

1.0 Security Space

1.1. *Basic Concept*

This document addresses the needs of the consumer and small business space concerning personal and corporate data. It outlines current practices and introduces the *dataSentinel* Storage solution. It describes the progression of technology that has led to the requirement for verifiable privacy.

1.2. *Supporting Documents*

The following documents pertain to the descriptions in this document.

1. *dataSentinel-10266-2-1* Distributed Storage Specification.doc
2. *dataSentinel-10268-2-2* Personal Information Space.doc

1.3. *Definitions*

- *Application*: A computer program that allows a user to view and manipulate data.
- *Hard Drive*: A mechanical device that stores large amounts of digital data.
- *Virus*: A malicious computer program that attempts to impair the user experience.

1.4. *Progression of the Computer Technology Market*

International Business Machines (IBM) dominated the early computer market for thirty years. During these early days of computing, designs were complex and highly interconnected proprietary implementations. IBM and its competitors each needed a complete in-house infrastructure to manufacture all the equipment and a large training force to bring it to the customer.

By the 1980's IBM had modularized their hardware designs to simplify the development of new models. They entered the personal computer market by outsourcing low-end modular hardware and an operating system believing that the value would be in the assembly and distribution of these systems. However, the hardware quickly became a commodity and many new companies began developing application software for the new Microsoft operating system. The market value then shifted to the operating system itself, as it became the component of greatest complexity. As in the early computers, the OS contained software components that were highly interconnected in a proprietary way and required the infrastructure of a large company such as Microsoft to build and maintain.

As the century turned, several new developments took place. The performance of computer hardware surpassed the requirements of the average user, so the PC industry moved from cycles of regular equipment upgrades to merely the replacement of older

computers. Ironically, a major driver of this market segment is the customer's preference to discard a computer choked full of excessive utilities, viruses and misconfigurations rather than go through the process of replacing the operating system.

Another development was the emergence of the open-source Linux operating system. Linux has gone through a modularization process that is similar in concept to what has happened to computer hardware. Linux variants that are assembled through integrator companies such as Red Hat and Debian have made the operating system itself a commodity.

It is unlikely that Linux would unseat Windows as the operating system of choice for the desktop computer as the interdependencies of current desktop software on Windows make the process of switching to Linux too painful. However, the emergence of mobile wireless devices represents a market in which the modularity of Linux can be leveraged to build acceptable solutions that cope with the limitations of mobile hardware. It is likely that a new generation of Linux based applications will appear on these devices that eventually spill over onto better mobile devices that supplant the desktop. Ultimately, market forces will force a commodity operating system upon all computing equipment.

Finally, the rise of the Internet as the dominant computer network for desktop and wireless equipment and the commoditization of hardware and software has pushed the frontier of development into distributed web applications. Google, eBay and Amazon have created large proprietary applications that run upon tens of thousands of inexpensive Linux machines. In this environment the hardware platform and operating system choices are irrelevant for both the service provider and the user. It is the overall distributed application that collects, manages and stores large amounts of public data that contains the value.

1.5. Storage capabilities in 2005

As computer technology has progressed, digital storage has been steadily improving. Early magnetic storage solutions stored a few megabytes on large drums. Current technology is approaching terabyte (1000 gigabyte) capacity per drive. Distributed web applications amass storage space measured in the petabytes. (1000 terabytes)

A look at a popular online computer sales site in 2005 will reveal personal computer bargains that offer systems with 300GB of hard drive storage for under \$1000. At the same time, a new vertical-orientation magnetic system promises to double the density of these drives. Clearly the average consumer or business user has plenty of cheap space to store pictures, videos, music or documents. Consumer electronics will ensure that this space will be filled. TVI technology places video on a hard-drive. It won't be long before a camcorder sports an IEEE 802 transmitter that allows video shot that day to be automatically uploaded to the customer's PC as soon as they return home.

This is good and bad. Clearly digital storage has been simplified for the consumer or small business operator by a plethora of software solutions that manage this data. All the data is in one place and could be mined (think Google desktop) to help find specific clips

and files. The bad news is that unlike collections of CDs, papers and VCR tapes, the entire accumulation can be wiped out instantly by a single hard-drive failure.

1.6. Backup Solutions

There is obviously a need to protect this lifetime collection of data against equipment failure. There are three prevalent approaches:

Security 0.0: Laissez Faire [graphic: ostrich with head in sand]

This approach has two tenets:

- My equipment will probably never fail
- If it does fail, I can pay someone to fix it

It is not that the owner does not place a value on their data. The problem is simply that the effort required to learn or execute a backup strategy is too large. There is also a naiveté that computer problems are repairable. Some are. A service might recover most of the data from a hard drive that has a bad sector but if the directory system is destroyed, the results might be thousands of indistinguishable data files that cannot be understood by the user's applications. Even this partial repair can be quite expensive.

Security 1.0: Local copies

The owner of the data can regularly burn copies of the data on to CDs or DVDs. This can be an ad-hoc process or it can be automated to a certain extent. Software utilities can track changes in the file system and prompt the user, but someone must be there to insert the CDs in the drive. For optimal protection against fire, etc., these disks should be stored off-site.

The small business owner might opt for a more sophisticated solution. By dedicating a computer to operate as a File Server and building it with Linux or Windows Professional, this machine can provide network drives for all of the other computers. If the users place all of the business data on these drives, then only the File Server will require backup. Tape drive equipment can be purchased that integrates with the File Server and provides a single tape for off-site storage rather than a set of disks. The backup procedure can be automated such that the only manual steps required would be to remove and replace the tapes.

This latter approach is more expensive. Not only are there costs associated with the hardware and possibly the operating system, but there are also personnel costs. Expertise is required for the administration of the File Server and the configuration of the tape storage subsystem. Should a data failure occur, knowledge is required to perform the reverse steps of obtaining the backup data from the appropriate tape and placing in the right position on the File Server.

Security 2.0: Offline Storage

There are hundreds of Internet companies that provide offline storage for data files. The simplest service installs a back-up program on a user's computer that automatically wakes up during off-peak hours to backup the drives to a remote computer on the Internet. The operator of the service performs tape backup on this data. The user is not inconvenienced by this process. Should their computer fail, the program is run interactively to allow them to pull back the files or directories that have failed.

Another class of services provides the equivalent of a Network Drive. As described above, the user sees a new drive on their system that looks and feels like their own hard drive. The difference is that this drive is actually hosted remotely on the service provider's equipment. The slower speed of Internet transfers versus direct hard-drive access is masked through caching techniques, and the user feels that they are working from a local drive. The big advantage of this type of solution is that the drive never appears to fail. It becomes the responsibility of the service provider to make sure that the contents of that drive are immediately restored from backup should an equipment failure occur. The user may experience an interruption, but once over, the file system is intact and no manual steps are required to pull backup files.

1.7. Why is Offline Storage unpopular?

Surveys reveal that most consumers do not actively backup their data. Small business operators do perform backups to CD or DVD, but in many cases this process is not performed regularly. As businesses grow, they turn to File Server technology and tape-backup processes and acquire the technical staff to operate them.

Paradoxically, the third option, an Offline Storage service, which seems like the ideal solution for individuals or small organizations, has not seen much uptake. The user effort is negligible and current offerings of this service are not expensive. A gigabyte of offline data storage cost as little as \$2 per month. The services usually protect data with the latest encryption techniques.

It is our view that the typical user does not have a sufficient level of trust in either the vendor or the technology to comfortably place their private documents and data on equipment owned by someone they do not know. Without an understanding of the technology, the average user suspects that given enough time, a malicious hacker can break the encryption of their files. They have heard news of the spectacular feats of these people in which credit card data is released and the personal data of celebrities is compromised.

1.8. Verifiable Privacy using the dataSentinel Storage Solution

Like computer hardware and operating systems, storage has now become a commodity. The cost of storage decreases each year and larger organizations that sell space are prepared to discount for larger volumes and repeat business. At the same time, increasing consumption of storage has made it more onerous for users to manage their own data backups. Consumers will ultimately want to turn to offline storage.

Users need to know their data is secure before they will place it on commodity storage. They need a conceptual understanding of how their data is stored and a simple way to monitor the mechanism. To this end, *dataSentinel* has developed a storage system that can be layered upon existing conventional commodity storage to provide verifiable privacy for the owner of the data.

The *dataSentinel* system is based on its InfiniDrive technology. It operates through mathematical principles. Each user file is broken in small blocks. The engine takes the name of the file and the Personal Encryption Code (PEC) of the user as inputs and produces a set of instructions to describe at which site each block of the file will be stored, what it will be called and how it will be encrypted. This set is unique for every file stored on a drive, and different for every PEC.

When a file is written to the *dataSentinel* system, InfiniDrive will write the blocks of the file to hundreds of different computers (Storage Peers in a large peer to peer network) across the continent on behalf of the user. When the user wishes to read the file back, InfiniDrive will regenerate the instructions by using the same inputs to the mathematical function, retrieve the blocks from the various Storage Peers, apply decryption and resequence the blocks back into the original file.

The high-level mathematical functions of InfiniDrive are designed such that the correct set of instructions cannot be reproduced without the user's PEC. The Storage Peers will only supply a block that is specified by name. These block names are based on a 64 bit binary number. Guessing the name of the next block in the sequence would involve counting through 2^{64} , an impossibly huge number. (18 billion, billion)

The *dataSentinel* Storage solution will allow the owner of a file to examine the collection of distributed blocks that make up any of their files, but only if they possess the PEC that wrote the file. They will be able to determine where geographically the Storage Peer containing each block exists, and more importantly, they will know if anyone other than themselves has read the block. The tools will inform them if any of the blocks of any files have ever been read and they will know with certainty that their data is safe.

1.9. Why is the *dataSentinel* Storage Solution better?

Conventional data protections systems are based on Symmetrical Key and Public Key Encryption technology.

Symmetrical Key Encryption is the uses a secret key of approximately 40 to 200 bits to alter the data of a file such that it is unreadable. The same key is used at a later time to unscramble the file back to its original contents.

Public Key Encryption uses two mathematically related keys: a public key and a private key. The user gives their public key to everyone but does not release their public key. A second party can encrypt a file with the public key and know that only the holder of the private key can decrypt the file to read it. Similarly, if the owner of the private key uses it to encrypt a file, the second party can be reasonably sure that the private key holder is the author of the file.

Public Key Encryption is much slower than Symmetrical Key Encryption. Typically, Public Key Encryption is used to privately transfer a Symmetrical key which is used to secure a transmission channel or the contents of a file.

The problem with these technologies is that they operate upon a complete file entity. Typically a data file is encrypted locally on the user's machine and then transferred and stored on the vendor's equipment. While the encrypted data file resides on the vendor's machine, it is intact and is vulnerable to a concentrated attack should the vendor's equipment be compromised by a third party. Typically all of the files are protected by the same Symmetrical Key. If an attacker is successful on breaking a small file then all of the rest of the larger files belonging to that customer are in jeopardy.

In effect, the customer is depending on the physical security of the vendor and the quality of the networking tools, practices and equipment to protect against hacking. Neither of these measures can be verified by the customer without a visit to the site or a detailed understanding and direct knowledge of the techniques that are used.

The *dataSentinel* Storage solution uses mathematics and geographical distribution to protect the user's data. For example, a file is broken into a hundred blocks is encrypted with perhaps a dozen or more unique symmetrical keys and stored across thousands of computers. A data thief would have to break into hundreds of computers containing millions of blocks without knowledge of the names of the blocks belonging to the file. They would face twelve decryption problems and have to place the blocks in correct sequence in order for the decryption to succeed. The results of such an attempt would not help with an attack on another file as the sequence and the keys would be totally different.

The *dataSentinel* Storage solution tools demonstrate to the user the complexity of the storage and the geographical distribution each file that has been saved in the network. More importantly, the *dataSentinel* tools can alert the user if attempts have been made to

read blocks of any file. If a pattern occurs, the user has plenty of time to create a new Personal Encryption Code and move the files into a new area based on fresh encryption. The knowledge the *dataSentinel* tools are constantly monitoring the security of their personal data gives them an assurance that conventional offline storage providers cannot offer.