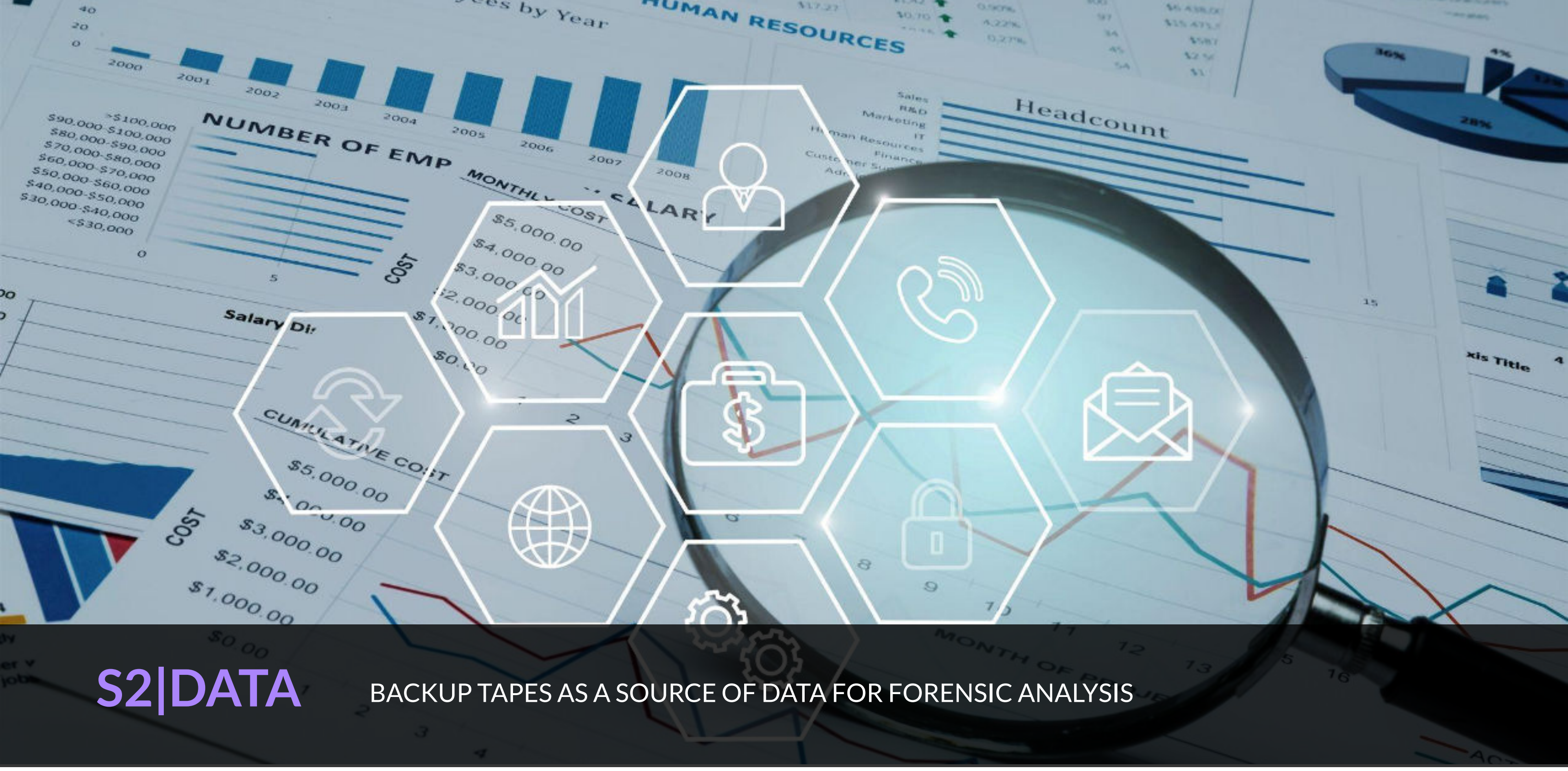
WHITE PAPER

Backup Tapes as a Source of Data for Forensic Analysis



S2|DATA BACKUP TAPES AS A SOURCE OF DATA FOR FORENSIC ANALYSIS

S2|DATA is in the business of helping its clients find out what is in your backups.. Once we identify a file, backup session or database of interest, it may be restored and handed over to the party required to produce it.

This service could be performed under the general term of requirements for:

- Business use
- Regulatory or Compliance use
- Discovery use

Even though S2|DATA typically provides this work as part of a broader discovery request for a litigation matter, it is not considered a sub-division of computer forensic analysis. Computer Forensics is the application of computer science to the courtroom or arbitration panel. The acquisition of data from PCs, laptops, smartphones, etc., do fall under the term Computer Forensics and specifically result in the performance of various analytics.

Electronic discovery focuses on the content of documents, emails and databases, and tells a story based on their content.

We consider this a fair distinction between the two storage mediums: data on disk is forensic analysis, but data in your backups is electronic discovery. It is fair to say there is good reason for the distinction. However, should there be a term “backup forensics”?

Data as a Snapshot in Time

The fact is that computer forensic evidence loses integrity over time. Timestamps are overwritten, logs are flushed, entire files are deleted. When putting together a documents story from years earlier, it can be much more difficult due to the loss of key metadata. This is not so with backups. While disk/ flash-based computers have unallocated sectors and deleted space to investigate, backup data typically does not. The main distinction is that backups are almost always a backup of active files on the file system rather than a backup of the entire disk, including data marked as deleted by the file system, but has yet to be overwritten. It's essentially a representation of data as a snapshot in time. Therefore, every backup could be seen as a forensically sound collection of files the second the backup is taken. Because backups are sequentially written, it is impossible to modify any section of data backed up data without corrupting the entire image or tape. While the potential of forensic analysis is reduced because many PCs aren't typically backed up at the system level. A significant amount of forensic analysis can still be performed by focusing on the file metadata and file content captured in server backups.

The metadata that typically is captured in a server backup includes the filename, file extension, file size, the path that leads to the files, its MAC dates (last modified, accessed, or created date), and of course, an MD5 hash can be created on the file to determine if it is located in other places. Many office documents internally capture metadata such as the last editor.

If working from a single forensic image or a single backup, knowing the last editor and the final content only tells a part of the story. Many documents go through a series of revisions. There may be numerous people editing the document and adding, removing, and changing the content. When a document is key to a litigation, building up the history of the document can often only be done by analyzing multiple versions of the document as captured in a series of backups. With that series of document versions, the legal team gets a full and more convincing story of who was editing the document and what changes have been made. Telling the documents' life story is the responsibility of computer forensics and having multiple versions available from backups allows for a more accurate tale to be told.

The metadata on files are a much deeper ocean, with even more potential stories to be told. In fact, some metadata can be created from file content to provide accurate forensic analysis and interpretation on a data set.

S2|DATA BACKUP TAPES AS A SOURCE OF DATA FOR FORENSIC ANALYSIS

Manipulation of Data

Backups are regularly performed by companies using tape and disk images, and each one has copies of files as of that point in time. A tape backup or disk image typically only saves complete files and not a complete disk image, so very few backup programs will save unallocated space. With a standard file system, a skilled person can modify dates to make the file look as if it was created or modified at a different time. This sometimes proves difficult, as some files save the dates internally as well as in the file system. With the a tape backup or image on disk, it can be shown that a particular file did or did not exist on a specific date with these contents.

Backup programs have different parameters surrounding when data is saved. The most superficial level would be just selected subdirectories. The highest level will include all system state files, and on these, it would be possible to analyze the registry. With regular backups, either total or incremental, it is possible to build up a picture of system usage, how often critical files have been changed, and each file version.

A backup is just a storage medium, so it can still be subjected to formal forensic analysis once files have been recovered. A straightforward tool is file signatures. This is a process where the start of a file is typically examined, and the expected file extension is determined. For example, if files that look like JPEGs are all called GPH, it is possible somebody has tried to hide them. With S2|DATA's proprietary software, TRACS*, this test can be done while scanning the backup.

Backup data and disks are both starting points for forensic analysis; they may give different but equally valid results. Disks have unallocated clusters, and backups have frozen data. Otherwise, they are very similar. The file metadata is identical for both disk and tape. We find that when the full story needs to told, discovering data from backups in addition to disk is highly effective.

TRACS – **Tape Restoration & Cataloging System**,
a proprietary software application developed by S2|DATA.

Authors –

- Michael Cotgrove – Senior Software Developer, S2|DATA
- Brendan Sullivan – CEO, S2|DATA