Silo: Immutable and anonymously addressable storage

Kee Jefferys* Sam Williams† Simon Harman‡

April 2020

Abstract

Silo is an abstraction on top of an immutable storage medium called Arweave. Silo provides a way to upload and share immutable data anonymously, by using a novel human readable content addressing scheme, and allowing data to be accessed using the Lokinet anonymous onion routing layer.

1 Introduction

For thousands of years humans have been attempting to share secret and confidential pieces of data without the knowledge of external parties[1]. However the methods to perform such an exchange secretly have always been difficult to scale, unreliable and prone to censorship. It is only in the past 10 years that this situation has begun to change: With the invention of blockchains economic incentives could be provided to store transaction data across a set of distributed nodes with recursive hashing as a way to address those transactions securely. More recently this concept has been built upon to allow for general data storage in storage layers like Arweave[2]. Arweave addresses the problem of the storage of such secret data, however the anonymous distribution and addressing of such data remains a problem.

Addressing remains as a problem because, secure hashes used for addressing content in the Arweave are long strings of seemingly random characters that cannot be shared quickly without needing to resort to pre-established digital communications channels. Anonymity remains as a problem because data is fetched directly from Arweave gateways who can log users IP addresses and the content they are fetching.

Silo attempts to resolve the addressing problem by deploying a naming scheme which can transform a content address and a decryption key into a human readable name, like TapDanceFish.17. To solve the anonymity problem Silo leverages Lokinet as an anonymous network layer to upload and retrieve data from the Arweave. Thirdly Silo encrypts all data before it reaches the Arweave to ensure content can only be read by users with the relevant decryption keys.

*CTO - Oxen Project : kee@oxen.io
†Founder - Arweave : sam@arweave.org
‡Director - OPTF : simon@oxen.io

2 Content Addressing

Typically in decentralised storage systems like IPFS or Arweave content is addressed by a hash. This allows a user to discover content by querying providers for content stored by hash and also allows that content to be verified once downloaded.

Silo keeps this hash based content addressing but adds an element of human readability so that names can be more easily shared. To achieve this users choose a "Silo Address", which consists of a word or collection of words, for example, TapDanceFish and a number, for example 17.

2.1 Silo ID

If we apply a hashing function to these words 2^N times where N is the number chosen by the user. The resulting 32 byte output can then split into two halves, the first half represents the Silo-ID and the second half represents the symmetric encryption key (Silo Key)

Once the Silo ID is obtained the Arweave software makes a blockchain transaction on the Arweave blockchain which consists of the encrypted content and the Silo ID tag, which ensures the transaction is indexed and searchable by its tag once included in the blockchain.

2.2 Silo Key

As previously mentioned the second half of the 32 bytes derived from the mnemonic address is used to derive the Silo key. The Silo key is derived by passing those 16 bytes into a password based key derivation function (PBKDF2)[3]. This expands the 16 bytes into an 256 bit AES key for which data can be encrypted and decrypted with. This encryption process ensures that data stored on the public blockchain is indecipherable to anyone that does not have the encryption key.

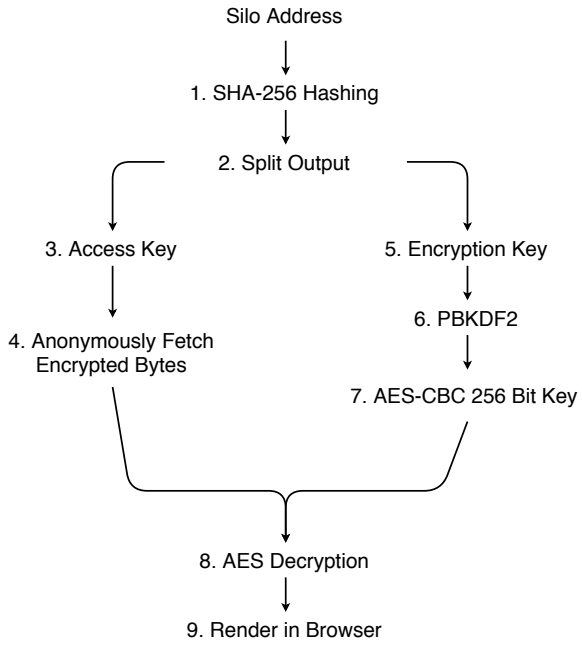

Figure 1: *How a human-readable Silo address is split and reconstituted to find and decrypt data stored on the Arweave*

2.3 Exponential hashing

Transforming the Silo Address into the Silo ID and Silo Key requires several rounds of hashing as indicated by the user prefixed number.

The reason for this hashing is to ensure that rainbow table attacks are difficult to perform at scale, for example if only an single English word was chosen for the mnemonic address, then an attacker would be able to iterate over the a list of the most common English words (about 30,000 words) and calculate the Silo ID and key for each piece of data encrypted under those addresses[4].

To strengthen Silo the scheme requires that the user suffix a number to their words, this number dictates the amount of hash iterations a user must complete before they reach the final Silo ID and Silo Key.

For example, TapDanceFish.17 will require "TapDanceFish" to be hashed 2^{17} (131,072 iterations) to achieve the correct output. This computation must be done in a serial manner, as the outcome of each hashing operation is dependent on the outcome of the last hashing operation. Based on the total words in the Oxford Dictionary there can be a total possible number of three word permutations [5].

$$\textbf{Permutations}, n^r = 171,476^3$$

$$\textbf{Permutations} = 5,042,083,489,338,176$$

by adding a number to each word combination in a range from 1 - 20, (with larger numbers the hashing becomes impractical for users wanting to access data) this increases the possible permutations and increases the time it would take to exhaustively pre-compute Silo IDs.

If users are properly educated on how to use this scheme it offers significant protection against rainbow table based attacks, while not significantly impacting the ability for users to quickly share the location and encryption keys for content.

3 Anonymous access

The Silo construct allows users to anonymously address and encrypt content, however it does not address how that content should accessed or uploaded. Typically Arweave users connect directly to a gateway node to access or upload content to the network. This allows the Arweave gateway to keep a log of the IP address of the connecting user and a log of content retrieved or uploaded to the Arweave.

3.1 Lokinet

Instead of accessing content directly Silo uses an onion routing network called Lokinet to obfuscate the link between a device's IP address and the Arweave gateway node.

Functionally this is achieved by running an Arweave gateway node as a SNApp (Service Node Application). SNApps in Lokinet are similar to hidden services in Tor. To connect to a SNApp the client first creates a path connecting three randomly selected Service Nodes, using this path the client will lookup the SNApp's .loki address, if that SNApp is reachable the lookup will return an introset, which presents the client with a number of pre-established endpoints, endpoints being the final hops in each SNApps paths. The client will then construct a new path which ends on one of the SNApps endpoints. From this point the client can begin sending and receiving packets from the SNApp.

In total there should be 8 hops, with data being successively encrypted for each hop, Because both the user and the SNApp create paths which end on the same endpoints, clients and SNApps maintain their anonymity, ensuring the Arweave gateway IP address is unknown and the clients IP address is unknown.

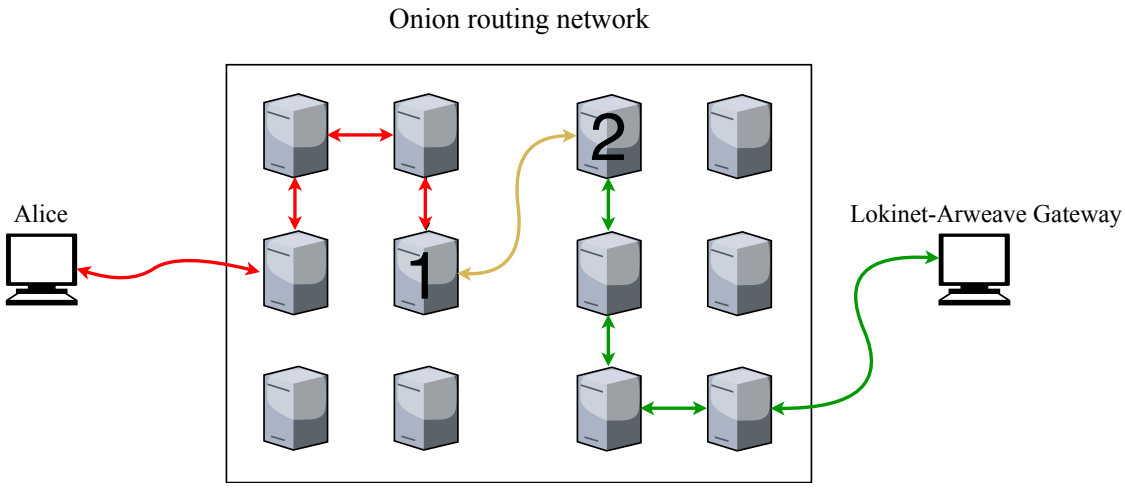

Figure 2: *Alice creates a path to the Lokinet-Arweave gateway endpoint(2), using this path Alice sends a request for data related to a Silo ID. The Lokinet-Arweave gateway sends this data to Alice's using the established path*

3.2 Silo Integration

Since users typically access content on the Arweave using a browser extension when users enter a Silo address they use a special format which signals the browser extension that it needs to resolve the name using the Silo process.

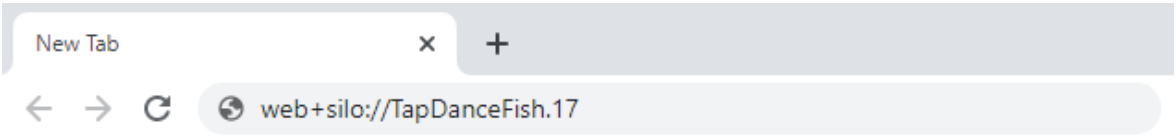

Figure 3: *Example of a Silo Address typed into a URL bar*

When the browser extension sees this format of address it calculates the Silo ID and Silo Key of the address.

Pre-configured in the client is a Lokinet Arweave gateway. The browser extension will attempt to do a DNS lookup for this Lokinet Arweave gateway using the .loki address. If Lokinet is running locally it will catch this DNS lookup and automatically create a path to the Lokinet Arweave gateway. Once this path is established the Silo ID is sent to the gateway and the relevant content is downloaded, once downloaded the client decrypts this data using the Silo Key derived earlier.

4 Future Work

4.1 Token Privacy

Silo focuses most of its efforts on protecting the privacy of users accessing data on the Arweave. Subsequently, it leaves some open questions about the privacy of the user uploading the data. Although users can upload data to the Arweave using a Lokinet gateway to protect exposure of their IP address, users cannot prevent the metadata produced by use of the Arweave token, since the Arweave blockchain is fully transparent. Although transparent blockchains can provide pseudonymity, much research shows that users can be tracked by interaction with centralised parties like exchanges [6, 7].

To limit metadata exposed to third-parties through the use of Arweave tokens there could be a bridge which allows users to swap between Oxen and Arweave tokens. Since Oxen is based on the cryptonote protocol and preserves anonymity during transactions this would provide a way to a user to source Arweave tokens where the origin of those tokens is unknown to a third party, as long as the bridge does not require metadata about the user swapping tokens.

4.2 Incentivization

Currently the Lokinet Arweave bridge is being run by the OPTF (Oxen Foundation). This presents a point of centralisation that could cause issues the OPTF is forced to discontinue service. Work should be done to investigate how we could better incentivize the operation of Lokinet Arweave gateways so as to prevent access issues if gateways are unavailable

5 Conclusion

Silo combines immutable storage with an onion routing network to provide access to archives of information privately and securely. Information access cannot be revoked from any user as gateway IP addresses are not discoverable to third parties, and the storage of data on the Arweave is immutable.

This means that once a piece of content is uploaded it is extremely difficult to censor and extremely easy to disseminate to other users by sharing a few words and a number. Silo welcomes a new era of data availability where information is free, anonymous and tamper-resistant.

References

[1] *History of cryptography*, https://en.wikipedia.org/wiki/History_of_cryptography.

[2] *Arweave: A Protocol for Economically Sustainable Information Permanence*, https://www.arweave.org/yellow-paper.pdf.

[3] *PKCS 5: Password-Based Cryptography Specification Version 2.0* (September 2000), https://tools.ietf.org/html/rfc2898.

[4] *Summary of results - Test your vocab* (May 10, 2013), http://testyourvocab.com/blog/2013-05-10-Summary-of-results.

[5] *Oxford English Dictionary, Second Edition, Volume 1*, 1989.

[6] *Deanonymizing Tor hidden service users through Bitcoin transactions analysis* (July 2019), https://arxiv.org/pdf/1801.07501.pdf.

[7] *An Analysis of Anonymity in the Bitcoin System* (2011), https://arxiv.org/pdf/1107.4524.pdf.