

**Title: Computer Forensics - Computer Use Policy Reviews in Classified Agencies**  
**Author: Michael R. Anderson, SCERC, Special Agent, IRS-CID (Retired)**

---

## **Outline**

**1-1 Introduction**

**1-2 Personal Computer Security Weaknesses - Historical Perspective**

**1-3 Security Risks - Windows XP and Notebook Computers**

**1-4 Risks Associated with Ambient Data Storage Areas**

**1-5 Risks Associated with Computer-Related Storage Devices**

**1-6 Concerns Specific to Classified Government Agencies**

**1-7 Forensic Search Practices in Classified Security Reviews**

**1-8 Risks Associated with Non-Text (Binary) Files**

**1-9 Conclusions**

## **Key Words**

The following key words/terms should be referenced for Internet related searches:

classified security reviews  
classified data leakage  
government computer use  
computer usage policy reviews  
computer forensics in government  
computer forensic searches  
government incident responses  
risks in government agencies  
computer security risks  
government security risks  
embedded text risks  
search term creation

## **Information About the Author**

Detailed information about the author can be found on the Internet at <http://www.forensics-intl.com/mra.html>.

## **1-1 Introduction**

Forensics, by definition, is the application of law to science. In the case of computer forensics, computer science is used to identify evidence in criminal cases and civil law suits. Computer forensics is a relatively new forensic science but its procedures and methodologies have been used for years in military and law enforcement agencies for intelligence gathering and to identify criminal investigation leads and evidence. Computer forensics moved from the secret world of



the military and law enforcement when New Technologies, Inc. (NTI) was created in 1996. Since that time, numerous commercial computer forensics training courses have come into existence and several colleges and universities have incorporated computer forensics topics in their curriculums. Several computer forensic software tools have also come into existence and computer forensics has become a popular and lucrative career field.

Today, computer forensics is the mainstay of corporate investigations and internal audits. Since 1996, military and law enforcement agencies have expanded their use of computer forensics because of the increased popularity of personal computers and the Internet. It is common knowledge that the U. S. Military put a high priority on technology-based intelligence gathering in the Iraq war and in the identification of weapons of mass destruction (WMDs). Modern weapons development relies upon computers and such activities leave a computer evidence trail of activities behind. It is no secret that the U. S. Military relied upon computer forensics tools and processes to identify such activities in Afghanistan and Iraq. Computer forensic tools and processes are used to identify computers that contain classified weapons data and to identify the Arabic (and English) names of individuals stored on computers.

Most people don't realize that computer forensic tools and processes are also used by some of the same U. S. government agencies to identify their own internal computer security risks and weaknesses. Personal computers have significant security weaknesses and a security solution for Microsoft-based personal computers is not anticipated in the near term. Until a security solution is developed, computer security risk assessments using computer forensic tools and methods will continue to be mandated by most classified U. S. Government agencies.

## **1-2 Personal Computer Security Weaknesses - Historical Perspective**

Personal computers were never designed to be secure. This lack of security is the direct result of the development of personal computers over the last thirty years. Personal computers came into existence in the 1970s and the first personal computers were popular with hobbyists who built their own personal computers from a kit. These computers had limited power and were difficult to use. Pre-assembled personal computers could be purchased but the personal computer market was primarily limited during the mid-1970s to technology savvy hobbyists. Thus, cottage industries were the spawning grounds for personal computer research and designs, e.g., the Apple computer grew from a small business operated from a garage.

No software standards existed in the 1970s and custom operating systems were, of necessity, written for each personal computer system that came into existence. The operating systems and application programs that evolved were not interchangeable from one personal computer brand to another. Because of file incompatibility between brands, growth of the personal computer market was stifled and brand loyalties were strong. In the mid-to-late 1970s, personal computers with brand names like Altair, Apple, Atari, Commodore, Heath-Zenith, TRS-80 and Osborne dominated the small and disjointed personal computer market in the United States. Personal computers were considered toys or electronic gadgets by most people. Their benefits were not defined and most people had no use for computers anyway. One possible exception was a personal computer developed by Wang Laboratories, Inc. (Wang), which marketed computer products dedicated to word processing and document management tasks.

The Wang personal computer-based word processing systems eliminated the need for carbon paper and the technology also provided spell-checking features that had not existed previously. Wang word processing systems quickly unseated the popular IBM Selectric typewriter in the private and public sectors. The U. S. Government became one of Wang's biggest clients and I fondly remember those big 12-inch Wang floppy disks from my government days. I was a Special Agent with the Internal Revenue Service, Criminal Investigation Division back then and we used the Wang technology to create and edit memorandums, search warrant affidavits and prosecution reports. Wang word processing systems saved much time in creating and editing documents but they also had a downside - Wang systems provided no security for the computer files they created.

The U. S. Government adopted the use of Wang systems without security because it really had no other option. The Wang systems were state-of-the-art technology at that time and the U. S. Government had no place else to turn for similar technology. Its only option was to use the new Wang word processing technology, without security features, or continue using typewriters, "whiteout" and carbon paper. Reluctantly, the U. S. Government adopted Wang's technology because of the productivity benefits. This decision by the U. S. Government had a significant impact on the future designs of all personal computers. Had the U. S. Government required

**Title: Computer Forensics - Computer Use Policy Reviews in Classified Agencies**  
**Author: Michael R. Anderson, SCERC, Special Agent, IRS-CID (Retired)**

---

Wang to incorporate security into the design of its products, a benchmark would have been set for the computer security designs of the future.

In the late 1970s, Tandy Corporation developed a word processing application for use with the Tandy TRS-80 brand of personal computer. This was an attempt by Tandy to move its personal computer into the business world. Other computer manufacturers followed suit and developed similar word processing programs for use with their brands of personal computers but compatibility from brand-to-brand was still a problem. Word processing files created with one brand of computer could not be read, edited or printed on another brand of computer and none of these personal computers were compatible with the popular Wang word processing computer systems. Therefore, the personal computer market remained segmented and there was no reason for computer users to switch their loyalties.

The development of the computer spreadsheet application also occurred in the 1970s. The first significant software spreadsheet application was VisiCalc and its functionality and benefits supplemented word processing software applications developed for personal computers. The spreadsheet applications provided computer users with new and powerful business calculation capabilities that had not existed previously. The spreadsheet applications were not a good complement to Wang's word processing systems because Wang's end-users were primarily secretaries who created and edited word processing documents. Spreadsheet applications were more suited to business professionals who created and used their own custom spreadsheets. As technology savvy business professionals began using personal computers to perform mathematical functions, it was a natural progression for the same professionals to begin using personal computer-based word processing software. Wang was never able to make the transition from a dedicated word processing system to a multi-purpose personal computer system and eventually the company became an artifact of the computer technology revolution.

Based upon its success, Tandy Corporation promoted its TRS-80 Model III personal computer in 1979 as a "business computer" and that promotion effectively moved personal computers from "toys" to business computers in the marketplace. Tandy's promotion of personal computers as tools for business quickly captured the attention of the computer mainframe giant, International Business Machines (IBM). Mainframe computers were expensive and well beyond the reach of many businesses. Tandy effectively changed the business mindset about computing with its TRS-80, Model III personal computer.

IBM followed Tandy's lead and conducted market research to determine the potentials for its own personal computer. Based upon its market research, IBM released the IBM Personal Computer (IBM PC) in October 1981 for government agencies and businesses. The IBM PC was one of the biggest technology successes in modern times. But IBM significantly underestimated the market demand for the IBM PC and it, like everyone else, did not anticipate the potentials of the Internet. If IBM had foreseen the huge demand for personal computers, it is

**Title: Computer Forensics - Computer Use Policy Reviews in Classified Agencies**  
**Author: Michael R. Anderson, SCERC, Special Agent, IRS-CID (Retired)**

---

likely that security would have been a significant design feature. However, IBM knew that the U. S. Government was willing to spend money for personal computers which did not have security based upon the successes of Wang Laboratory. They also knew that the U. S. Government was the biggest user of computer technology and it was likely that private sector businesses would follow suit. That is exactly what happened and there was no business reason for IBM to secure the IBM PC initially.

Since 1981, the personal computer has become a powerful worldwide analysis and communications tool. Sensitive business and government documents are created and printed using these small computers. E-mail is routed around the world via the Internet. Database programs are used to store and access business information. Spreadsheet applications are used in financial calculations and PowerPoint is used to make most business presentations. Even photography has moved from film to digital flash memory. These wonderful technology tools are also used in classified government research and intelligence gathering activities. They are even used to track the financial trail of terrorists in the War-On-Terror. Regrettably, personal computer systems still aren't secure but, as previously discussed, they were never intended to be secure. As a result, an added layer of security risk exists today in classified government agencies and computer forensic tools and methods are typically used to reduce those risks.

### **1-3 Security Risks - Windows XP and Notebook Computers**

The creators of the original IBM PC never imagined that their primitive computer design would become the mainstay of worldwide commerce and a critical component used in the operation of U. S. Government agencies. Because of the need to provide upward compatibility for files and software, the basic foundation for the original IBM PC still exists in most of today's personal computers. Although the Microsoft NTFS-based operating systems (Windows NT, Windows 2000 and Windows XP) are more robust and provide better network security and auditing capabilities, no substantial security improvements have been made at the data storage level. Unfortunately, these advanced NTFS-based operating systems can create a false sense of security for computer users and management in classified U. S. Government agencies.

Microsoft-based personal computers can easily be compromised and password and logon controls can quickly be circumvented using computer forensic tools and methods. For example, NTI's TextSearch NT (forensic search tool) can completely evaluate and document all data storage areas on a Windows XP-based computer system. This U. S. Department of Defense (DoD) certified forensic tool was designed for internal government security reviews but its uses could be twisted, in the wrong hands, to compromise computer security. TextSearch NT operates from a DOS formatted floppy diskette and no logons or passwords are required to circumvent the minimal security afforded by Windows XP (or Windows NT and Windows 2000). The reader should note that this is the reason that some of the more powerful computer forensics tools are not made available for purchase by the general public.

Security problems are compounded when portable notebook computers are used with classified government data. Portable notebook computers are frequently used in classified executive briefings and sometimes in military and intelligence field operations and they can be taken anywhere. Most notebook computers also feature hibernation modes of operation that create added security risks. Some computer manufacturers call them suspense or sleep modes and the features rely upon a special file or a special partition that essentially captures and stores the work session when the computer is "asleep".

Notebook computers usually go into a sleep or hibernation mode when keyboard activity has not been detected for a predetermined period of time. This is a convenient feature that helps conserve battery power, but the tradeoff is an added security risk. When the hibernation or suspense feature is triggered (either manually or automatically), the work session is frozen in time and part of the data from the work session is stored in a hibernation file or partition. You can think of it as an electronic bookmark used by the computer. The data stored in the hibernation file can consist of any data tied to the work session and, when the computer is awakened, its operation is restored using the data contained within the hibernation file. The hibernation file is not securely deleted after the computer is awakened and the file contents remain behind for discovery using forensic tools and techniques. If classified data were involved

**Title: Computer Forensics - Computer Use Policy Reviews in Classified Agencies**  
**Author: Michael R. Anderson, SCERC, Special Agent, IRS-CID (Retired)**

---

in the work session before the computer went into hibernation, then it is likely that the hibernation file (or partition) contains classified data. Hibernation files are potentially huge. The hibernation file of the notebook computer I used to write this chapter has a capacity of over 203 million bytes of data. That is the equivalent of approximately 507,000 printed pages!

## **1-4 Risks Associated with Ambient Data Storage Areas**



Most personal computer users are unaware of the background processes involved in the operation of the computer. The processes are transparent and they potentially involve the leakage of sensitive computer data into 'special' data storage areas. These obscure storage areas are referred to as ambient data storage areas and include file slack, Windows swap files, Windows page files, temporary working files, work session suspense files and unallocated (erased file) storage space. Even the mere viewing of sensitive files on floppy diskettes or over the Internet can result in data seepage into ambient data storage areas and the computer user doesn't have to save any work to disk for the process to occur. Because of a general lack of knowledge of the security weaknesses of personal computers, government employees can unintentionally transfer classified data to unclassified computer systems. Ambient data storage areas constitute the biggest risk for classified data leakage and they are described as follows:

**File slack** is defined as the data storage space between the end of a file and the end of the last cluster assigned to the file. Files are stored in uniform blocks of data called clusters and a more specific definition of clusters can be found on the Internet at <http://www.forensics-intl.def19.html>. Rarely does the size of a file exactly match the data storage capacity of the number of clusters assigned to the file. File slack is the residual storage area which exists in the last cluster assigned to the file and following the contents of the file. File slack consists of two separate components called RAM slack and drive slack, discussed in following paragraphs.

Word processing documents, spreadsheets, databases and E-mail messages are all stored in files on personal computer storage devices. The same is true of many temporary files that are created transparently by software applications and the operating system. File slack is created when a file is saved (closed) and it is a significant security risk on all Microsoft-based personal computers. You should also be aware that the data potentially stored in file slack is typically beyond the knowledge and control of most government computer users.

**RAM slack** is created from the buffers on Windows and DOS-based systems. Buffers can be thought of as plumbing used by the operating system and the number of buffers on a Windows/DOS-based system is set in the CONFIG.SYS file, e.g., buffers=30. The buffers reside in the Random Access Memory (RAM) of the computer system and the contents of the buffers can potentially contain any data or data fragments created, viewed, or printed during a computer work session. RAM slack is written to the first sector of the last cluster of the file. In Microsoft-based computers, sectors are small storage blocks that hold 512 bytes of data. Clusters are made up of varying numbers of sectors depending upon the size of the storage device and the operating system involved. RAM slack will always be in the first sector of the



last cluster of the file but RAM slack cannot contain more than 512 bytes of data. Ram slack is only a concern on Windows, Windows 95, Windows 98, and DOS-based systems because Windows NT, Windows 2000 and Windows XP-based computer systems automatically scrub all relevant data contained in RAM slack.

**Drive slack** is a security risk on all Microsoft-based personal computer systems because it is not automatically scrubbed by the operating system. Unlike RAM slack, large quantities of data can reside in drive slack because its storage capacity is not limited to one sector, i.e., 512 bytes. Drive slack can potentially store up to 63 sectors of data (32,256 bytes of data or the equivalent of approximately eight printed pages), depending the operating system involved. Information stored in drive slack can contain remnants of previously deleted files and other information that resided on the storage media before the file was created.

The following example will to help clarify file slack and its components. Assume that you create a file by writing the word "Hello" and no other data is contained in the actual file. The file is only five bytes in length. Assuming that the clusters assigned by the operating system to the file are two sectors in size, the data stored to disk and the file slack would be represented as follows in Figure 1 below:



Figure 1

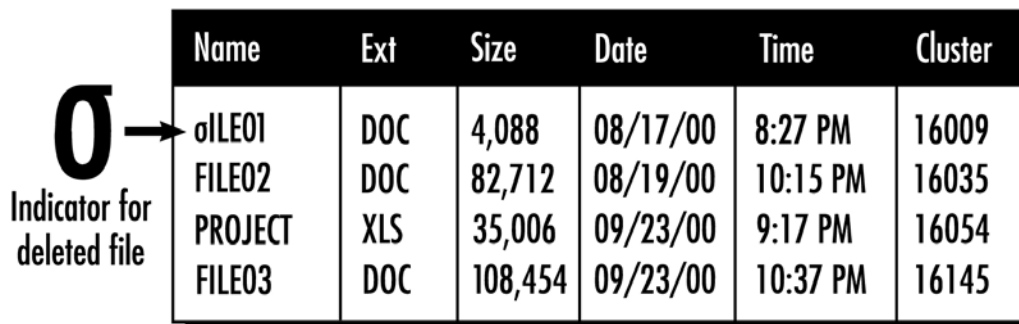
The data is identified by the word "Hello". RAM slack is identified by the "\ " symbol and drive slack is identified by horizontal line pattern. The "(EOC)" marker has been listed to identify the end of the last cluster of the file but it has been provided for illustration purposes and such a marker is not actually stored on disk at the end of the assigned cluster. Rather, the end of the file is recorded in the directory area. In this example, only one cluster is needed to store the small five byte file containing the word HELLO.

**Windows swap and Windows page files** act as an extension of memory for the operating system and are used when more memory is needed by the operating system. This happens when multiple software programs are running at the same time or when extremely large documents or graphics files are viewed or edited. These files act as a “scratch pad” for the operating system when more memory is needed. Windows swap files and Windows page files are huge. Depending upon the operating system configuration, the size of the swap or page file will be between approximately 50 and 700 million bytes. In the case of Windows, Windows 95 and Windows 98, the file is called the Windows swap file. For Windows NT, Windows 2000 and Windows XP , the file is called the Windows page file.

Most computer users are unaware of Windows swap and Windows page files and they don't realize that fragments of their work products may transparently be written to these special files. Essentially any work performed in a Windows work session can end up in the swap or page file, including fragments of created files, edited files and even fragments of files that were merely printed and not created or edited on the computer system. Swap or page files also capture fragments of Internet web browsing activities, Internet E-mail addresses and messages, and even the directories listings displayed in a Windows Explorer session. Passwords and logons may also be written to the Windows swap and Windows page files. These special files are a wonderful source of investigative leads in computer evidence-based cases, but they are a significant security risk in classified government agencies. More information about these files is available on the Internet at <http://www.forensics-intl.com/def7.html>.

**Unallocated file space** is another ambient data source that should be of concern to government computer users. When files are "deleted" on personal computers, the data associated with subject file is not erased. Rather, the space assigned to the file becomes unallocated by the operating system and the storage space is made available for new files. However, the data from the former file can actually linger on storage media for months or even years. The same is true of the name of the former file. Only the first byte of the file name is overwritten and it is replaced with a lower case Greek sigma character.

An example of a deleted file directory file listing follows in Figure 2 below:



Name	Ext	Size	Date	Time	Cluster
σILE01	DOC	4,088	08/17/00	8:27 PM	16009
FILE02	DOC	82,712	08/19/00	10:15 PM	16035
PROJECT	XLS	35,006	09/23/00	9:17 PM	16054
FILE03	DOC	108,454	09/23/00	10:37 PM	16145

Figure 2

In the above example, the deleted file was named FILE01 before it was erased and the only change to the file name is the first character which was replaced with the with a lower case Greek sigma character. Note also that the particulars about file dates, sizes and attributes remain behind.

Unallocated file space should not be confused with free storage space. Free storage space is the space that resides beyond allocated hard disk drive partitions. Unallocated file space differs because it is contained within a logical hard disk drive partition and it can contain both erased file data and the file slack associated with erased files. Unallocated file space is the largest source of ambient data and it can potentially involve several billion bytes of data on a large personal computer hard disk drive. Therefore, the security risks associated with unallocated storage space can be significant in a classified government agency.

**Temporary files** are created and used by most software applications. The operating system also creates and uses various temporary files to perform various tasks, e.g., the printing of files. These specialized files are typically created as the result of background processes and the user is usually unaware of their existence or purpose. When the need for a temporary file has passed, it is typically deleted by the background operation that created it. However, the data associated with the erased temporary file remains behind in unallocated file space. In this regard, erased temporary files are no different than other erased files.

Many Windows-based software applications create temporary files to facilitate sorting functions, to create indexes, and for directory scrolling in Windows Explorer, etc. Temporary files potentially store fragments of the data processed during the computer work session. All allocated and erased temporary files should therefore be considered a security risk in classified government environments because the likelihood exists that they may contain classified information.

**Partition gaps and free space** can be a security risk on previously used personal computer systems. Partition gaps and partition free space risks are some of the reasons that the U. S. Government requires that security risk reviews include the search of all physical sectors on the subject hard disk drive. Individual hard disk drives are referred to as physical hard drives (physical devices) and the data storage area of a physical hard drive can be broken into smaller storage components which are called logical hard drives. This is typically done with commercially available hard disk drive partitioning software when operating systems are upgraded on used computer systems or the computers are transferred from one person to another. During the upgrade process, not all of the storage capacity of the physical hard drive may be needed and therefore a smaller partition is used. In such cases, partition gaps can also result between allocated partitions on the same physical hard disk drive. As an unintended result, artifacts of the legacy data may remain behind in the partition free space or in partition gaps when multiple partitions are involved on the same physical hard disk drive.

An example of a partition gap is illustrated as follows in Figure 3:

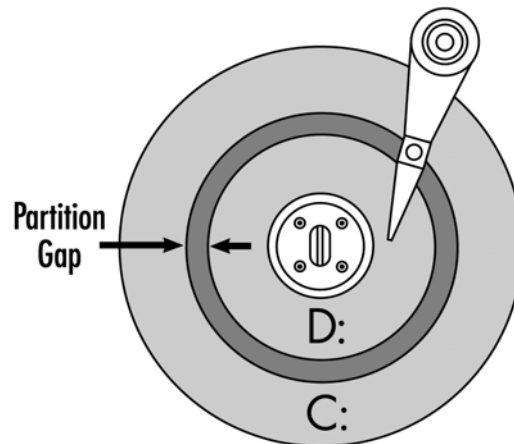


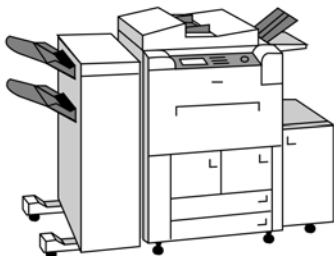
Figure 3

After partitioning, the resulting drive partitions are referred to as logical drives and, in the example above, they are referred to as drives C: and D:. Multiple logical disk drives can reside on one physical drive and multiple operating systems can also reside within different logical drives on the same physical hard disk drive.

Many risks associated with partition gaps and partition free space can be identified through the use by a DoD certified forensics text search utility which has the capability of searching all areas of the physical storage device.

**Hibernation mode files and partitions** create added security risks for government employees who deal with classified information, as mentioned previously. Most notebook computers automatically capture and save a computer work session using hibernation mode files or dedicated partitions. This conserves battery power and it provides a convenient way for computer users to resume their work sessions where they last ended. When the hibernation option is triggered, the notebook computer bookmarks the last work completed and puts the computer into a sleep mode. Computers that have this feature essentially capture the work session using a special file or, in some cases, a special partition. When the hibernating computer is awakened, the computer user can resume his/her work session at the point where he/she left off. This provides the computer user with benefits but significant risks are created because artifacts of the suspended work session can remain behind on the hard disk drive for an indefinite period of time. The risks are substantial because hibernation files are huge, i.e., over 200 million bytes in size, and they should be factored into government computer security risk assessments when notebook computers are used to process or analyze classified government data.

## **1-5 Risks Associated with Computer-Related Storage Devices**



**Photocopy machines** bring other risks to classified government agencies because copied classified data can be inadvertently stored on these machines. Photocopy machines typically rely upon personal computer technology in their operation and many of the devices contain internal computer hard disk drives. Some photocopy machines can also be used as printers in a networked environment. If the photocopy machine is used with sensitive government data, then it is likely that the data will migrate onto the photocopy machine's internal hard disk drive. For this reason, photocopy machines should be treated as personal computers by government computer security specialists. The security risks are the same as with personal computers and, in some cases, the risks may actually be greater because some of the computer files created by photocopy machines are binary in nature and their contents can't be evaluated using forensic text search utilities.

**CD's and DVD's** create additional risks in classified government agencies because of their portability and large storage capacities. These storage devices easily interact with personal computers and they can be duplicated without leaving a trace of the duplication process behind. For these reasons, physical security is usually required for these storage devices in classified government agencies. However, computer security risk assessments should include the review of non-classified CD's and DVD's in classified government agencies, because of the potentials of classified data leakage onto these storage devices.

**Personal Digital Assistants (PDAs), digital cameras and cellular telephones** are all capable of exchanging data with personal computers and they also create additional and significant security risks in classified government environments. Because of their portability and non-essential needs in classified agencies, these devices are typically not allowed in classified government environments. However, exceptions are made in some classified government agencies and computer security specialists need to be aware of the potentials for classified data leakage into the internal storage areas of these computer-related devices.

**Floppy diskettes, USB memory sticks, and Iomega Zip disks** all interact with personal computers and they are portable data storage devices. Floppy diskettes were allegedly used in the Robert Hanssen spy case in the transmission of U. S. Government secrets to the Russians. Even though their storage capacities are relatively small, floppy diskettes still pose a significant risk in classified government agencies because of their portability. As stated, their storage capacities are fairly limited when compared to newer technologies, e.g.,



**Title: Computer Forensics - Computer Use Policy Reviews in Classified Agencies**  
**Author: Michael R. Anderson, SCERC, Special Agent, IRS-CID (Retired)**

---

USB compatible flash memory devices and Iomega Zip disks. If these devices are present or allowed in classified agencies, they should be considered a risk because of the potential for leakage of classified government data.

## **1-6 Concerns Specific to Classified Government Agencies**

The potential for leakage and the unintentional transfer of sensitive data to unclassified computer storage devices is of great concern in classified government agencies. Unfortunately, classified data leakage is a common occurrence in classified government agencies. This is because most government employees are unaware of the inherent security weaknesses associated with personal computer usage; specifically the potentials for migration of sensitive government data into ambient data storage areas. Security weaknesses are inherent in the design of personal computers and most of the security risks are not obvious to computer users. This is because the risks are highly technical in nature and beyond the knowledge of most unsophisticated computer users who just want to get work done. For these reasons, most classified government agencies make it a priority to regularly scan non-classified computer systems to identify the leakage of classified data. When the classified data leakage is identified, the data must be securely eliminated with DoD certified data scrubbing software tools or other approved methods.

Classified computer security risk assessments are typically conducted with DoD tested and certified computer forensic search tools that are executed from a bootable floppy diskette or a bootable USB storage device. By using floppy diskettes and/or USB devices, running under DOS or Linux, a government computer security specialist can simultaneously review several personal computers. The practice also allows for the search of all sectors of the targeted hard disk drives involved. It is a U. S. government requirement that such searches include the search of every sector of the storage media and this requirement cannot be accomplished from a computer network or over the Internet. Therefore, government security reviews are usually conducted onsite and multiple computers are accessed for risk at the same time.

Most government computer security review policies and procedures are outdated because they were developed under the assumption that classified data leakage would be limited to ACSII (text-based) data and files. Unfortunately, most of the policies were created before Microsoft PowerPoint and digital photography became popular in the workplace. For this reason, many of the current government security review policies and procedures need to be updated to account for new risks tied to current technology. In addition to text-based files, the policies should take into account threats tied to compressed files, e.g., Zip files, and graphics files, and proprietary file formats which are binary in nature. Such files can potentially mask the fact that they contain classified data and they do not respond well to text-based computer forensic search tools. Some government security review procedures don't take into account disk fragmentation. Fragmented data can result in targeted search terms being split between clusters and that can have a negative affect on ASCII text-based security reviews. For these reasons, existing policies should be reviewed to make sure that they are current with technology. If they are outdated, security review policies and procedures should be updated or supplemented to take into account new security risks associated with new file formats, storage devices and other technology advances which have been adopted for use in the workplace.

**Title: Computer Forensics - Computer Use Policy Reviews in Classified Agencies**  
**Author: Michael R. Anderson, SCERC, Special Agent, IRS-CID (Retired)**

---

Computer forensic search tools cannot adequately search some types of files, e.g., graphics files, PowerPoint files, compressed files and PDF files, etc., and some suggestions have been provided by the author in hopes of enhancing the quality of classified government security reviews and risk assessments. Until security is designed into personal computers, security reviews will remain as a standard practice in classified government agencies. It is also likely that similar security review practices will be adopted by other U. S. government agencies in the future as U. S. Homeland Defense measures are implemented and expanded. As of this writing, the demand for government computer security specialists exceeds the supply of qualified individuals with computer forensics knowledge and experience. Private sector businesses and Fortune 500 corporations will likely encounter the same shortages of qualified computer security specialists as they deal with recently enacted laws regarding security of information, e.g., HIPAA, Gramm-Leach-Bliley and Sarbanes-Oxley. These laws require many businesses and organizations to establish safeguards and controls over the security and privacy of financial, health and public corporation insider information. These new laws are discussed in more detail elsewhere in this book.



## **1-7 Forensic Search Practices in Classified Security Reviews**

Most classified U. S. Government agencies require the periodic review of all personal computers located in close proximity to computer systems which store and process classified data. Computer forensic search tools are the first line of defense in classified agencies for the identification of classified data leakage. The U. S. Government requires that these reviews include the search of every sector of the storage media for classified text. Typically, DoD tested and certified forensic text search utilities are used for this purpose, e.g., NTF's TextSearch Plus and TextSearch NT and government developed tools like D-Scan. As mentioned previously, government security reviews are lacking if they rely totally upon text searches for the identification of classified data leakage. Forensic search tools are extremely helpful in security reviews but they provide no benefits when classified data is potentially stored in non-text file formats. The techniques discussed in the following pages are provided with the intent of helping improve the accuracy of risk assessments in government agencies.

### **The Creation of the Search Term List:**

The terms used in the computer forensic search are extremely important to the success of any computer security risk assessment which involves the use of a computer forensic text search tool. The creation of the search term list is one of the most important parts of the security assessment. Lists that include small words, short terms and abbreviations tend to generate false search results. Lists that include long strings of text may be missed in the search process due to disk fragmentation. The effectiveness of the security scanning process is only as good as the design of the search term list and the computer forensic search tool used in the search. In any case, the list of search terms should be crafted with much care and thought by a person who is trained in the use of computer forensic search tools. It is helpful if the person conducting the search has knowledge of personal computer technology and inherent personal computer security weaknesses.

Search term lists can be created using DOS Edit, Windows NotePad or even a word processing program. Depending upon the computer forensic search utility used, search terms are usually stored in ASCII format and each search term is terminated with a carriage feed/line feed sequence. DOS Edit and Windows NotePad generate this type of file output automatically and a word processor can be used to generate such a file using the "File Save As" feature. When using a word processor, the list of terms is saved in ASCII DOS Text format.

Ideally, short words and abbreviations should not be used in the search term file. This is because the letter combinations associated with some relevant classified words or terms may also be found in common forms of data found on most computer systems. The following is an example of a flawed search term list that would likely result in hundreds or even thousands of false hits in a security review which relies upon a forensic text search utility:

**An Example of a Poor Search Term List:**

troll  
lion  
lie  
secret  
soft  
copy  
poly  
roso  
program

This listing of search terms may appear to be acceptable but a close examination reveals several problems with the terms in the list, e.g., these terms are sub-strings of larger words or they are found in system files that are common to most personal computers. If this list were used in the search of a hard disk, hundreds of false leads would likely be identified by a text-based computer forensic search tool. To illustrate this important point, please consider the following:

1. **troll** is included in the words controller, controlled and trolley.
2. **lion** is included in the words battalion and dandelion
3. **lie** is included in the words believe, client, lien, earliest, families and allied.
4. **secret** is included in the word secretary and the titles Secretary of State, Secretary of Commerce, Secretary of the Treasury, Secretary of Defense, etc.
5. **soft** is included in the name Microsoft Corporation and personal computer searches will identify thousands of occurrences of this word on most systems. The same situation exists with the search term "**roso**" which is also included in the name Microsoft Corporation.
6. **copy** is included in the words copyright, copying and photocopy.
7. **poly** is a slang term for polygraph and is included in the word polygon.
8. **program** is a term that is stored in numerous locations on all personal computers, e.g., the error message contained in all Windows based programs "program cannot be run in DOS mode".

**Title: Computer Forensics - Computer Use Policy Reviews in Classified Agencies**  
**Author: Michael R. Anderson, SCERC, Special Agent, IRS-CID (Retired)**

---

This hypothetical list of search terms would need to be fine-tuned to enhance the search results and to eliminate as many false hits as possible. An example of the perfected search term list might look something like the following:

```
troll<space>  
<space>lion  
<space>lie<space>  
secret<space>  
<space>soft  
<space>copy<space>  
poly<space>  
<space>roso or roso project  
weapons program
```

The strategic insertion of spaces in the search term list is important when one or more of the targeted search terms are a part of larger words. This technique is called “space bracketing” in computer forensics and it is very effective. However, the space bracketing technique can create problems when punctuation immediately follows the term on the targeted storage device. Other problems occur when plurals of a word in the search term list exist on the targeted storage device.

To illustrate some of the problems associated with space bracketing, consider a search for the word “secret”. Let’s assume for a moment that we are conducting a security review at a U. S. Embassy and we want to avoid false hits associated with the title, “Secretary of State”. We would modify the search term list by adding a space behind the word “secret”. However, the addition of the trailing space would cause the forensic search software to skip the word “secrets” and it would skip the word “secret”, if it appeared as the last word in a sentence. This is because the trailing punctuation, i.e., a period, would defeat the search results for some forensic search tools. DoD certified forensic search tools, e.g., TextSearch Plus and TextSearch NT, will not be fooled by trailing punctuation. To avoid missing the word “secrets” in the search, we could add that word to the search term list. Thus, the search term list would include both the word “secret” (followed by a space) and the word “secrets” to enhance the results of the search.

Repeated test scans are usually required to perfect the search term list to avoid false positives and to ensure that relevant search terms not missed. It is not uncommon in computer forensics to “fine tune” the search term list during the search as false positive findings are identified. Once an accurate search term list has been created, it can be used on multiple computers in the security assessment. Every change to the search term list should be given much thought because even minor changes can affect the accuracy of the security search results.

## Logical Verses Physical Text Searches

The U. S. Government requires that all sectors of storage device data be searched to identify the potential leakage of classified government data onto unclassified computers and related storage devices. This policy is based upon sound computer forensics logic because of the possibility that classified data could reside between or beyond partitions on the storage device. As mentioned previously, this can occur when used computers are upgraded with new operating systems or when hard disk drives are repartitioned during hardware maintenance. Physical keyword searches involving the search of each sector will also identify allocated and “deleted” file names that may also contain classified terms. However, remember that the names of erased files will omit the first character in the file name. It will be replaced with a lower case Greek sigma character. If the potential exists for file names to be a security risk, then this needs to be taken into account when the list of search terms is created.

Logical text searches of storage volumes will not identify risks associated with file names, data stored in the MFT on NTFS-based systems and data potentially stored in the partition gap or partition free space. However, logical searches can provide benefits that do not exist with physical searches when disk fragmentation is involved. To illustrate this situation, assume that our search term list includes the classified term, “Aardvark55”. The length of this search term is not excessive and it could be relevant in a classified security risk assessment. For the purposes of this example, assume that the hard disk drive involved is well used and the data storage areas are fragmented. On well-used disk drives the potential of disk fragmentation is high and the potential exists for targeted search terms to be split between consecutive clusters. This is illustrated as follows in Figure 4:

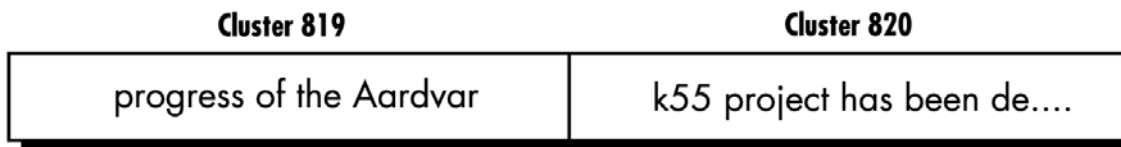


Figure 4

For the purposes of this example, assume that the file has not been deleted and that it is still an active file. If we were to perform a physical text search for the term Aardvark55, the data would be identified by the forensic search utility. The same would be true if we were to performed a logical search for the term using a forensic search utility.

Now let’s look at the same fact pattern but factor in disk fragmentation. For the purposes of the example and to illustrate the affect of disk fragmentation, assume that the data previously stored in cluster 820 was written by the operating system to cluster 822. This would be the case, on a

used hard drive, if clusters 820 and 821 were already occupied with data from another file. Due to disk fragmentation, the data would be stored on disk as illustrated here in Figure 5:

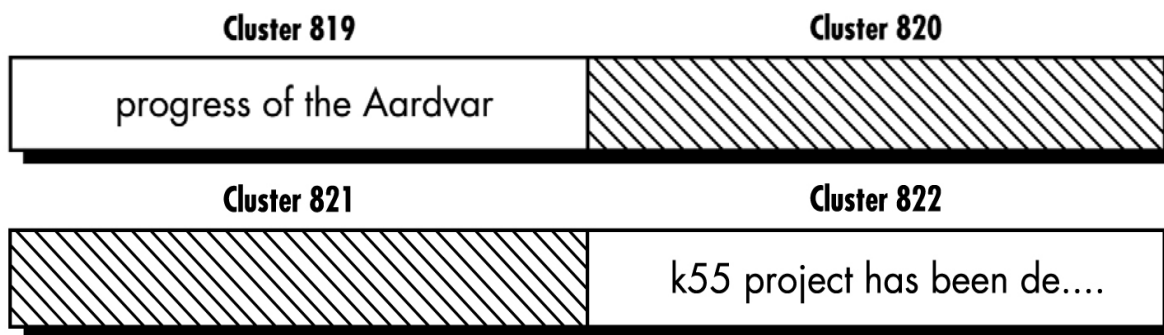


Figure 5

If we were to conduct a physical text search, in this case, the search term, "Aardvark55", would not be identified. A logical text search would identify the search term, though, because the FAT (or MFT on an NTFS system) would connect the clusters. However, neither the physical text search nor logical text search would identify the search term if the file was deleted. This is because the FAT (or MFT on an NTFS system) would no longer recognize clusters 819 and 822 as part of the same file.

The problem with fragmentation can be addressed in two different ways. First, it is recommended that both physical and logical searches be conducted using a DoD certified forensic text search utility. Such tools will accommodate both physical and logical searches and they can be run automatically in batch mode. Second, it always makes sense to split targeted search terms in the search term list to enhance the search potentials because of search weaknesses attributed to fragmentation. In the hypothetical Aardvark55 example, we could include Aardvark55, Aardv and k55 in the search term list. Although this involves three separate search terms and it will likely generate redundant hits, the practice helps ensure that targeted data will not be overlooked. However, you must be aware that even this technique will not ensure positive findings when deleted files and fragmentation are involved, as stated previously.

## **1-8 Risks Associated with Non-Text (Binary) Files**

Forensic text searches are not effective when certain types of files are involved. The files which cause concern include but are not limited to: compressed files, graphic files, embedded text files and compound files which include combinations graphics files and text. Forensic text searches are also ineffective if files have been encrypted using any number of commercial and government encryption tools. The topic of encryption has not been discussed here because encrypted data in a classified government agency is not a security risk.

### **Graphics File Formats**

Digital photography has become popular and graphics files are abundant on the Internet. In today's modern world, some cellular telephones have the ability to take pictures and store the image in the form of graphics files. Because of advances in technology, we have become a more picture based society. It is not uncommon for hundreds or thousands of graphics files to be stored on a well-used personal computer system. This causes problems in classified government agencies because forensic text-based search utilities cannot identify classified terms and words stored within graphics files. For this reason, it is currently necessary to identify and manually review the contents of all graphics files as part of a classified government security risk assessment given the state of current technology. **As of this writing, no forensic search tools can reliably search graphics files** and identify targeted key words or strings of text. The problem is compounded by the fact that **some text-based file formats allow the embedding of graphics files in the files**. This is the case with word processing files, PowerPoint and spreadsheet files. This means that computer forensic search utilities have limitations concerning the identification of all data leakage in classified government security risk assessments.

In the past, I have recommended that file headers of known graphics files be included in the search term lists. Some forensic search tools do this automatically for the common graphics file formats, e.g., BMP, GIF, JPG, TIF, etc. However, in high stakes computer security risk reviews, I recommend that all graphics files be captured from the target media and that they be reviewed using a graphics file viewer. Although this is a tedious process involving potentially large volumes of files, decisions can be made quickly by reviewing thumbnails of the extracted graphics files. The extraction process can be simplified through the use of a physical access capture of all graphics files contained on a target hard disk drive using a tool like NTI's Graphics File Extractor and the process can be performed using a portable USB storage device. Floppy diskettes cannot be used because they do not have enough storage capacity.

## **Compressed File Formats**

File compression programs are popular and they are used to combine and compress one or more files to save space. They were originally created to shrink file size back when files were downloaded from computer bulletin boards. Back then, modem transfer rates were very small and long distance phone rates were high. To save money and time, file compression was used extensively with programs like ARC, PKARC and LHARC. Today the most popular file compression program is PKZIP but other popular programs include RAR, ARJ, ACE, etc. In addition to compressing file contents, modern compression programs also provide security features through the use of encryption.

It is unfortunate, but forensic search utilities have no ability to reliably and quickly search compressed file formats. However, some forensic search tools do identify these files by either their file extensions or file header signatures. When the compressed files are identified in this fashion, a manual review is required. This can be a time consuming and tedious process and this is especially true if security features have been used with the compression. Because of the risks and problems created, many classified government agencies forbid the use of these programs.

## **Embedded Text and Obscure File Formats**

Not all files are stored as text and this complicates the job of the computer security specialist who uses forensic search utilities in computer security reviews. In general, text searches involve the search for upper and lower case characters, numbers and punctuation. All of these characters can be entered from the keyboard but that is not to say that a computer application will store text in the same order it was entered.

Computer applications typically allow work to be stored in the form of a file. Unfortunately, there is no universal standard for file formats and this is true of even commonly used computer applications like word processors. Because of this lack of standardization, file formats vary dramatically and some use special characters to identify unique characteristics about the stored data, e.g., underlining of text, bold type, dropped capitalization, italics, etc., in word processing files. Some word processing files contain a mixture of ASCII text blended with binary characters. This situation can cause problems in classified security reviews that are based purely upon the search of key words and strings of text in ASCII form. All computer security specialists should be aware of this fact and they should realize that **forensic text search tools have limitations**. I have provided some examples of different file formats as food for thought for computer security specialists. This information is intended to help in the crafting of search term lists and it is also intended to demonstrate the limitations of forensic text searches in classified government reviews.

**Title: Computer Forensics - Computer Use Policy Reviews in Classified Agencies**  
**Author: Michael R. Anderson, SCERC, Special Agent, IRS-CID (Retired)**

---

To illustrate the weaknesses of forensic text searches, in classified government reviews, let's assume that we have included the term "little lamb" in our search term list. The assumption would be that this is a term of interest or it could hypothetically be a classified project name. With this in mind, I created a one sentence WordPerfect document that consisted of - "Mary had a **little lamb** whose fleece was as white as snow." Note that the targeted search term, little lamb, was bolded in the sentence that I created. The file was then saved in the form of several different word processing file formats. You would think that this simple sentence would be stored identically in each file format but that is not the case.

I have provided various examples here to illustrate how file formats differ from one application to another. Please note that I filtered all binary and control characters and replaced them with a period. Any replacement character could have been used but I did this so that the results could be easily printed and reviewed. The results follow:



### AMI PR0 Version 3.0

Most word processing applications add information to the stored file which assists in the editing process. Therefore, word processing files are usually larger than the stored contents and this is the case with the AMI Pro file format. The sentence, "Mary had a little lamb whose fleece was white as snow." is 57 bytes in length but the file size is 1,358 bytes in length and the filtered output is listed as follows:

---

```
[ver]...4..[sty]....[files]..[charset]...82...ANSI (Windows, IBM CP 1252)..[prn]...PCL / HP  
LaserJet..[lang]...2..[fopts]...4...1...2880...0..[Inopts]...0...Body  
Text..[docopts]...5...0..[tag]...Body Text...2...[fnt]....Times New  
Roman...240...0...32768...[algn]....1...1...0...0...0...[spc]....33....273...1...0...0...1...100...[br  
k]...4...[line]....8...0...1...0...1...1...1...10...10...1...[spec]...0...0...0...1...1...0...2...0...0..  
.[nfmt]...272...1...2.....,....$.[lay]...Standard...513...[rght]....15840...12240...1...1440...1440  
...1...1440...1440...0...0...0...0...2...1...1440...10800...10...1...720...1...1440...1...216  
0...1...2880...1...3600...1...4320...1...5040...1...5760...1...6480...1...7200...[hrght]...[lyfrm  
]...1...11200...0...0...12240...1440...1...1...3...1...0...0...0...[frmlay]...1440...12240...1...1  
440...72...1...360...1440...0...1...0...1...1...0...1...1440...10800...0...[txt]..>...[frght]...[lyfr  
m]...1...13248...0...14400...12240...15840...1...1...3...1...0...0...0...[frmlay]...15840...122  
40...1...1440...360...1...14472...1440...0...1...0...1...1...0...1...1440...10800...0...[txt]..>..[  
elay]..[11]...0..[edoc]..Mary had a <+!>little lamb<-!> whose fleece was as white as snow...>..
```

---

Note that the targeted text, **little lamb**, is surrounded with special characters which are used by the application to identify that the text is to be displayed and printed as bold text. In this example, a forensic search utility would easily identify the targeted search term.



**MS Word Version 4.0**

Microsoft Word has become the most popular word processing format and MS Word, Version 4.0 is one of the older versions of the program. In this example, it created a fairly small, 772 byte file and the filtered output is displayed as follows:

---

1.....  
Mary had a little lamb whose fleece was as white as snow.....  
.....w.....  
.....=./....2...\$.....  
.....<.....=.....  
.....

---

In this example all of the text is displayed and a forensic search utility would easily identify the targeted search terms.

### MS Word 97/2000 for Windows

Microsoft has made many improvements to the MS Word program over the years. Current versions allow the tracking of changes and the incorporation of graphics, etc. These improvements come with a penalty in the size of the file. In the example, the converted file is 12,292 bytes in size and relevant parts of the filtered file are displayed as follows:

---

```
.....R.o.o.t.
.E.n.t.r.y.....@...".....P.e.r.f.e.c.t.O.f.f.i.c.e._M.A.I.N.....
....&.....P.e.r.f.e.c.t.O.f.f.i.c.e._O.B.J.E.C.T.S.....
.....@..."@...".....R.o.o.t.
.E.n.t.r.y.....;".....D.a.t.a.....
.....1.T.a.b.l.e.....O.b.j.e.c.t.P.o.o.l.....
.....;".....;".....W.o.r.d.D.o.c.u.m.e.n.t.....G.....bj
bj.....&.....P.....].....
.....g.....
.....$.....R.....
.....bt'.....
.....S.E.Q. .C.H.A.P.T.E.R. \.h. \.r. .1..M.a.r.y. .h.a.d. .a.
.l.i.t.t.l.e. .l.a.m.b. .w.h.o.s.e. .f.l.e.e.c.e. .w.a.s. .a.s. .w.h.i.t.e.
.a.s.s.n.o.w.....*.....B..X.....mH..CJ..5..mH..CJ...U..CJ..mH..U.....
.....1$...$.../
..=!..."#...$...%.....[.....(..@....(.....N.o.r.m.a.l.....CJ.
.mH.....<.A@...<.....D.e.f.a.u.l.t.P.a.r.a.g.r.a.p.h.F.o.n.t.....P.....
.....P.....P...@.....G.....
T.i.m.e.s.N.e.w.R.o.m.a.n..5.....S.y.m.b.o.l..3&.....A.r
.i.a.l..#.....1..1.....!#.....
.....
```

---

This is one of the most popular word processing programs and you should note that the application stores the sentence with binary characters inserted between each letter. Many of the older forensic search utilities would not recognize the targeted key words stored in this fashion. However, a DoD certified forensic search utility like NTI's TextSearch Pro or TextSearch NT would identify the targeted strings of text.

**OfficeWriter Version 6.2**

When the file is converted into the OfficeWriter file format it resulted in a 1,681 byte file and the filtered output is displayed as follows:

---

```
6.0
01/08/04:..wp.
                                85 110 10 10 10      n  nfo n6..      ..
.1.1.....
.....
.....1.1.....
.....
.....1.1.....
.....Mary had a .little lamb. whose fleece was as white as snow..
```

---

In this case the sentence is stored on disk as text and a forensic search utility would easily identify the targeted string of text. In this case, the bolding is tagged with binary characters before and after the term “little lamb”.

**Professional Write Version 2.2**

The sentence is converted into a 1,151 byte Professional Write file and the filtered output is displayed as follows:

---

B.....N.....  
.....#(-.2.7.<.A.F.....  
.....  
.....  
.TYPE  
4.2.00.....  
.....1.....K  
.K.....N.....#(-27<AF...Mary had a ..... whose  
fleece was as white as snow....N.....#(-27<AF.....

---

In this case the targeted strings of text are not visible because the word processing application converts them into binary characters to identify them as bolded text. For this reason, a forensic search utility would fail to identify the targeted strings of text.



### **WordPerfect Version 5.1**

The older version of WordPerfect was an industry standard for many years. When our sentence is saved in this file format the resulting file is 410 bytes and the filtered output is displayed as follows:

---

```
.WPC[.....2.....Z..B..0.....I.....S...Canon i9100.....X.N...\.....  
P.....X.P..X.N...\..... P.....X.P(.....9.....Z.....6.T.i.m.e.s. .N.e.w. .R.o.m.a.n.  
.R.e.g.u.l.a.r.....X.....3|x...Mary had a ...little lamb...  
whose fleece was as white as snow.
```

---

Because all of the data is stored as text, a forensic search utility would have no difficulty identifying the targeted strings of text in this case. Note that three binary characters are used, before and after, to designate the bolded text. Also, note that the connected printer is identified and recorded in the file.



**WordPerfect Versions 6/7/8/9/10**

WordPerfect has also made improvements to its word processors over the years at the expense of a larger file. When our sentence is converted into this file format the resulting file is 1,721 bytes in length and the filtered output is displayed as follows:

---

```
.WPC=.....Q.|R&....s.Z...v.]..ri.....,.....=H;Uf..)....U....?...4.\8z.-.}....'.6.(3....a.gG.z.I..
....B.+M.....la.r...hM+..?.i@..n.VD...?D7...7...O.C.B"@..N...].ug...D.yQY.....:.....{.@....8}..".@
W..L!...[.A&.VjK.l.Y.Z...Y[.p.#i.w..w.l]o.H.1H..)=....'ir.d...F...\z.lF.....<.m.j...Q5O..C.*1)..h.".
Am..@.fm...GX.e.Z`.....y$.S.....#B.qH=....<C.;.Rq.f...B..|Az..`I.=N.uM.9kK...7GX^.....BR.X.5
.lA...V.r....NO)Gi.PF...K.....}z....z....PB'.V.[...9.Hd.....#.....U...N...c....%.....0.....:
....^.....w.....4.....$...m.....&....C.a.n.o.n.
i9.1.0.0.....
.....
.....0.*.*0...0.....
.....(.....9.....Z.....6.T.i.m.e.s..N.e.w..R.o.m.a.n.
.R.e.g.u.l.a.r.....X.....($.....USUS.,.....(0.....'0.....N.thF.}....3|x.....
.....U.....!.....USUS.,....._.....Mary.had.a....little.lamb....whose.fleece.w
as.as.white.as.snow.
```

---

In this case, binary characters have been inserted between each word. Because our targeted string of text “little lamb” contains two words, this can have a negative affect concerning forensic text searches. Many of the older forensic search utilities would not recognize the targeted key words stored in this fashion. However, a DoD certified forensic search utility like NTI’s TextSearch Pro or TextSearch NT would identify the targeted strings of text.

### **WordStar Version 2.1**

Several years ago WordStar was an industry standard in the word processing field. When our sentence is saved in this format the resulting file is 396 bytes in size and the filtered output is displayed as follows:

---

```
. WS2000.1.00.... PRINTER..... 0 ...!Release.3.00... !... 0 0...i 0 0i.^ 1^.._  
1_[.K.....K.....$.)...3.8.=.B.G.L.Q.V.[.`.e.j.o.t.y.~.....[.a  
66a.\ 0\.] 0]..b11b..e0e..f1f..g1g..h00h..`1 3 2`...:033:...s011111s..v..z 9 3z..{ 9 9{..b00b.Mary  
had a ...little lamb... whose fleece was as white as snow.
```

---

Because all of the data is stored as text, a forensic search utility would have no difficulty identifying the targeted strings of text in this case.

Only a few of the different file formats have been listed. You should be aware that file formats can vary dramatically from one program to another and even from one version of the same program to another. For that reason, it is important that you are familiar with the computer applications used in your organization and the files that they create. The degree of risk is dependant upon a variety of variables and this is an important one.

## **1-9 Conclusions**

Unfortunately, personal computers were never designed to be secure. For this reason, inherent risks exist when sensitive data is stored on personal computer storage devices in government and business environments. Some of the risks associated with data leakage can be identified through the use of computer forensics tools and methods but there are limitations. The limitations deal primarily with data stored in obscure file formats, graphics files and compressed files which do not lend themselves to forensic text searches. When these types of files are involved the risks are increased in classified government agencies and manual reviews of specific files and data types are required. Once data leakage has been identified, the subject data can be eliminated through accepted data elimination processes.

Questions on any of the topics covered can be directed to Michael R. Anderson via E-Mail at [mrande@teleport.com](mailto:mrande@teleport.com).