Tape based storage within the School of Informatics

Version 1.1 – 17/11/14

Within the School of Informatics and its predecessors, the preferred medium for the longer term storage of data has long been magnetic tape of various formats. Tape based storage has served many purposes: disaster recovery, off-site storage, restoration of accidentally deleted data and archiving (in the days when one month of full backups were retained theoretically for ever but in practice until we no longer had a working drive capable of reading the tape format).

In the last few years, the storage landscape has changed radically and so has the way in which tape is used within Informatics. Data which cannot be easily re-created is now mirrored nightly to off-site disk providing a first line disaster-recovery capability, and, after analysis of the requests received by User Support for recovery of data from backups showed that practically all such requests concerned data less than a year old, all tapes are now recycled after 13 months meaning that the archiving function has been eliminated. This leaves tape based backups fulfilling the following three functions in Informatics:

1. Second line disaster recovery. If data is lost and for some reason, the nightly disk mirror is unusable, disaster recovery can be performed from the tape backup. The robustness of this facility would be considerably enhanced if the tape library were located somewhere other than next to the racks containing the mirror servers. As the School expands into other sites, it would make sense to move the tape library from the KB server room.
2. Restoration of data less than a year old. Potentially data which has been inadvertently lost within the last year can be retrieved from tape. In the last three years, we have received an averaged of less than one restore request per month. Though there is of course no way of telling how important the data restored was, such an under-used service seems like a poor use of resources.
3. Detection and restoration of corrupted data. It may be that due to malicious action, subtle hardware failure (for instance bad memory in the cache of one of the disk arrays) or software error, data held on disk becomes gradually corrupted in some way. Resorting to the disk mirrors is useless because the corrupt data will have been faithfully copied across. But, providing that the corruption occurred during the tape retention period, tape storage offers the possibility to go back through earlier versions of the data until the start of the corruption is found. This may be useful in both restoring the uncorrupt data and in determining the causes of the corruption.

Future Costs

The 5 year warranty purchased with our current tape library expires in July 2015. If we are to continue to use tape as a storage medium, we have two options:

1) Pay for maintenance cover. Quotations have been obtained from two sources and come out to roughly £9000 per year.
2) Replace the existing library. Little attempt to ascertain the cost of this option has been been made as yet but there is good reason to think that the costs will not have changed much from the last procurement exercise where a new library and 5 years warranty cost £50000. The new library would certainly come with a later generation of LTO drive. If we procured a library with LTO5 technology, we could in theory continue to use our current set of LTO4

tapes gradually replacing these to take advantage of the increased capacity of LTO5. If we went for a library with LTO6 drives (LTO6 was released in December 2012), then we would only be able to read from our existing LTO4 tapes resulting in the added expense of buying a large number of new tapes. The current tape library holds 500 tapes and latest generation LTO tapes are roughly £40 each (excluding VAT) giving an additional cost of £20000 to fill the library. Of course not all of these tapes would have to be purchased at the same time.

To sum up. To continue to use tape within the School, we are faced with an expenditure of at a minimum £45000 over the next 5 years to continue to use our existing equipment and potentially £70000+ to replace our existing tape facility. In practice, during those 5 years, our current library will almost certainly be end-of-lifed leaving us with no option but replacement. Judicious extending of the warranty would allow us to procure the new library at an financially optimal time.

Questions to be answered

Does the continued use of tape justify the expenditure of over a quarter of the annual computing budget at some point in the next few years? Certainly, the underused data restoration service for lost and deleted files seems to offer little support for the continued use of tape. The other uses are harder to quantify. The first source of disaster-recovery data will always be the disk based backups and gradual and at first undetected corruption of vital data may never happen. But one incident where a tape based second line of defence allowed the School to quickly restore services in the event of a disaster or to avoid loss of reputation due to data corruption might easily justify all the expense of continuing to use tape based technology.

In addition, it may be that new uses for tape based storage can be identified in the next few years. Data curation is becoming more and more relevant to the School and our forthcoming audit of Medium and High Risk data will undoubtably identify sets of data with differing retention, storage and expiry requirements which are too diverse for the central services to cope with. Although tape is a poor medium for data curation in general since it is obviously impractical to delete some of the data on a given tape while retaining the rest, a minor extension to our existing backups setup would allow us to back up different data sets to separate tape pools allowing us to retain some date beyond the commodity 13 month retention period and even remove the relevant tapes from the tape library for remote storage. This would be far harder to accomplish with, say, disk based storage.