Study on data protection for used hard disks

Data, Data Everywhere

Dipl.-Inform. Olaf Kehrer
O&O Software GmbH, Berlin – April 2004

Nothing is more important than protecting data from unauthorized access. Thanks to extensive reports in the media, most PC users are now aware of the dangers presented by viruses and Trojans. Do they also know that deleted data can be restored to PCs with widely-available software? Do they act on this knowledge and delete data properly? This study tries to answer this question and comes to the conclusion that both private and business users are extremely careless about their private and vital data.

In the year 2004, Europe is moving into the age of purely electronic communication. Letters are now sent by email rather than post. Many Europeans carry out all their banking activities over the internet, as it's more convenient and often cheaper. The German government has signaled that electronic ID cards and medical records are the next step on the way to a secure and reliable data exchange, for the good of all citizens. At the CeBIT, the world's largest IT fair, German Chancellor Gerhardt Schroeder announced that these changes would take place as early as 2006.

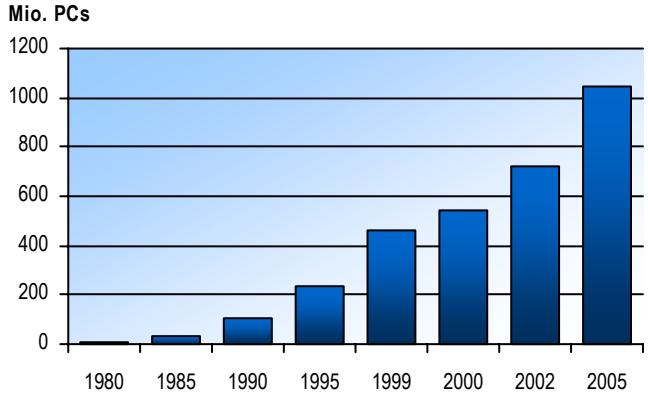
One of the most important tasks for today's information technology is the protection of data against unauthorized access. The number of viruses is increasing daily. These viruses attack anything that is accessible over the Internet, from private PCs to large processing plants. Most users protect themselves with anti virus programs, firewalls and other applications in order to make an attack on their computer as difficult as possible.

The appearance of the ILOVEYOU virus in mid-2000 brought these dangers into the public eye. With Windows XP and Windows Server 2003, Microsoft has built in far more security measures against such attacks than ever before. Many protection measures against viruses and Trojans have now been integrated, and Microsoft Outlook – pretty much the standard email software – is now much better protected than ever before.

But what happens to the protection of data when users decide to replace their PCs? Are hard disks really wiped before the computer is sold or given away? How aware of the danger are today's PC users? Do they have any idea how easy it is to restore data they thought they had erased?

To answer this question, the Berlin software house O&O Software GmbH bought 100 hard disks on eBay. These were then examined and tested to see if data were still accessible and how easy it is to restore such data.

Fig 1: Numbers of installed PC systems worldwide



Source: COMPUTERS-IN-USE FORECAST, eTForecasts[5]

The real #1 danger

Of course, all PC users are aware of the dangers presented by malicious virus programs that infect the home PC and perhaps publish private data for the whole world to see. Almost everyone is now aware of this danger, and more and more programs are being written in order to stop it in its tracks. Constant updating of such software is now almost an everyday activity for the PC user.

There is still one danger that no software can prevent: user carelessness. This is particularly observable when dealing with data when the trusty PC is replaced by a new one. In this day and age, PCs are not simply thrown away. In many cases the computer's fate is a completely different one – the eBay auction.

Everyone should be aware of the need to delete data from the computer before it is passed on someone else. Nobody wants their personal documents to fall into the wrong hands. The hard disks are therefore formatted and – look! – all the data is gone. The operating system is no more; the computer refuses to work once started up. Everything is just as it should be. Or is it?

No-one would throw letters and documents that contain important or personal data into the waste bin. People tear them up or, even better; turn them in to confetti with the help of a shredder. There was a time when such machines were only found in offices, but now they can be picked up for a few dollars and are used in private homes. This is understandable – who wants their neighbor finding bank statements, love letters or documents about termination of employment?

More than 150 million new PC systems sold worldwide in 2003

According to eTForecasts, more than 1 billion PC systems will be in use worldwide by the year 2005 (fig. 1). The Gartner Group and IDC report that more than 150 million PC systems were sold, 7.5 million of those in O&O's home country of Germany. Most of these systems were replacements for old ones. Most of the systems being replaced were not broken or faulty in any way, but simply were not up to the rising IT standards we know today. Most of these old systems were sold on or simply given away. The computers in question are not just private systems for home users, but company computers. The employer replaced them and sold the old ones on in order to cover some of the cost of buying the new systems. Often a part-exchange is possible, or the computer dealer simply takes the old systems off one's hands. Sometimes the computer itself is taken apart and destroyed, and the valuable components are then sold on individually at a profit. These often end up on eBay, where anyone can get their hands on them at a fairly good price, in order to replace components for his own PC or if he is putting the PC together himself. In this way a circulation of PC components is created – components only leave this cycle if they are hopelessly old-fashioned or broken beyond repair.

Sensitive data – on eBay at bargain prices

What happens to data on old hard disks and other data carrying media when they are no longer required by their original owner? Is the data properly erased, or does the original owner think that formatting the hard disk is enough?

In January 2003, two researchers from MIT (Massachusetts Institute of Technology) published a study which involved the purchase of hard disks from eBay and the subsequent examination and search for restorable data. The result of this study was that the majority of hard disks still contained data – from private correspondence, to data from an ATM including customers' bank transactions. [7]

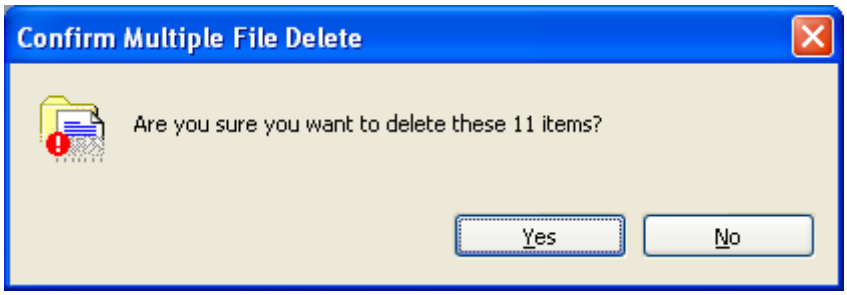
Has this study with its shocking results lead PC users to change their behavior? After over a year, have people reacted to the report's findings?

In order to answer this question, O&O Software GmbH bought several hard disks over a period of a few weeks at the beginning of 2004. These hard

disks were chosen completely at random. Functioning hard disks were bought as well as some that were broken and unusable.

These hard disks were examined with the help of widely-available software in order to see if they contained any restorable data. The result: from the 100 hard disks, only ten were securely and properly wiped. The other 90% all contained data from the previous owners – from illegal software to MP3 music files, to personal bank account details and love letters.

Fig. 2: Emptying the reycling bin doesn't really delete data



Medical patient data and business documents

Our biggest find was a hard disk that had obviously been used by a German health insurance company. It contained not only internal staff emails, but also internal data about customer claims for medical treatment along with the correspondence with the relevant doctors and medical data about the patients.

This data would have allowed an unauthorized person to discover patients' addresses and use this information for personal gain. The release of this information can of course have legal consequences.

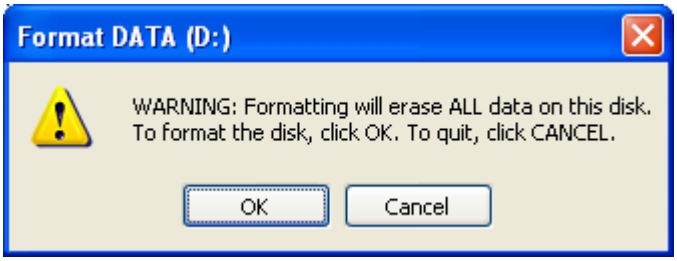
The search for reasons

Once we had evaluated the results of our study, we asked why users do not correctly delete their hard disks so that the data cannot be restored.

Many are simply not aware of the danger, as they believe that deleting data is the same as destroying it once and for all. The recycling bin in Windows is called that for a reason. When files are deleted via the Windows explorer, one can restore them easily from the recycle bin. When the recycle bin is emptied, Windows leads one to believe that the files will then be securely and finally erased. (fig. 2)

The same is true when the hard disk is formatted. The warning that Windows displays leads the user to believe that all files on the disk will be permanently destroyed. (fig. 3) However, this is not the case. Windows simply rewrites the boot sector and creates a new main directory. All other data remain and can be restored easily.

Fig 3: The warning before the formatting is misleading



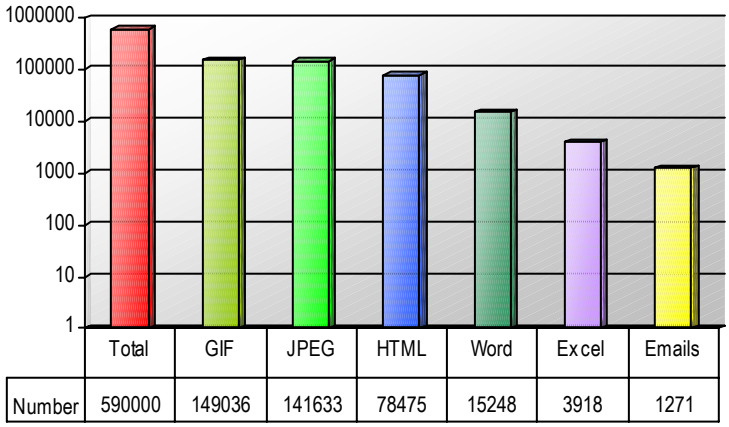
One further reason was discovered during the study – hard disks that are faulty and therefore cannot be used in the normal way, as Windows does not recognize them. These disks are simply discarded and replaced with new ones. Earlier, these disks probably ended up in the rubbish bin, but nowadays one can always find a buyer via eBay. Even these hard disks can contain secrets – all it takes is basic technical know-how and

Fig 4: Types of files found

In total, 590,000 files could be restored. Most of these were graphics and internet sites. (GIF, JPEF, HTML).

Over 15,000 Word and almost 4,000 Excel documents were readable. WE could also read 60 Outlook mailboxes with a total of 1271 emails.

The y-axis is given by the logarithm.



	Total	GIF	JPEG	HTML	Word	Excel	Emails
Number	590000	149036	141633	78475	15248	3918	1271

information the user thought was gone forever can be restored.

The ignorance surrounding the whole problem cannot be underestimated. The sellers of used hard disks know that the new owner can read data from them, but do not see this as a serious problem. During the course of the study, we received three hard disks and were ready-to-use and without the slightest fault. Do these people also sell on their plastic folders without taking their documents out of them? This is a case of gross negligence, as all data can easily be read here even without restoration software.

186 GB of data found

100 hard disks were bought with a total capacity of 526 GB. Fifteen of these were technically faulty, and were therefore disregarded, as restoring data would have been a disproportionately large effort. Only ten of the hard disks had been properly wiped; all others could be read either immediately or after a partly automated restoration process (fig. 5)

All together, over 590,000 files with a total size of 186 GB could be restored and read, including 15,248 Word documents and 3,918 Excel files. Furthermore, 60 complete Outlook mailboxes (PST files) were found, including the email correspondence.

No data that we found had been encrypted in any way. They could be opened and read simply by Word, Excel or Outlook.

Looking into PC users' filing cabinets

Most data came from private users who had saved their personal documents, pictures, photos and emails. With the help or restoration software, it would have been possible to fully take on the PC user's identity. The following files were found on one hard disk:

- The scanned-in signature of the PC user
- Job applications and résumés
- Scanned-in employee testimonials
- Scanned in ID card and ATM card
- Legally binding bank instructions
- Internet and email passwords

Another user had scanned in his PIN letter for his online banking account, so that it would have been possible to carry out online transaction in the name of the PC user. Of course, we did not attempt to do

this. The following data were also found and could be restored:

- Court judgments and prison release documents from the previous owner
- Over 10,000 MP3 music files, various illegal software copies, original licenses for software products, etc.
- All kinds of private emails
- Pornographic material of all descriptions.

In summary, not only private and sensitive information was found, but also information that could attract the attention of the criminal authorities. No-one would let a complete stranger at their private filing cabinet in order to look at the private files there. But these hard disks contained exactly the same sort of information.

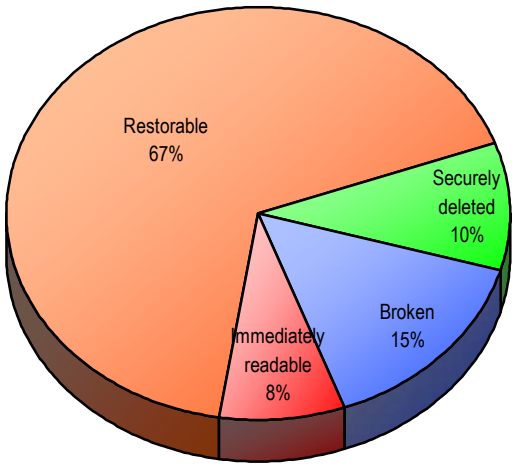
The health insurance hard disk with patients' medical information

It is stupid enough that private users carelessly – and unfortunately also accidentally – pass on their personal data. But the data we could restore from company computers could have even more dramatic repercussions.

We were able to obtain a hard disk previously used by a clerk at a major German health insurance company. It contained the entire email correspondence of this worker – internal as well as external correspondence with doctors, including the relevant medical data belonging to the patient:

- Email archive, that could be opened and read
- Letters to doctors declining to pay for suggested treatment, including all medical data and health insurance numbers
- Internal strategy papers and working practices, for example concerning the level of cash reserves at the health insurance office
- Parts of the internet as an offline copy

Fig 5: 75% of the hard disk was readable



This hard disk was the first one to reach us. However, it got even worse: a few days later we received another hard disk from another seller from the same health insurance company, but another office. Again, sensitive and important data were restorable, for example about the distribution of extra payments according to an internal formula. This information was explicitly marked "highly confidential".

As early as 1997, John Markoff in the New York Times reported the case of a woman who bought a used PC and found 2,000 private medical records from a pharmacist on the hard disks. [3]

It is clear that the implications of this article have been ignored in Germany, although – or perhaps because – it was published a full 7 years ago.

Strategy papers and other company secrets

The other company hard disks contained large numbers of confidential, vital and even strategic files which could cause a great deal of damage of they had fallen into the wrong hands.

For example, strategies were detailed that could make use of the weaknesses of a large national transport company in order to gain business advantages.

We also found internal bonus systems, cost details, and even internal communications from a board member from a global pharmaceutical company detailing the turnover breakdown for the previous year and the business goals for the following year.

Also amongst the hard disks was one from a head-hunter – again, the data were readable. As well as the names, positions and addresses of the target employees we could also read their profiles and the salaries new employers were willing to pay.

The list goes on and on. There was hardly any type of file we didn't find. We even found a letter from an O&O employee from our partner support department to a prospective partner on one of the hard disks. At least now we know our communications arrive.

Preventative measures

Before one can even talk about appropriate preventative measures, one has to be clear about where the danger of restoring deleted data comes from. In Appendix A, we have given a detailed description of the technical background to this process. We will therefore simply summarize here by pointing out that it is theoretically possible to restore even data that has been overwritten. This does involve a degree of technical skill and effort, but for around $2,000 hardware is available that will do the trick, with the relevant software thrown in.

In the next paragraph we describe a few possible ways of protecting oneself from data spying.

Data encryption

One of the most elegant ways to protect data is to encrypt it. This means that all data are saved to the hard disk in encrypted form. Only when a username and password is entered is access to data granted.

With Windows 2000, Microsoft added EFS (Encrypted File System) to the existing file system fort his purpose.[4] In Windows XP, this is only available with the Professional Edition, so home users don't benefit from this technological advance. They have to purchase and install this software separately. The installation and configuration process can be so complex that many users give up. There is also the problem that if the user forgets their password, the data become inaccessible. Users are more afraid of this than of their data falling into the wrong hands.

The main advantage of data encryption is the fact that the data are permanently protected, even if the PC is stolen. It also involves relatively little time and effort on the part of the user, as the operating system automatically encrypts and decrypts once the user is successfully authenticated.

Encrypted data can be restored just like all other data, but they are unusable, as he information cannot be read. Users who want even more peace of mind should choose one of the following options.

Physical destruction of the hard disk

The physical destruction of the hard disk is one of the most secure methods. This ranges from the demagnetization of the disk with industrial electromagnets to the drilling of holes and the smashing of the disk, there are various ways of doing this. All of these methods have one thing in common – the hard disk is then unusable and belongs in the waste bin. However, the costs are relatively high and the processes cannot be carried out in the garden shed without serious safety risks.

In many companies, the physical destruction of the hard disks is out of the question, as they are leased together with the rest of the PC. The whole system must be given back in one piece at the end of the lease.

Software for the secure deletion of data

The cheapest, least complex and most effective method of secure data deletion is that of specialist software. A range of programs can be found on the market that all facilitate the secure deletion of data. Special processes are used that are sanctioned and suggested by such authorities as the US Department

of Defense and the German Office for IT security (BSI). [1], [6]

One of the most famous algorithms is the extended NISPOM (US DoD 5220.22-M ECE), which defines a sevenfold overwriting process. This uses a combination of random values, predefined values and their complements.

The method developed by Peter Gutmann is now regarded as the most secure method of secure deletion. This involves overwriting the data up to 35 times. Restoration of the data by any software method is thereby made impossible.[2]

For this purpose, O&O Software GmbH offers two products: O&O BlueCon, which, for example, facilitates the secure deletion of entire partitions. Even system partitions can be wiped with this program – this is not normally possible when the system is up and running. The second product is O&O SafeErase, which is integrated into the context menu of the Windows explorer. The user can thereby delete files, directories and partitions securely. Both programs offer five different ways to erase data, including the two methods already described.

Conclusion

The protection of sensitive data in Germany, both for home and business users, shows a strange contradiction. Almost everyone uses a virus scanner to protect themselves from external attacks and their consequences. At the same time, the handling of important data after the hard disks are thrown away is careless and negligent. The results presented here represent a cross-section snapshot of the daily practice of hard disk sales. If one takes this into account, the data we found on the 100 hard disks are a clear sign that the awareness of both home and business PC users has to be dramatically raised.

With regard to home users, this carelessness can partly be explained by the impression Windows gives – namely that data are deleted when the hard disk is formatted. However, the professional system administrator should be aware of the situation. After all, who hasn't restored files a user has accidentally deleted? It is their duty to ensure that data are not passed on along with old PC systems and hard disks and fall into the wrong hands. Anything else is dangerous and negligent.

Sensitive patient data

The example of the health insurance company shows clearly that careless behavior occurs when disposing of data carrying media. How else would it have been possible for us to buy two hard disks from two different resellers that originated from two different offices? How great is the probability of obtaining two hard disks from the same company from only 100 hard disks in total? There is obviously a serious demand that the company check its procedures; patient medical data is amongst the most sensitive information that exists. Who wants their medical data available for sale on eBay?

The warranty danger

The danger of data misuse appears not only when hard disks are resold or given away. When a computer is repaired, one usually hands over the whole computer – including the hard disk. In this way, private data can get into the wrong hands. One should therefore be careful to whom the computer is entrusted for the purposes of repair. It is best to get a written assurance that data will neither be copied nor read in as far as this is unnecessary for the repair.

To make completely sure that no danger exists, one can remove the hard disk before bringing the computer in for repair. However, this is only possible if it does not invalidate the guarantee or make the repairs impossible.

Omnipresent data carriers

More and more equipment includes data carriers- from PCs and notebooks to cell phones and PDAs (Personal Digital Assistants). All these gadgets can

be connected and synchronized so that personal data, appointments and documents can easily be found in unexpected places.

In particular, the power of the cell phone is growing. The boundaries between cell phones and PDAs are being eroded, so that greater volumes of information are constantly available. Cell phones are normally protected against abuse with a PIN code, so that at least the data are protected if the phone is lost or stolen. It is indeed quite difficult to restore data from cell phones. However, Toshiba recently demonstrated mini hard disks for cell phones that are to be introduced by the end of the year. In this way, cell phones can store several gigabytes of data. The problem of secure deletion becomes particularly important.

The danger of data misuse

The danger of data misuse is omnipresent. Account access information can be used to gain access to online bank accounts and perform money transfers .If the correct PIN and other access codes are used, the account holder is liable for any damages, as it is difficult or impossible for him to prove that he did not authorize the transaction.

The personal documents we found (ID card, driving license, birth certificate) could enable criminals to take over the user's identity. There were also incriminating files (illegal copies) which could be used to blackmail users. Would a family man really want his pornography collection, including pictures with animals, to be sent to his wife?

If the data we found were misused, legal consequences (civil and criminal) could follow from the violation of copyright and the damages that stem from this.

Solutions

Every PC user has to know that deleting and formatting within Windows does not destroy data. In this study, we have introduced various methods of restoring data. The minimum solution is data encryption, which makes the data unusable. For greater security, software solutions for secure deletion are necessary. There are many Internet sales outlets offering such software at a low price.

If the user does not wish to invest this effort, he should perhaps destroy his next hard disk rather than selling it on. Here care needs to be taken. Throwing the hard disk in the waste bin does not destroy it for good.

Acknowledgements

I would like to thank my colleagues Frank Labedzki, André Weiß, Matthias Günther and Fatihelyasin Erdas for their support during the carrying-out of this study. Not only did they undertake the several-week-long process of purchasing the hard disks, they also carried out the time-consuming process of data reconstruction and evaluation of the results.

I would also like to thank my colleague Frank Alperstädt his constructive criticism during this study.

Special thanks to Liz Disley, who translated the original German text into English.

About the author

The IT specialist Olaf Kehrer is he managing director of the Berlin firm O&O Software GmbH, which has as one of its specialist areas secure deletion of data. Amongst other tasks, he is responsible for the development of new technologies and products for data security. This includes the products O&O BlueCon, O&O SafeErase and O&O UnErase – apart from the data deletion functions described in the study; these products also restore and repair Windows systems.

About O&O Software GmbH

O&O Software GmbH has been developing Tools for Windows since 1997. These tools are sold in more than 80 countries in a variety of different languages. Customers include home users, SMEs, public institutions and global business concerns. The product portfolio includes applications for performance optimization, data restoration and the secure deletion of data. O&O products have been judged as technologically leading in numerous tests. You can find more information on our website or from us directly.

O&O Software GmbH
Am Borsigturm 48, 13507 Berlin, Germany
Internet: http://www.oo-software.com/
Email: info@oo-software.com
Tel: +49 30 4303 43-00
Fax: +49 30 4303 43-99

Appendix A: The saving and deleting of data on hard disks

The following text has already been partly published in another article by the author on tecchannel.de.[8]

How data are saved

Before one can delete data, one has to know where these data are stored. Often it isn't just the file itself that has to be deleted.

When a file is copied, moved or compressed, the original version of the file is often retained. Particular care is advised by, for example, so-called version systems. These deliberately keep old versions or a file in case the user wants to compare or reverse changes. This is particularly the case with the Windows 2003 Server operating system with its 'shadow copy' system. This is supposed to prevent the user from changing or deleting a file by accident. Changes to the file are therefore kept in special storage areas on the hard disk, so that old versions can be restored. In this way, it is also necessary to delete the (shadow) file in order to completely delete the data.

Windows also creates copies of data: temporary files include intermediate versions of the actual file. Swap files include storage areas that do not fit in the main memory but are to be moved in later. Temporary files are generally deleted when the relevant program is closed down, but in this case too deletion is just the release of storage space on the disk – these data can also be restored.

Hidden data carriers

Data also hide themselves in some other places which the normal user cannot normally access. One problem is the so-called cluster tips. Each hard disk is divided up into clusters when it is formatted. These are the smallest units on a hard disk that the operating system can use. Given the large size (tens of GB) of today's hard disks, clusters of 64kb are becoming quite usual. For the operating system, this means that even if a file is only 12KB, it will still occupy a space of 64 KB. The rest of the cluster is then unused.

This is normally not a problem, but storage areas are often released for use and then overwritten with new data. Let's imagine that a file has a size of 62 KB and occupies one cluster. This file is then deleted, the data remain but the directory entry disappears. Now a new file is written to this cluster. Let's imagine this is 10 KB. In this case, only the first 10 KB of the file are overwritten and the other 52 KB remains. This example can of course be transferred to any file, as larger files are separated into clusters so that the last cluster is usually not fully occupied. These data fragments are known as cluster tips. The problem with this is that one cannot access these fragments, as the cluster in question is marked as belonging to another file. Only with the help of specialist software can these fragments be erased.

Data between the ‚lines'

The saving of data to a hard disk involves the magnetization of very small iron particles which have a value or either 0 or 1. These particles are transferred to the surface of the disk and are divided into rows so that the hard disk head can read and write the data. However, data are not just written to the main row on the hard disk, but also just beside these rows. These outside rows therefore also contain data. This is normally now a problem, as he hard disk filters this „whispering" out. However, potential attackers can use this data traces in order to restore data. Earlier, simple processes such as the readjustment of the hard disk head were used fort his purpose. Today these data traces are more difficult to reach thanks to greater storage intensity. For this purposes, a great deal of effort and technical know-how is required. It is safe to assume that

only very well-equipped data rescue firms and secret services are able to use these data traces.

Deleting data

Deleting is more complicated than it sounds. Moving a file to the recycle bin and emptying said bi does not really remove it from the hard disk. The directory entry is removed, but the data themselves remain on the hard disk and can therefore be restored. Even formatting partitions and low-level formatting at the BIOS level does not represent as data can still be restored, even if the effort required is then greater.

Overwriting once or twice can be counterbalanced with the use of an error filter, and old data can thereby be brought to the surface. This is thanks to the physical effects which result from the analog signals of 0 and 1. These never really represent 0 or 1, but are distorted to 0.05 or 1.05. The hard ware counterbalances this defect with tolerance levels meaning that 1 can be saved as 0.95 or 1.05. By using these variations, a microanalysis of the relevant data signal can provide information about previous data values. If a 0 is replaces another 0, this gives a different signal strength than if a 0 replaces a 1. This process is not exactly easy or cheap, but it shows that simply overwriting the data does not delete them. For this reason, the most common erasure methods are always overwritten with a data value and its complement in order to make this 'difference method' impossible.

Appendix B: Bibliography

[1] DEPARTMENT OF DEFENSE, DEPARTMENT OF ENERGY, NUCLEAR REGULATORY COMMISSION, CENTRAL INTELLIGENCE AGENCY, *"National Industrial Security Program Operating Manual"*, 1995, 1997, 2001; http://www.dss.mil/isec/nispom.htm

[2] PETER GUTMANN, *"Secure Deletion of Data from Magnetic and Solid-State Memory"*, Usenix Assoc., 1996; http://www.cs.auckland.ac.nz/~pgut001/pubs/secure_del.html

[3] JOHN MARKOFF, *"Patient Files Turn Up in Used Computer"*, New York Times, 04.04.1997

[4] MICROSOFT, *"Encrypting File System for Windows 2000"*, Microsoft Inc., July 1999; http://www.microsoft.com/windows2000/techinfo/howitworks/security/encrypt.asp

[5] EGIL JULIUSSEN, PH.D., „COMPUTERS-IN-USE FORECAST", eTForecasts, June 2000, http://www.etforecasts.com/products/ES_cinuse.htm

[6] BUNDESAMT FÜR SICHERHEIT IN DER INFORMATIONSTECHNIK, *"IT-Grundschutzhandbuch"*, BSI, 2003; http://www.bsi.de/gshb/deutsch/menue.htm

[7] SIMSON L. GARFINKEL, ABHI SHELAT, *"Remembrance of Data Passed: A Study of Disk Sanitization Practices"*, Massachusetts Institute of Technology, 2003; http://computer.org/security/

[8] MIKE HARTMANN, OLAF KEHRER, *"Daten sicher löschen"*, tecchannel.de, April 2003; http://www.tecchannel.de/software/1161/index.html