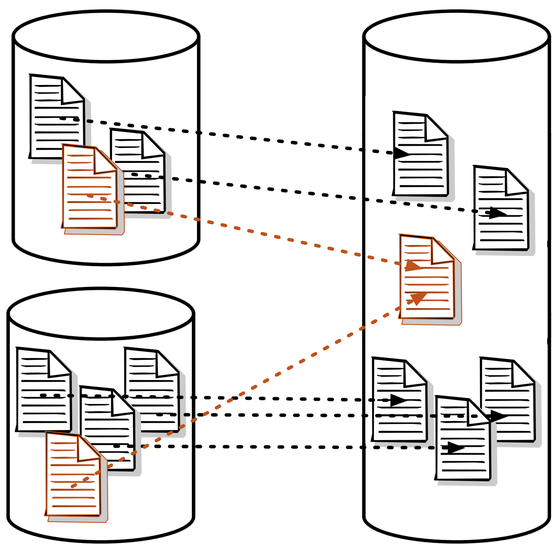
Fast and Secure Laptop Backups

High-speed personal data backup with space-efficient Convergent Encryption

THE UNIVERSITY of EDINBURGH informatics

De-duplication - even if many people have a copy of the same file, *de-duplication* techniques (or *content-addressable storage*) allow us to store just one copy on the server.



For typical laptops, we are seeing space savings of 40-50%. In some cases, savings of up to 80% have been reported (*).

Perhaps even more important than space savings, are the corresponding time savings - if someone else has already installed a particular application or document, then it will already be present in the backup system and will not need to be re-copied.

Convergent Encryption techniques allow us to de-duplicate files, even when they are encrypted by different users. The files are encrypted on the client and inaccessible to others users, or the service provider.

We are using the strong Salsa20 encryption algorithm.

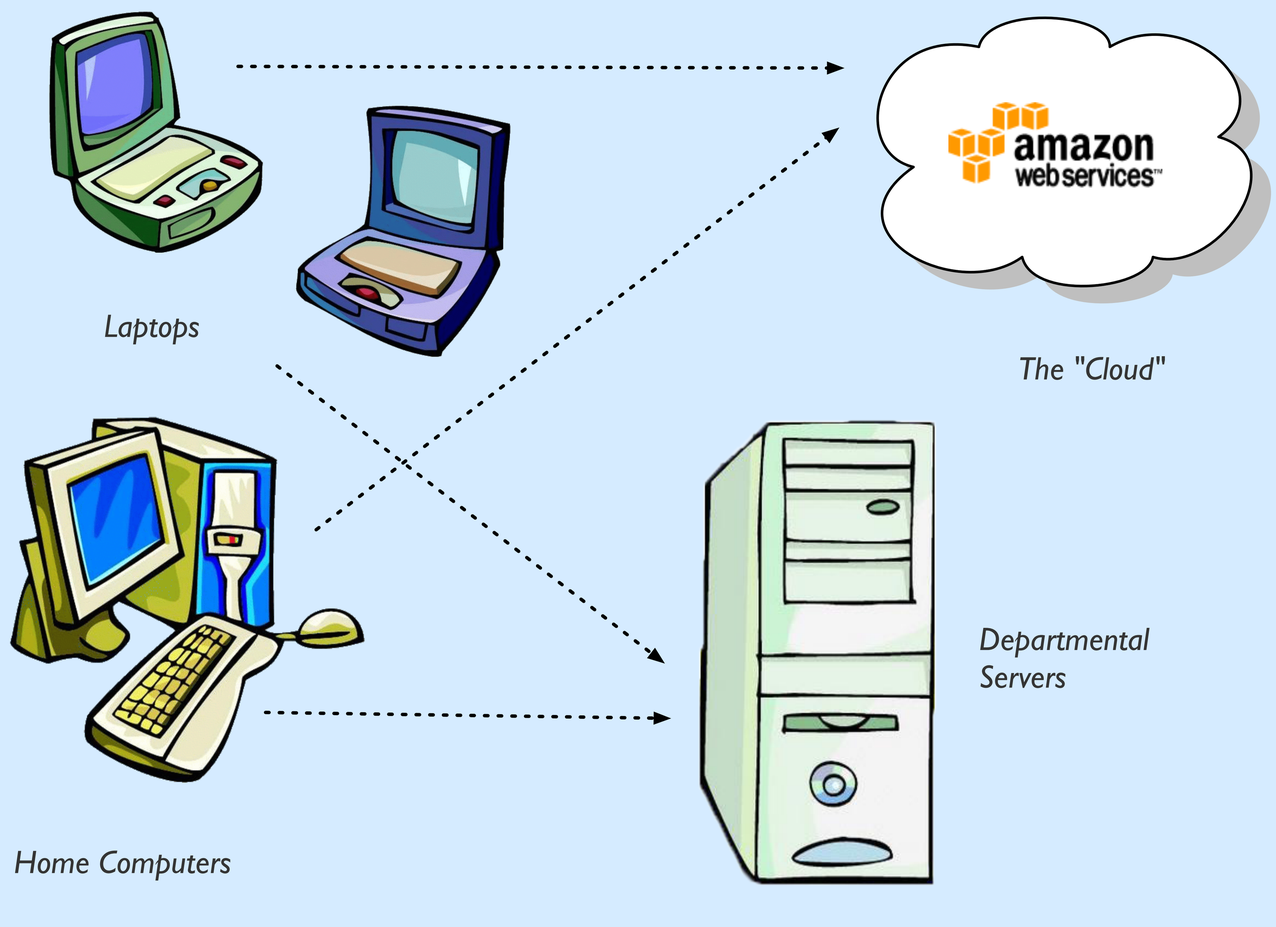
(*) Austin T. Clements and Irfan Ahmad and Murali Vilayannur and Jinyuan Li
Decentralized Deduplication in SAN Cluster File Systems
Proceedings of the 2009 Usenix Annual Technical Conference

But Cloud storage is comparatively slow and expensive, and I don't trust it ...

Our client software will work with a departmental server running free software such as Eucalyptus.

But I only need to backup my "user files" - I don't need to backup the operating system files ...

How confident are you that there aren't some critical configuration files among those OS files ?



But I use Time machine for my Mac backups, so I don't need this ...

Time machine backups are not stored "off-site", so if there is a fire, or a flood, all the data will be lost.

Time machine backups are not normally (*) encrypted either, so if there is a theft, your private data will be exposed.

(*) Unless you are using Filevault

But my company SAN does de-duplication and so does ZFS ...

The de-duplication will be performed on the server-end, so it doesn't save any time when doing backups over a slow link such as home broadband.

The encryption also occurs on the client, so encrypted backups are not possible in conjunction with the de-duplication.

But there are lots of other "cloud backup" services already available ...

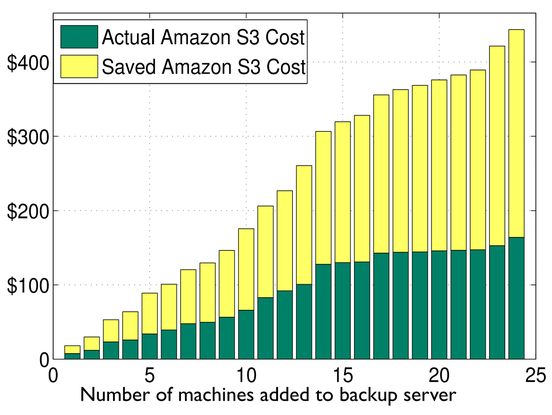
If they don't use de-duplication, then they will require a lot of storage and be very slow:

"I have a home Internet backup service and about 1 TB of data at home. It took me about three months to get all of the data copied off site via my cable connection, which was the bottleneck. If I had a crash before the off-site copy was created, I would have lost data" (*)

(*) Henry Newman
"Why Cloud Storage Use Could Be Limited in Enterprises",
Enterprise Storage Forum

Do you have any data to prove that it works?

A pilot study among local Mac users shows up to 30% of data can be shared, and up to 50% cost saving can be achieved when backing up to Amazon S3:



Over a six-month period the estimated accumulated saving is $1,340 for backing up 4 TB data from 24 machines to Amazon S3. That shows a cost reduction by 35-40%

Collaboration Opportunity

Having developed a running prototype, we are looking to develop relations with potential investors and technology partners to commercialise this technology.

Current Status:

- Working demonstrator on Apple Mac
- Developing commercialisation plan

We are looking to:

- Build relations with future investors
- Talk to partners who can help us demonstrate and share the benefits of this new backup technology

For further information, contact Le Zhang <zhang.le@ed.ac.uk>