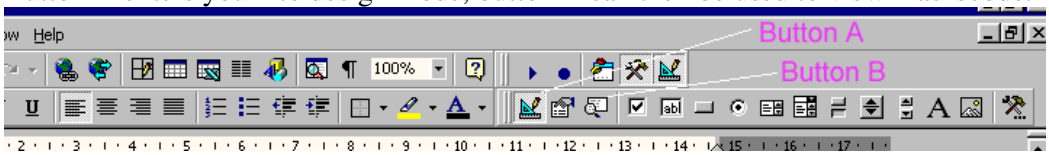

Fig 4.4 Word 97 VB Toolbar to help remove disabled macros.

Detection Failure?

Out of 45 word Virii tested with GW-Scan, two were not detected. The first was a German virus using German macronames, and couldn't infect an English copy of word anyway. The second was just a polymorphic routine and not actually a virus. If it does not hook onto any of the standard macronames, it does not become a threat within Word documents, as it does not copy.

As expected, the first virus can be detected if you add German macroname search strings (see section 4.6 – International Use) to the program. The second can be detected if you include the text 'AdvancedPolymorphic' as a search string. This gives a problem to a generic program that relies on standard macronames to assume the presence of a virus. The GW-Scan program will require an update to detect it unless you change the structure of the program code. Because both of these documents cannot infect an English /American copy of Word, they do not present a problem.

A change made to accommodate this was reading the macro search strings from a separate file, not inside the program itself. The ability to edit, add and remove the search strings from the file will allow the user to find macros like this one, without requiring a program update.

The only obstacle is to warn the user of the search strings these viruses use, which can be achieved with an E-mail distribution list, web site or a fax on demand service. The services contain the new macroname to scan for so the user can add it to their search string list. This is useful for national companies who share their documents between international sites and could unintentionally, infect them.

The foreign wordlist of macro search strings should get around this problem, and the product will be fully generic and should not require published search strings. The only drawback is finding the correct list of macros for German, French etc for over 50 different countries.

4.5 Limitations of the program

False Alarms

Because of the generic nature of the program, there are two possible conditions that will trigger a false alarm (the scanner suspecting a virus when there isn't one).

The first condition is when a document contains one of the macro search strings in the normal document text. This document contains several occurrences of the search strings that GW-Scan uses. These will trigger as a virus alarm, and if cleaned, the text within the document will be changed from Auto… to xuto… for example. Therefore, documents that describe macro viruses may not harbour a virus - they just appear to. If the macro section of the document can be distinguished from the document text, this technicality is resolved but requires structural understanding.

The second condition applies when there are innocent macros in the document created by the user or a third party that use the same macronames as the search strings. Using template add-ins allows the use of different macronames and may prevent this.

In both cases, files can be excluded from the scanning process by choosing which filenames to leave alone. The scanner checks to see if the current filename to scan appears in the 'exclude files' listbox and if found will skip to the next file to scan.

This can make the cleanup process more difficult when a macro virus is found. If the excluded document becomes infected it will not be checked by the scanner and will re-introduce the virus if used again. Keeping a record of the excluded files will help identify which macros were created by a user, and which are added by a virus.

A good precaution would be to keep copies of the macrocode in separate non-MSWord format (text file .txt or basic file .bas). After overwriting the Normal template with a fresh copy from GW-Scan, using Word97 you can disable macros when you load the file, then put them back in manually from the .txt or .bas files. Obviously, as more files are excluded from the scanner, the manual work required on those documents will increase.

Removal Failure

More sophisticated word viruses use classes to execute their macros, so that names of the macros can be different from the standard ones. There are sometimes problems when macros are renamed within a file, but these just appear as word basic errors when you load the document. The document text appears when you have pressed the OK for the error, and you can select the whole document and cut & paste the contents into a fresh document and save it. This removes the macros from the document.

Future versions of Word

If Microsoft decide to change their macronames, structure of the document or produce a different NORMAL.DOT, the program will have to be updated to accommodate these changes. They may change the document format in such a way that prevents a generic approach altogether. This limitation cannot be foreseen, only time will tell.

The project is aware of the Office 2000 application suite by Microsoft. More research is needed to understand the structure of the document when macro names are stored, and which version of word is reported in the file. A copy of Normal.Dot from the Office 2000 suite may be required for a global template overwrite to function correctly.

User Interaction Vs Automation

The project can extend into a TSR version of GW-Scan, so less user operation is required. People like things done automatically, because it's one less thing to worry about during the working day. It will also make sure that the checking of documents actually gets done.

This will be an aim beyond the scope of this project, but not beyond the scope of the product idea. It will need a complete re-write to achieve this, and testing needs to be thorough to prevent any interference to or from the other programs in memory. This will make the product more marketable and popular. If the product were written again, it would take the form of the TSR strategy used by the existing anti-virus competition.

4.6 GW-Scan Improvements

Here is a list of possible ideas to extend the project, this list is by no means complete but gives an insight into the depth of possibility that a word macro virus scanner could incorporate and use. There are also additional tips to improve the security of Word.

File association changes

File association is the term used to explain which particular software is loaded to open a file type. For example .TXT files are opened with the Windows notepad, .DOC files are opened with WordPad or Microsoft Word if installed. If the entry for .DOC files pointed to a special version of my virus scanner rather than Word, it could scan the file for dangerous macro names and warn the user of their presence. The user would then decide if they wanted to load the document, change the macronames, delete the document, or load the file into a viewer for manual analysis. This would protect the user from documents attached to e-mail, without any interaction required.

Software Security Patches

Although Word97 has a macro-warning feature, it does not provide any solution to prevent a virus from disabling the option, without installing extra service packs. There are now service packs SR-1 and SR-2 which must be installed to add to the stability of the office suite, and a separate patch to prevent newer viruses changing normal.dot without the user first entering a password. (Included on the last page)

With the previous trends in Microsoft security, their legal situation on exportable cryptography limits the password protection and renders it vulnerable to future attack. It would be easy to bypass the protection if the virus created it's own normal template! Password protected documents can also be unprotected, in addition to their password being decrypted by a third party. Using third-party strong encryption solves this problem.

International Use

The international profit to sell such a program is possible when the product can use their language. This is the perfect opportunity for the word virus scanner to be shipped with a variety of language files that contain the text strings & images used throughout the program. Users select their language on-screen and may choose the name of their country or click on a picture of their national flag. The word macro viruses exist from many language versions of word, and in their country the names of the search string macros are usually different. Searching the document for macros in different languages will detect most word viruses across the globe.

The structure of the program needs to accommodate this by storing all internal messages into external language files for each country. The same will apply to the search strings, but kept in a separate file from the text to allow user changes to the search strings used. Reading an initialisation file on start-up tells the program which language to use, depending on which was selected during setup.

5. CONCLUSION

Extra Considerations

A series of ever-changing factors will continue to effect the anti-virus world, with further solutions necessary to solve new problems.

- The future

The increase of more computer viruses to come will reflect on the ease of which they can now be coded. Virus generators are readily available, to assist new inexperienced programmers to build a devastating virus causing damage to data. Although the best way to defeat the virus scanner is to build a very basic generated virus, and add to the code using modules from other viruses. It will then be capable of identical or altered operations from the original virus.

Documents containing word macros will continue to spread and hide themselves well. If Microsoft took the initiative to write a decent generic macro protection tool, the problems may disappear over time.

A project such as this may extend to achieve these goals, but first must delve deeper into the data structure of the document. This project will not stop here, it will continue to be a quest for useful research and tips to improve efficiency.

- The size of anti-virus software

In the advent of a wide selection of new viruses, the anti-virus software will grow in disk space required. To identify and repair a virus, the new antidote code must be added to the virus definition database where records of all other viruses are kept.

Here it is possible to slow down the scanner or have greater hardware components to run it. The suggested method to keep the virus checking quickly is to upgrade the current computers and use some of the fastest and affordable technology. Many companies have already phased out the early Pentium, and replaced it with Pentium II or the recent Pentium III. Large amounts of memory are required for faster scanning, to limit the use of the windows swap file. Use of SCSI hard disks also reduces time.

- The law

Unfortunately, the law will not be enough to prevent the threat of a virus, but it has previously scared people from committing the offence and will punish those it finds accordingly. Christopher Pile from Plymouth was convicted for writing & spreading the Smeg.Pathogen and Smeg.Queeg viruses in 1996, serving an 18-month prison term. The acronym stands for Simulated Metamorphic Encryption Generator, and it's complicated polymorphic routines caused huge problems for many anti-virus companies who were unable to detect it reliably even months after it was released. Due to the anonymity of advanced computer hackers, this will not prevent the problem but may reduce the amount of people willing to try it.

- Success of research and Anti-Virus Policy

The mystery of how a company can defend themselves against this type of security attack has been revealed, but does require constant work to keep it running smoothly. The attitudes of management and staff have to be right for such a policy to work, but the guidelines produced have proven to work for many other companies with a range of skilled and unskilled computer users.

The information provided about viruses should be a good enough introduction as to how they work, where they come from and how to deal with this problem. It requires experience and calmness to deal with these problems, which again is a staffing issue. With these people in place, and co-operation from the users, the guidelines will produce a sound policy and prevent the majority of computer virus attack.

- Meeting objectives of the GW-Scan product

This product works as intended, even if the macros do need removing afterwards. It acts as an early warning system to the hundreds of new word macro viruses appearing every month, and helps remove it before virus-scanners are able to cope with it a month later.

It is clear that the product is excellent as an early warning system for a new type of word virus, but not to be used as a defensive product. Coupled with a real-time TSR scanner like Mcafee or Dr Solomon, it would be useful in detection as well as an emergency clean-up for new and unknown word macro viruses.

The users do need to read the documentation properly to understand that office includes templates with automatic macros, and a list of excluded files needs to be generated for every computer that it's set up for. If this is not done, and installed as a regular anti-virus scanner – it will cause panic and distress when the user believes that their office templates contain a virus. As a result of this, a new copy of exclude.txt was generated which excludes most office templates and the password protection for the normal template (protect.dot). The conversion of automacro names described in this document from a normal text format to a bitmap prevents the scanner from finding a false alarm in this document!

The original goal was met, and a steep learning curve was achieved with Visual Basic, computer viruses and word macro virus code. It has been a very enjoyable project to work with, and more research will be undertaken in this field.

Bibliography

Paper based resources

Computer Viruses – A High Tech Disease
Ralf Burger
2nd Edition 1988
ISBN 1-55755-043-3

Dr Solomon's Virus Encyclopaedia
Alan Solomon
Edition 5.1 May 1997
ISBN 1-897661-16-9

Dr Solomon's Virus Encyclopaedia
Alan Solomon
Edition 2 February 1992
ISBN 1-897661-00-2

Computer Anti-Virus Strategy
Gregory Charles Day
July 1996
Final Year Project (Business Information Systems)
Goldsmith Library, Milton Campus

On-line resources

Dr Solomon's Anti-Virus Toolkit – http://www.solomon.com

Provides information, latest attacks and threats, virus descriptions database, virus alert e-mail subscriptions and allows an early warning fax alert service to be purchased. Now merged with Network Associates (producers of Mcafee) this website will be merged will Mcafee and may cease to exist before long.

Mcafee Anti-Virus information - http://www.mcafee.com

Contains advertising and virus alert e-mail subscriptions. This is a very useful site that alphabetically categorises virus names and description of their effects. All Mcafee software can now be downloaded from their ftp site using the password provided by Mcafee upon purchase. The new version of the scanner uses the award winning Dr Solomon's scan-engine.

Urban legends E-mail hoaxes - http://urbanlegends.miningco.com/library/weekly/aa030798.htm

E-mails are sent about viruses that don't exist just to scare people and create a false chain letter. This page was produced to tell you if a new hoax exists giving you time to warn everyone in your company that they shouldn't be alarmed by the e-mails.

Anti virus procedures - http://www.ey.co.zw/antiviru.htm
http://www.cai.com/virusinfo/policies.htm

Guide-lines to help you implement and maintain your own anti-virus policies. Explanations of equipment, staff, strategy and tools required doing the job.

Epulse Virus risks http://www.epulse.co.uk/virus.htm

Virus information about the risks that viruses present to any organisation, and preventative measures to combat the problem.

National Computer Security Associate (NCSA) – http://www.ncsa.com

The Computer Virus Prevalence Survey provides useful statistics on the costs and effect of computer viruses, the rate of increase over recent years and which viruses were the most popular in a particular era. It is a useful source for graphs.

Appendix A

Glossary of Terms

BBS	Bulletin board system – a computer connected to a modem that allows other computers to dial into a menu system, for information and file transfer. E-mail services and individual user areas are possible.
BIOS	Basic Input Output System – A chip inside the computer controlling the screen, disk, keyboard and other functions of input and output.
CMOS	Complimentary Metal-Oxide Semiconductor – the type of chip that the BIOS uses, and can store date / time, hard disk sizes and other functions that the user selects. Settings are kept for the next bootup while the PC is switched off by using the battery on the motherboard.
Encrypted	Data within a file that is scrambled to prevent casual viewing.
FAT	File Allocation Table, contains filenames, starting sectors and filesizes so that the operating system can locate its programs and data.
In The Wild	Describes a virus whose infection has been reported at least once outside of a computer virus research laboratory. Some viruses only exist within the anti-virus community, and are not seen elsewhere.
Modem	MOdulatorDEModulator – A method to connect two computers via a phone line through a hardware card or box connected to both.
Motherboard	The main circuit board that bridges PC components together.
PC	Personal Computer (assuming the standard Intel architecture).
Plaintext	Data within a file in a normal unencrypted state.
SCSI	Small computer systems interface, which provides faster data transfer for data storage. Needs a SCSI card unless already on the Motherboard.
Virii	The plural to “virus” meaning more than one virus i.e. a collection.
Viruses	Also a plural, now the adopted standard but still also known as Virii.

Appendix B – Virus Statistics

Anti-virus companies and third party organisations that convey where the most common problems lay publish these statistics. It can help predict how many people are needed to cope with the increase of viruses within the anti-virus industry, and the human resources to be allocated to particular sectors of the company.

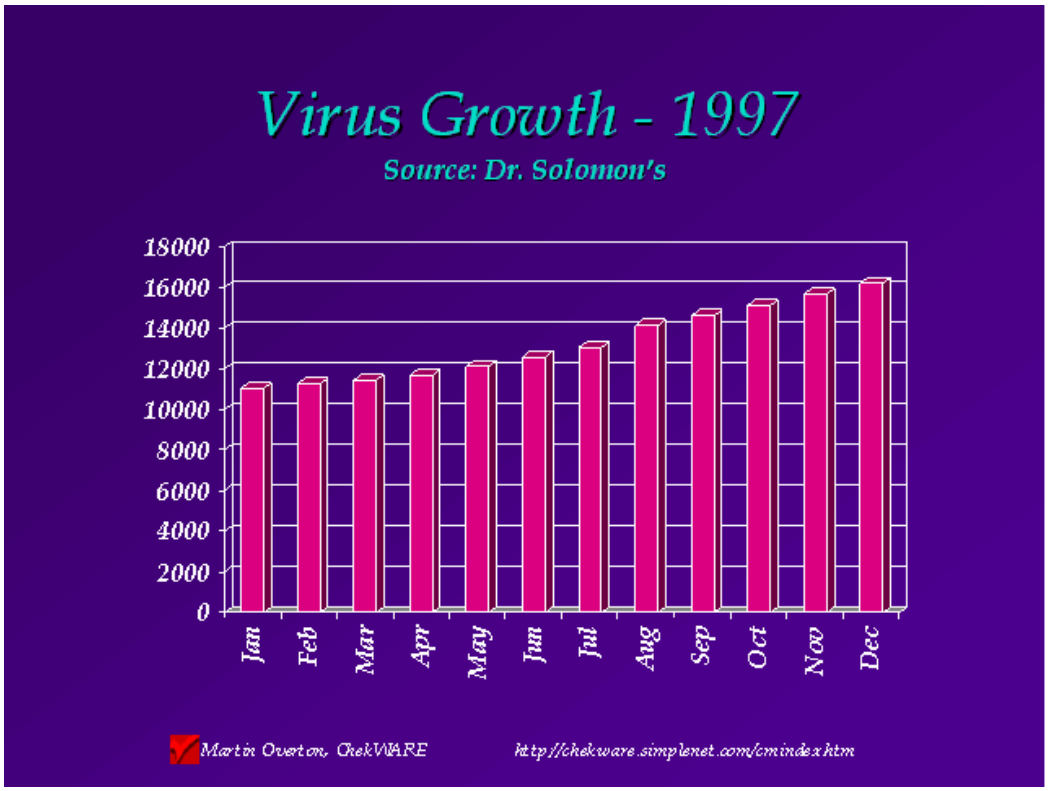

Fig 5.1 This shows a steady virus increase, a problem known as 'glut' to many anti-virus researchers meaning increased work required to incorporate the scanning and repair code to their scanners.

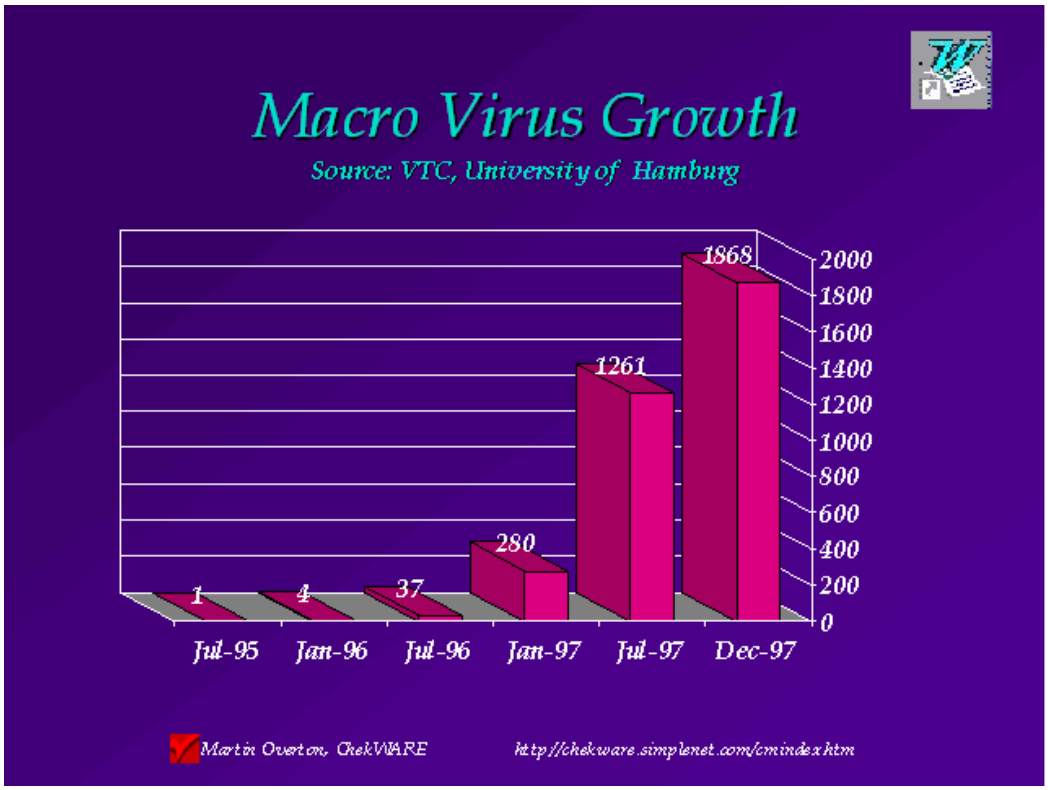

Fig 5.2 An overwhelming increase of the macro virus has arisen since the very first Concept virus in 1995. Now the biggest virus problem there is.

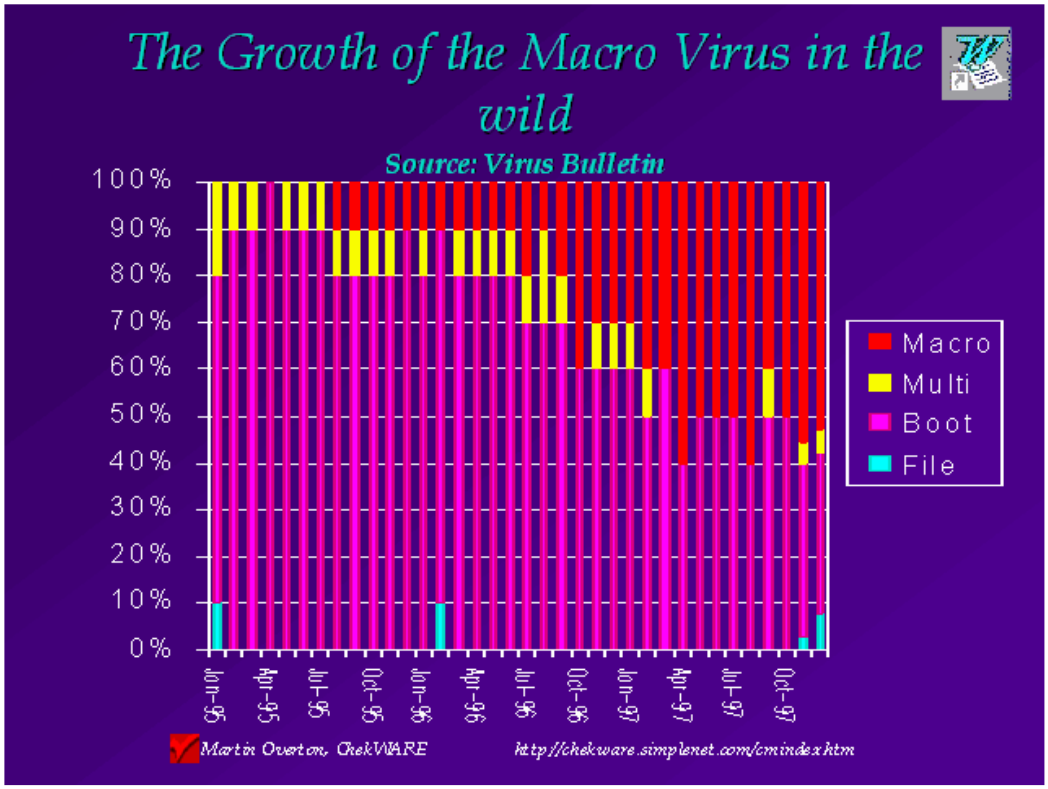

Fig 5.3 The reported viruses have a breakdown of different types. Clearly the macro virus is on the increase, not just new strains but reported cases of infection in the business and home environment.

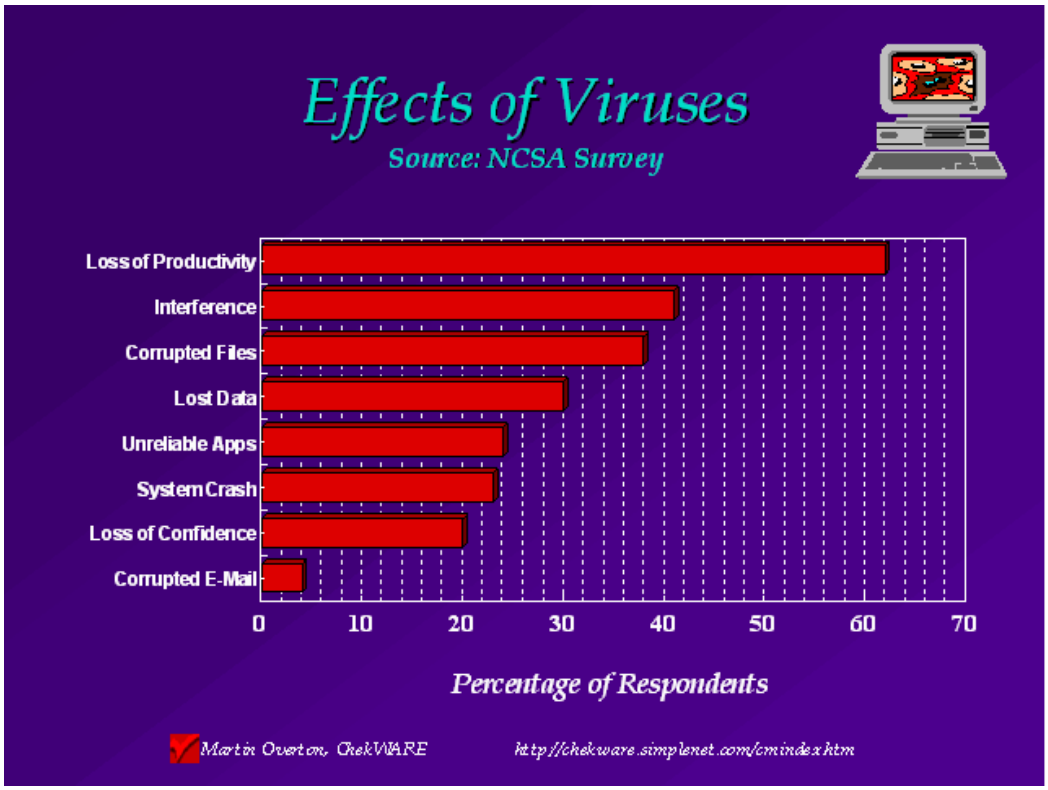

Fig 5.4 Loss of productivity may be due to the amount of downtime the machines need as a result of a virus clean up.

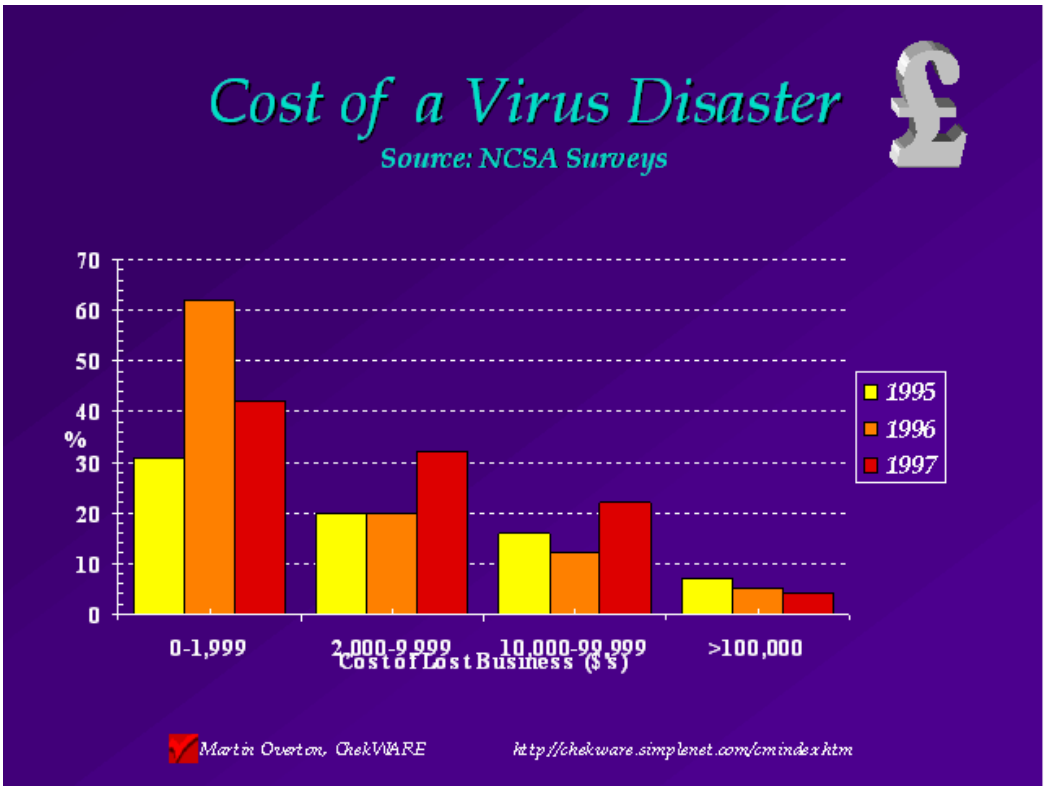

Fig 5.5 Dependant on the type of business and the nature of required computer dependency, the cost of a virus differs from business to business.

Appendix C

Further Reading

Here is a list of links and books available to order from any good bookshop. This will cater for many different types of detection, prevention and understanding. An alternative way to read these without purchasing them is to go to the local library and have the receptionist search for a range of book titles with the words "computer virus". They can be borrowed from other libraries and lent to you for a few weeks.

Anti-Virus Knowledge

Computer Virus Prevalence Survey
By the National Computer Security Associate (NCSA)
ISBN 0-7881-3726-3
Costs $30 printed copy or free in PDF format from the NCSA website:
http://www.ncsa.com
The NCSA provide useful information about which viruses are 'in the wild', and the statistics for cost and totality of various viral infections.

Robert Slade's Guide to Computer Viruses : How to avoid them, how to get rid of them, and how to get help (Second Edition) Springer, 1996 ISBN : 0- 387-94663-2

Virus: Detection and Elimination. Runem Skardhamar. AP Professional. 1996. ISBN : 0-12-647690-X.

The Giant Black Book of Computer Viruses. Mark A. Ludwig. American Eagle. 1995.

The Computer Virus Crisis. Fites, Johnson, and Kratz. Van Nostrand Reinhold Computer Publishing. 1988. ISBN: 0-442-28532-9

European Institute for Computer Anti-Virus Research. http://www.eicar.com

Computer Virus Help Desk. Courtesy of the Computer Virus Research Center. Indianapolis, Indiana. http://www.iw1,indyweb.net/~cvhd/

A Guide to the Selection of Anti-Virus Tools and Techniques. W.T. Polk and L.E. Bassham. National Institute of Standards and Technology Computer Security Division. Friday, Mar 11; 21:26:41 EST 1994. http://csrc.ncsl.nist.gov/nistpubs/select/

Virus Knowledge

The Virus Creation Labs: A Journey Into The Underground. George Smith. American Eagle Publications. ISBN : 0-929408-09-8. Also reviewed in Net Magazine, Feb 96.

Future Trends in Virus Writing. Vesslin Bontchev. Virus Test Center. University of Hamburg. http://virusbtn.com/OtherPapers/Trends

Appendix D - **PROGRAM OBJECT CODE**

EXECUTABLE COPIES OF THE SOFTWARE:

OLE (OBJECT LINK EMBEDDING ALLOWS GW-SCAN TO BE LINKED BELOW THIS TEXT)
DOUBLE CLICK ON ALL OF THE LINKS BELOW ONE AT A TIME AND FOLLOW THE PROMPTS TO INSTALL GW-SCAN version 2.2.0 – INSTALL PROGRAMS IN ORDER 1,2,3

1. VISUAL BASIC 5 RUNTIME LIBRARIES
Visual Basic 5 Run-time Libraries

2. GENERIC WORD VIRUS SCANNER VERSION 2.2.0
GW-Scan 2.2.0 Object Code

3. GW-SCAN PROGRAM SHORTCUT COPIES TO WINDOWS DESKTOP
Desktop program shortcut to GW-Scan

4. GW-SCAN VERISON 2.2.0 SOURCE CODE – TRULY BRITISH!
GW-Scan 2.2.0 VB5 Source Code

A professional installation program has also been created, available on request.

Microsoft Password protection on the main template means that you must enter a password before you use Word. This prevents the casual user from risking an infection on a colleague's computer. Click on the link below to load the document.

Password protect Normal.Dot

How to obtain the latest installation program

Please contact The.Invisible.Man@Cyberdude.Com **to receive an e-mail from the author that links you to the latest version. There will be a dedicated web site to this project in the near future. All users who e-mail the author requesting the notification of the new web site will be added to the program update e-mail distribution list to inform users of changes to the product. The program is already released as an open source program, the latest source code is available for free on request for the purpose of improvements and new ideas.**

Please read the disclaimer carefully before using the project media.