This paper is not intended to focus attention on any specific backup program. Rather, it strives to give the reader insight into some of the underlying technology and recent innovations in backup systems.

Current Technology

To meet the demand for a centralized "Server Centric" backup policy, software developers have created some noteworthy applications. Many utilize "clients" or "agents" residing on workstations that permit the server access to remote workstation files during an enterprise backup session.

However, there is one major underlying factor that diminishes the effectiveness of most of these programs: Full file incremental backups. More specifically, any minor change in a file requires the backup of the entire contents of that file. There are obvious ramifications as the size of data files increases and the network bandwidth to back them up decreases.

This gives rise to an important observation: While workers may be creating larger files, daily changes to those files are, on average, small. This leads to the obvious conclusion that if there were a procedure in place to permit the extraction and backup of only those portions of a file that changed from day to day, backup size and time would dramatically decrease.

The Next Step

While not new, the concept of backing up only binary changes to data has, nonetheless, eluded backup software developers. However, if one delves into the actual mechanism of such a function, one quickly realizes that the process is not as straightforward as first observation might suggest. In fact, it is fairly complex. It is this complexity that has relegated the concept to being just that, a concept, until now.

Recently, programs that perform "televaulting" or off-site backups have been receiving significant press. Utilizing standard telecommunications devices such as modems and ISDN adapters, these backup applications collect and backup changed data to a remote site. However, if one scrutinizes this process carefully, one quickly realizes that, using current technology, such an application would have little use in a large business environment.

To increase acceptance of televaulting as a viable backup solution for the corporate world, developers have invested a significant amount of time and expense into improving the underlying technology. Two significant innovations have come from these efforts. Both of which permit discrete data changes to be backed up instead of the entire file.

Block Technology

The first innovation to come from the development of the latest backup software is referred to as "Block Technology". In one form or another, Block Technology has been around for some time and was originally developed as a method for mirroring data from one hard drive to another.

In essence, the Block Technology process evaluates changed data by breaking a file down into discrete blocks of information. These blocks are typically between 1 and 32 kilobytes in size. Through the use of a cyclic redundancy check (CRC), Block Technology compares each block of a modified file with the corresponding block in the previous version of that file. When the process detects a difference, it extracts a copy of that discrete block, not the entire file. In practice,

changes in files will usually result in a number of blocks being copied. However, the cumulative size of these blocks will be less than that of the original file. This has the effect of reducing the total backup size and time.

However, observing Block Technology in action reveals that it produces larger file sizes than one would expect. This is, in part, due to the use of fixed block sizes. If only 100 bytes of data has been changed in a file, but the block size is 4 kilobytes, the entire 4-kilobyte block is extracted. Combine this with similar changes to other blocks and one will observe that the size of the extracted data can be significantly greater than the actual size of the changed data.

Binary Patching

The second backup technology making headlines today is "binary patching". Originally developed over 8 years ago as a method for upgrading software, binary patching has received widespread acceptance by many of the world's largest hardware and software manufacturers including IBM, Compaq and Microsoft.

To cut costs and decrease the time to market, manufacturers distribute their updates as tiny files or "patches" containing only the binary difference between the old and new version of their software. Once received by the client, these patches are applied or merged into the existing file instantly upgrading it to the latest release. An obvious advantage is that the size of the upgrade is reduced significantly. This permits clients to use modem dial-up connections to obtain major software updates instead of the more traditional forms of distribution such as floppy disk or CD-ROM.

Although binary patching may sound similar to Block Technology, it differs in one significant aspect: Binary patching does not evaluate a file as a collection of discrete blocks, rather as a continuous string of binary data.

Utilizing a complex algorithm and special memory management, binary patching is capable of comparing files and extracting "patches" of binary data that represent only the specific changes to those files. Simply put, If only 1 kilobyte of data has actually changed in the file, then only a 1-kilobyte patch is extracted for backup, thus eliminating the overhead imposed by Block Technology methodology.

Observing the binary patching process, one can quickly see a significant decrease in backup size over that of the Block Technology system. This is clearly demonstrated in table 1, which outlines the results of a carefully designed and executed test.

Empirical Comparison

To better understand the effectiveness of Block Technology versus binary patching, a simulated workflow model was created that closely approximated that of the average business-computing environment. Table 1 outlines the results of applying this workflow model to a group of 5 file sets that one might find in the average corporation.

While it is obvious that each technology produced backup files substantially smaller than the original, it is evident that binary patching significantly outperformed Block Technology in every instance. Moreover, while the results may seem inconsequential at this level, when multiplying these figures by the large number of users an average corporation might have, the difference becomes staggering.

FIRSTBACKUP
YOUR DATA IS OUR FIRST PRIORITY

TABLE 1: Backup Technology Comparison

	Full Backup	Group 1	Group 2	Group 3	Group 4	Group 5
Base Line[+]	15 Mbytes	4.4 Mbytes	5.2 Mbytes	4.8 Mbytes	6.8 Mbytes	4.1 Mbytes
Binary Patch	4 Mbytes	250Kbytes	360Kbytes	**69Kbytes**[*]	**512Kbytes**[**]	242Kbytes
Block Technology	4 Mbytes	**561Kbytes**[*]	1.3Mbytes	1.1Mbytes	1.5Mbytes	**1.6Mbytes**[**]

+Uncompressed size of changed files; *** Best case**; **** Worst case**

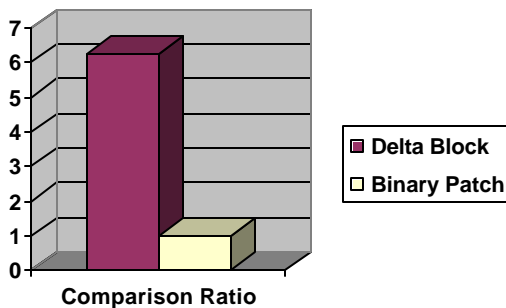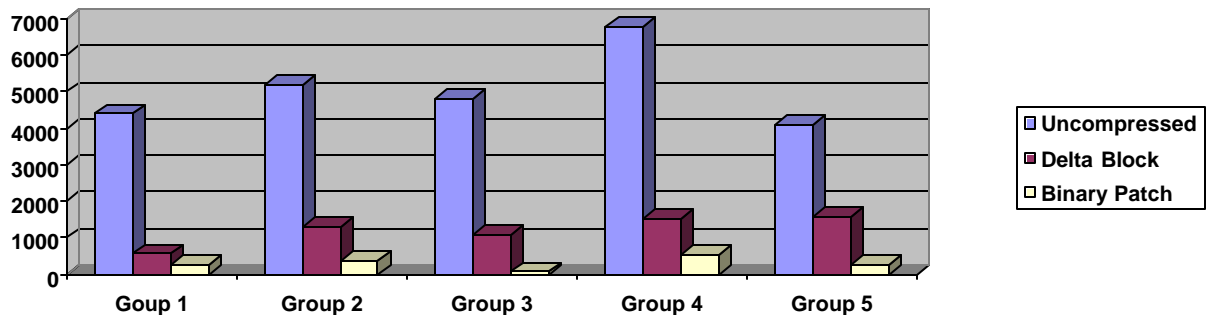**Chart 1:** Bar Chart displaying results from TABLE 1

Chart 2: Average difference between Block Technology and Binary Patch

This is a ratio showing the average difference between the incremental backup sizes of both Block Technology and Binary Patch technology. This clearly shows that Block Technology-based applications will require a significantly greater amount of bandwidth and server storage to accomplish the backup of the same amount of user data.

CONCLUSION

In the past, corporate and IS professionals alike have viewed data backup as a necessary evil. While most professionals understand the ramifications of a poorly conceived backup strategy, many hesitate to develop more formalized procedures in light of reduced resources and network bandwidth. Furthermore, even IS managers with adequate resources often relegate the backup process to a subordinate task because of the increased management required.

In an attempt to simplify the backup process and reduce data flow over current network infrastructure, backup systems incorporating Block Technology and binary patching have been developed. Because of their ability to extract small changes that occur in data files, these new technologies will lead the way toward improved backup procedures for all who adopt them.

However, as in any market, one finds that not all products or technologies are created equal. A unique few will rise and stand above the competition. Binary Patching is such an example. With its ability to extract only that data which has changed within a file, binary patching promises to lead the industry by greatly reducing the burden that current backup systems place on networks and IS professionals alike.