Comparison of Dropbox to a Distributed Storage System

Shawn Wilkinson
me@super3.org

Abstract

Cloud storage services have grown to fill the growing need for consumers to store and organize their data. While these services work well enough for most needs, they have several inherent flaws related to cost and security. This paper proposes a system that allows users to sell their computer storage space back to a network in return for an digital incentive. This converts what would be unused storage space into a low-cost storage network. Furthermore, layers of encryption and hashing are added to provide unbreakable digital security.

Introduction

In the 1980s you could purchase a Morrow Design's 10 megabyte hard drive for a low price of $3,695 [1]. Of course if you offered that deal to someone now, most people would laugh. One could pick up a several gigabyte thumb drive for under $10 at the nearest electronics store, and probably have more digital storage space than existed in the 1980s. This reality demonstrates the rapid increase of storage media, as it struggles to keep up with the increasing demand for data in a computer driven world.

Services like Dropbox, Google Drive, and Microsoft OneDrive (so called cloud storage services) were built to fill the growing need for consumers to store and organize their data. Other services such as, Amazon S3, Google Cloud, and Microsoft Azure provide for the same need, on a much larger scale, to enterprise customers. While these services work well enough for most needs, they have several inherent flaws related to cost and security.

Costs

Hardware Costs

Dropbox, the online file storage service, lists the pricing for their "Pro" plan at $99 a year for 100 GB [2]. At approximately $1 per GB per year, the cost seems reasonable until you compare it to the costs of raw hardware. For example, I have a $60 2TB hard drive and reasonably fast internet connection. If I sold my hard drive space at Dropbox prices, I could charge $2,000 a year.

If I turned off my computer or my hard drive died, the files I was storing would be inaccessible. For a more realistic and fair comparison, we would need to add more people. A group of people including myself, Alice, and Bob connect our computers together to form a distributed storage network. They all bought the same 2TB external hard drive that I did for $60. We can set any file stored on the network to be mirrored on 3 computers by default, creating a 3x redundancy. Dropbox uses Amazon S3 for storage, with a similar 3x redundancy [3][4].

Now, if a file stored on this network is on Alice's, Bob's, and my computer will be safe unless we all decide to turn off our computers simultaneously. If Bob turns off his computer, another computer will simply download the file from Alice or myself, and return to a redundancy of 3x. By creating the redundancy at the network level, we ensure that hardware failures or disconnected computers are irrelevant to preserving the files.

When we account for the price of the hardware, we end up with a potential profit of $1,820 per year split among Alice, Bob, and myself. Using these estimates, this storage network is ~12.5x cheaper than Dropbox.

Power Costs

We established the concept that our theoretical network is significantly cheaper if we simply factor in hardware costs. Let's next examine power costs for specially running a computer for our distributed storage network.

So what does it cost to run a hard drive? According to some hard drive power consumption stats from Tom's Hardware [5], we get an average of 7.16 watts per 3.4 TB, or about 2.1 watts/per terabyte. According to the U.S. Energy Information Administration, the average U.S. price per Kilowatt hour is $0.1209 [6]. Using these stats, lets figure out the raw cost for a 100 GB storage package on our network. Our 100 GB consume ~0.205 watts for a total consumption of 1.796 kilowatt-hours per year. So the raw power costs for our 100 GB storage package is $0.22 per year.

Sunk Costs

Because we are running our distributed storage network mostly on personal computers, for the time being, we can using sunk cost to further reduce our cost levels. Whereas a cloud storage provider, like Dropbox, has to maintain servers, data centers, building space, and employees, we have none of that overhead. So what is a sunk cost? A sunk cost is “a retrospective cost that has already been incurred and cannot be recovered.” [7] If you have a computer with a 1TB hard drive, and you are only using 200 GB of space, you could sell that extra 800 GB of space to this storage network. At our Dropbox rates thats around $800 a year or $67 a month. That is a pretty convincing inventive to get someone to provide their extra storage space.

Incentives

Incentives

So since internet, computer, hard drives, power may be paid for regardless of using the distributed storage network, our costs approach zero. We do, however, need to provide some incentives for the user to participate in our network. If you leave your computer on specifically to help run the distributed storage network, you are of course using power, so we have to at least cover this in our incentives. If we want users to run dedicated computers for our network, we need to at least be able to cover hardware costs.

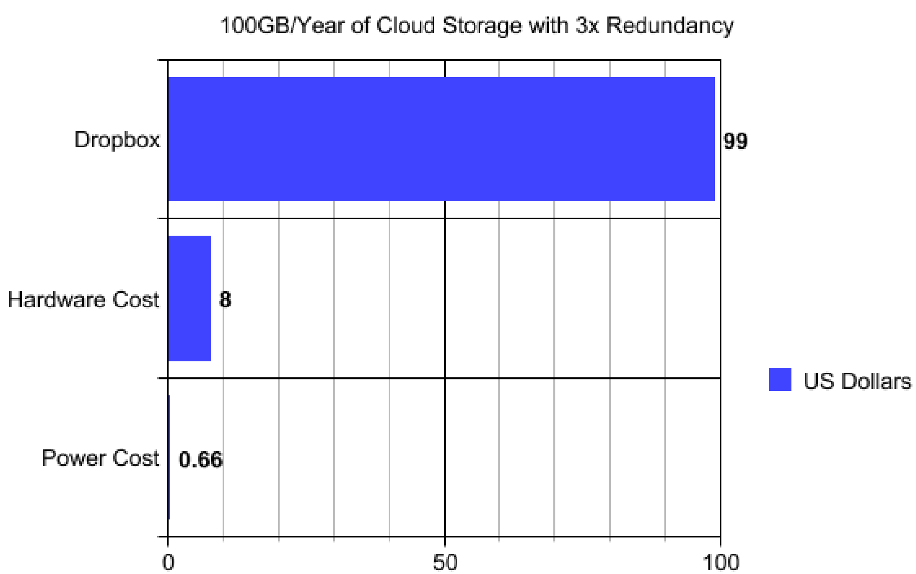

Figure 1

As shown in Figure 1, there is a large gap between our core costs, and how much we could charge for the storage space.

Tokens

Of course we could not redeem 10 GB of storage for our power bill. To solve this we use a digital token [8]. This token could be a specialized cryptocurrency like Spacecoin. This token would be given in exchange for providing storage, bandwidth, and maintaining the network. Alice could provide 10 GB of storage to the network, and in return get a token worth that amount of space. Alice could now store ~3 GB of her own files on the network (remember we need to account for redundancy), sell her Spacecoins for Bitcoin or US dollars on an exchange, trade her Spacecoins for good or services, just hold her Spacecoins for when she needs storage later, or hold her Spacecoins hoping she can sell them for more later.

Digital incentives has made the Bitcoin network the most powerful computing network [9]. We aim to mirror that in our distributed storage network.

Solving Security Pitfalls

Trust

Data security is a large concern. If I store my file on another person's computer, how can I know they won't access or modify my file? Currently users rely on a trust model for data security. When I upload my files to a service like Dropbox, I assume that the files are securely uploaded and no one but myself is able to access them. Despite efforts to keep data secure, these cloud storage providers experience many failures and data breaches [10].

Without complex or expensive encryption software, the average consumer faces difficulties storing data in a secure way. Companies and corporations, especially those who deal with sensitive data, choose to use the private cloud, at increased cost and reduced scalability, in order to keep a better handle on their data. Even with these methods, the data remains susceptible to a range of traditional security attacks.

Our distributed storage network works in a trust-less environment. I have no guarantee that someone on the network might have malicious intent. In order to address the inherent discomfort individuals and companies have with storing files on a stranger's computer, we using basic encryption and hashing methods.

Our Solution

All data, regardless of content, is encrypted on the client side, that is on your computer, before being transmitted through the network. At the storage destination, the file will retain its encrypted state. We use the hash of the file to verify that the file is not modified in any way. If even a single character of the file is modified the hash of the file will be different. Furthermore, unless the file is very small, it will be split into parts and distributed among the network. Using these methods, our file cannot be accessed, modified, or reconstructed by an unauthorized party.

Conclusion

We have laid out a rough outline for a distributed storage system. This proposed solution has major advantages in terms of costs and security. There is also several additional benefits in terms of scalability, privacy, and use cases that have not been addressed in this paper. We will save more technical specifics including proof-of-storage for another paper.

References

[1] J. Chung, "10 Vintage Hard Drive & Memory Ads That Question Value For Money," http://www.makeuseof.com/tag/10-vintage-hard-drive-memory-ads-that-question-value-for-money, 2013.

[2] "Dropbox - Upgrade account," https://www.dropbox.com/upgrade, 2014

[3] "Does Dropbox keep backups of my files?," https://www.dropbox.com/help/122/en, 2014

[4] "Amazon S3 Redundancy," https://forums.aws.amazon.com/thread.jspa?threadID=143942, 2014

[5] M. Masiero, "Seagate Desktop HDD 4TB Review: Big Capacity At 5900 RPM," http://www.tomshardware.com/reviews/desktop-hdd.15-st4000dm000-4tb,3494-6.html, 2013

[6] "Electricity Monthly Update," http://www.eia.gov/electricity/monthly/update/end_use.cfm, 2014

[7] "Sunk Cost", https://en.wikipedia.org/wiki/Sunk_cost, 2014

[8] D. Johnson, "The Emerging Wave of Decentralized Applications," https://github.com/DavidJohnstonCEO/DecentralizedApplications, 2013

[9] S. Cowley, "Bitcoin more powerful than fastest supercomputers," http://money.cnn.com/2013/05/23/technology/enterprise/bitcoin-supercomputers/, 2013

[10] "Dropbox gets hacked ... again," http://www.zdnet.com/dropbox-gets-hacked-again-7000001928/, 2012

Further Reading

G. Maxwell, "StorJ," https://bitcointalk.org/index.php?topic=53855.msg642768#msg642768, 2013

J. Garzik, "StorJ, and Bitcoin autonomous agents," http://garzikrants.blogspot.com/2013/01/storj-and-bitcoin-autonomous-agents.html, 2013

J. Garzik, "Blackbox: Bitcoin-enabled, decentralized cloud," http://garzikrants.blogspot.com/2014/01/blackbox-bitcoin-enabled-decentralized.html, 2014

M. Hear, "Mike Hearn, Bitcoin Developer - Turing Festival 2013," https://www.youtube.com/watch?v=Pu4PAMFPo5Y, 2013