Metadisk: Blockchain-Based Decentralized File Storage Application

August 1, 2014

Abstract

Metadisk is an open source software project seeking to prove conceptually that cloud storage applications can be made more decentralized, more secure, and more efficient. In addition, Metadisk provides a prototyping platform for a fully decentralized network. In pursuit of this goal, we propose developing a web application that provides an interface for non-technical users, and an underlying API for native applications and feature extensions. A cryptocurrency will serve as both an incentive and payment mechanism while a separate blockchain will be used as a datastore for file metadata. This application will seek to operate autonomously as a peer-to-peer network of nodes running open-source code, depending on a public blockchain for information rather than a central database. Metadisks primary objective is to provide a stable testing platform for a peer-to-peer cloud storage network called Storj. Metadisks ultimate objective is to provide a set of tools that will allow Storj to integrate more easily with traditional platforms and users.

Introduction

Cloud storage is a marketing term that exploits the popular craving for novelty. It has caught on because to ordinary users, the cloud sounds like a newer technology when compared to the internet, client-server, or as a service. This rebranding of existing technology is simple misdirection, not magic. When data is stored in the cloud, it is transferred over TCP/IP from the clients computer to the hosts server in a data center. It is the same old client-server model that has existed since the days of the mainframe and dumb terminal. That server then copies it to other servers to comply with industry standard redundancy policies where three copies are made. The current model of cloud storage through centralized institutions that are entrusted with private information is inherently insecure in many ways. Information thieves, spies, and censors can seek to copy or destroy data stored on the host servers through political strategy, legal tactics, and technological means.

Fig 1. The Standard Model for Cloud Applications

The distinction between these three categories has become increasingly blurry over time. It is now clear that personal privacy and enterprise information security can only be achieved if off-site data stores can be protected from attacks originating in and operating through each category. Having easily identifiable central points of attack built into the model is a problem solvable through decentralization and automation. Other inherent security flaws in the current cloud storage model are the types of payment mechanisms currently in widespread use by cloud storage services. These mechanisms are neither private nor information secure because most online payment technologies store and leak information about payer and payee.

We need a cloud storage model that is not based on trust between client and host. All potentially private data, including filename, date and other metadata, must be encrypted before any transfer takes place from a clients computer to the cloud. There will be no centralized point of attack using political or legal attack vectors. All incentive payments for both resource providers and consumers will be automated and made in a pseudonymous cryptocurrency. It is time for the cloud to truly become a cloud, made up of a vast multitude of resource droplets that are added and subtracted as the cloud forms, moves, and changes shape.

The overarching design principles enabling a decentralized storage network have been known for several years. Projects like MaidSafe [1] and Tornet [2] have outlined possible solutions. Unfortunately, achieving the security, scalability, and cost efficiency of a truly decentralized storage system will require software of immense technical complexity. We must design the nodes and network in an extremely secure manner as we can trust neither the lines of communication nor the nodes themselves. Nodes on the network must collaborate to achieve the level of redundancy and performance of current centralized networks. Furthermore, the software must run in an unmanaged environment, very different from the current cloud networks.

All of these are solvable problems using a variety of existing technologies such as BitTorrent Sync [3], Bitcoin [4], public key encryption, and cryptographic hash functions. Metadisk aims to simplify the development of a decentralized storage network by allowing integration with other existing open source projects in a modular fashion. Such a network must be built incrementally, with software that consists of interoperable, easily replaceable modules that support a wide variety of hardware, including the hardware as a service model that existing cloud storage service providers offer.

Using a model similar to that employed by Bitcoin, where miners are paid block rewards for contributing cryptographic hashing power resources to the network, Metadisk can use cryptocurrency to reward and buy storage space and bandwidth on the network. This model harnesses the powerful free market force of individual self-interest to drive the network to grow and become more efficient while remaining decentralized. For example, if another faster method of file transfer is found, the network will gravitate toward that method until someone produces an even faster method. Nodes must be able to work in a constantly changing and improving environment.