

University of Konstanz
Department of Computer and Information Science



Master Thesis

Building Decentralized Applications with Data Ownership

in fulfillment of the requirements to achieve the degree of
Master of Science (M.Sc.)

Harsh Kedia

Matriculation Number :: 01/752437

E-Mail :: <harsh>.<kedia>@uni-konstanz.de

Field of Study :: Information Engineering
Focus :: Applied Computer Science
Topic :: Distributed Systems

First Assessor :: Prof. Dr. M. Waldvogel
Second Assessor ::
Advisor :: Prof. Dr. M. Waldvogel

Any dedications or other fancy stuff???

Abstract

Abstract

Contents

Abstract	i
List of Figures	iii
List of Tables	iv
1 Introduction	1
2 Background	2
2.1 Pretty Good Privacy (PGP)	2
2.2 Public Key Infrastructure (PKI)	2
2.3 Distributed Hash Table (DHT)	3
2.4 Blockchain	3
3 Concepts & Design	4
3.1 Introduction	4
3.1.1 Centralized	5
3.1.2 Distributed	5
3.1.3 Decentralized	5
3.2 Enabling Technologies	5
3.3 Application Concepts	6
3.4 Data	6
3.4.1 Storing Data Directly in a Blockchain	7
3.4.2 Storing Data in a Distributed Hash Table	7
3.4.3 Storing Data in a Cloud in Encrypted Containers	8
3.5 Identity	10
3.5.1 Decentralized Public Key Infrastructure (DPKI)	11
3.5.2 Decentralized Identifiers (DIDs)	12
3.5.3 Related Work	13
3.6 Value	14
3.7 Computing	16
3.7.1 Related Work	17
3.8 Bandwidth	18
3.8.1 Related Work	18
3.9 Application Design	20
3.9.1 Decentralized Application Architecture	20
3.10 Summary	21
4 Building a Decentralized File Sharing Application	22
4.1 Introduction	22

5	Results	23
6	Discussion	24
7	Conclusion	25
A	Acknowledgements	26
	References	29

List of Figures

3.1	The three way of modeling web applications	4
3.2	How traditional web applications work	6
3.3	IPFS Stack	8
3.4	How users interact with Gaia storage ¹	9
3.5	The token ensures that a write is authorized ²	9
3.6	Overview of Gaia and steps for looking up data. ³	10
3.7	Zooko's Triangle ⁴	11
3.8	The URN Specification	12
3.9	The DID Specification	13
3.10	Blockstack Authentication Flow ⁵	13
3.11	Bitcoin as a state transition system ⁶	14
3.12	Transactions in the Bitcoin Network ⁷	15
3.13	Cloud computing service models ⁸	16
3.14	Overview of Golem Network Architecture ⁹	17
3.15	The Internet ¹⁰	18
3.16	IPv6 Segment Routing Header ¹¹	19
3.17	The Althea Network ¹²	20
3.18	Creating a Decentralized Identity	20
3.19	Decentralized Application Architecture	21

List of Tables

Chapter 1

Introduction

Humans have evolved over thousands of years building systems which deals with land ownership and property rights. With the advent of Internet our lives has become more and more digital, but we have no experience in managing data ownership. It's clear that data is becoming the new currency in today's digital economy. Big tech companies understood this a long time ago and therefore offered their services free of charge in exchange of our data which they then used to generate profits, control our perception about how we see the world and also tamper with public affairs like the election. There's clearly a need to define data ownership and build systems which enable users to own their data.

With data ownership comes the question of digital identity. How can we identify ourselves over the internet? With username and passwords we can uniquely identify ourselves when using a service, but then we have to create an identity for each service we want to use. It has another drawback, i.e. our passwords are stored on a central server which is prone to hacking. There exists systems like *Google Sign-in* or *Facebook Connect* which allows us to carry our identity across multiple services but then again this identity is not owned by the user but by Google or Facebook. Therefore, there is a need for a self-sovereign identity which is owned by the user and can be verified independently by anyone.

To define a model for Data Ownership, lets looks at Land Ownership. In a land ownership model, at any given point in time, a property has a fixed Geo-location while the owner can be anywhere in the Geo-space. Conversely, In a data ownership model, at any given point in time, a user has a fixed identity while the data can be anywhere on the internet.

Blockchain along with Public key cryptography allows us to build a Decentralized Public Key Infrastructure (DPKI) thereby empowering users to create self-sovereign identity. Combining self-sovereign identity with encrypted storage enable us to build systems where users own their identity as well as their data.

Explaining contents of each chapter.

Explores the emerging protocols which enable a decentralize web.

Chapter 2

Background

2.1 Pretty Good Privacy (PGP)

PGP¹ is a encryption program which uses public-key cryptography[1] to provide cryptographic privacy and authentication for data communication. It can be also used to sign messages such that the receiver can verify both the identity of the sender and integrity of the message.

It is built upon a Distributed Web of Trust in which a user's trustworthiness is established by others who can vouch through a digital signature for that user's identity[2].

There are a number of inherent weaknesses which prevented the widespread adoption of PGP. These include the following[2]:

- Trust relationships are built on a subjective honor system.
- Only first degree relationships can be fully trusted.
- Levels of trust are difficult to quantify with actual values.
- Issues with the Web of Trust itself (Certification of Endorsement).

2.2 Public Key Infrastructure (PKI)

PKI is a system for creation, storage and distribution of digital certificates which can be used to verify ownership of a public key[3]. In today's Internet, third parties such as DNS registrars, ICANN, X.509 Certificate Authorities (CAs), and social media companies are responsible for the creation and management of online identities. Thus our online identities lie in the control of third-parties and are borrowed or rented rather than owned. This results in severe usability and security challenges[4].

There is a possible alternate approach called *decentralized public key infrastructure (DPKI)*, which returns control of online identities to the entities they belong to. By doing so, DPKI addresses many usability and security challenges that plague traditional public key infrastructure (PKI)[4].

¹https://en.wikipedia.org/wiki/Pretty_Good_Privacy

2.3 Distributed Hash Table (DHT)

Distributed Hash Tables (DHTs)[5] provides a way of routing data in peer-to-peer systems. It works by enabling a lookup service similar to hash tables where (key, value) pairs are distributed across the network such that any peer can retrieve the value associated with a given key efficiently. DHTs forms a base infrastructure on top of which others decentralized applications are built. Kademlia[6] is a popular DHT which provides an efficient lookup across massive networks. S/Kademlia[7] is another DHT which extends Kademlia to protect nodes against malicious attacks.

2.4 Blockchain

The current Internet Protocol stack consists of four layers: the *Link Layer* puts data onto a wire; the *Internet Layer* routes the data; the *Transport Layer* persists the data; and the *Application Layer* provides data abstraction and delivers it to the end user in the form of applications. All four layers work seamlessly for exchanging of data, but not value. Bitcoin[8] and other cryptocurrencies help define the fifth Internet Protocol layer which enables the exchange of value as fast and efficiently as data[9].

Exchanging value across the Internet presents two challenges. First, every participant in the network must agree upon a shared state and Second, the asset being exchanged should have a clearly defined owner. These challenges are commonly referred as the *Byzantine General's* problem[10] and *Double-spending* problem[11] respectively. Blockchain, the technology underlying Bitcoin and most cryptocurrencies solved the above problems by means of decentralized consensus².

At a higher level, blockchains are append-only, totally-ordered, replicated logs of transactions[12]. A transaction is a signed statement that transfers the ownership of an asset from one cryptographic keypair to another. Peers³ in the network, append new transactions by packaging them into a block and then executing a leader election protocol which determines who gets to append the next block[13]. This election protocol is determined by the underlying consensus algorithm of the blockchain. Each block contains the cryptographic hash of the previous block along with some transactional data.

²[https://en.wikipedia.org/wiki/Consensus_\(computer_science\)](https://en.wikipedia.org/wiki/Consensus_(computer_science))

³A Node having the full copy of the blockchain.

Chapter 3

Concepts & Design

3.1 Introduction

An Application software or *app* is a computer program designed to perform a specific set of tasks or actions for the end user. There are countless number of applications in use today and the majority of them are web applications following a centralized client-server model[9].

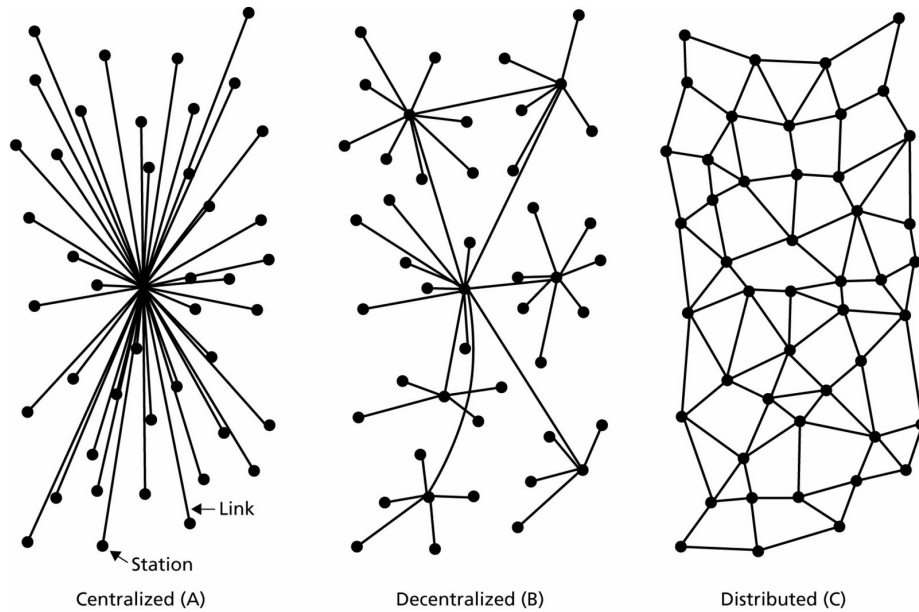


Figure 3.1: The three way of modeling web applications

Figure 3.1 shows a visual representation of three different ways of modeling web applications[14]. Here, *Centralized* and *Decentralized* refers to level of control, while *Distributed* refers to differences of location. Both centralized and decentralized systems can be distributed as well.

3.1.1 Centralized

It's currently the widespread way of building software applications. In this model a central server control the flow of information and governs the operation of individual units. Since the control is centralized, these types of systems suffer from single point of failure risk.

3.1.2 Distributed

In a Distributed model, the control still resides with a central server, however, the computation is spread across multiple nodes or servers.

3.1.3 Decentralized

In a Decentralized model, there is no central point of control as it's spread across all the servers running the application. Applications built using this model don't have a single point of failure and are inherently fault tolerant.

3.2 Enabling Technologies

The document *Information Management: A Proposal*^[15] written by *Sir Tim Berners-Lee*¹ conceived the ideas for what would become the WorldWideWeb. It's main goal was to enable information exchange between computers in an accessible way at CERN².

HTML³, URI⁴ and HTTP⁵ were the fundamental technologies that defined the foundation of the Web. HTTP connected every computer on the planet with a common protocol. The HTTP protocol guidelines defined a set of trusted servers that translated a web address into a server address. Furthermore, HTTPS⁶ added another layer of trusted servers and certificate authorities. People would host personal servers for others to connect to, and everyone owned their data^[9]. As the Web evolved, applications servers⁷ became the common way of interactive with the Web and the centralized model of data ownership as we know it today was born^[9]. It was conceptually and programmatically easier to maintain an application server and profit from user's data that utilize it.

Blockchain is the primary technology that enables the creation of applications with a decentralized model of data ownership. It puts the users of an application in control of their data thereby enabling a more open Web, as it was originally intended⁸.

The blockchain helped solve the Byzantine Generals Problem^[10]. This problem describes a situation where all participating nodes in a distributed network must agree upon every message that is being transmitted between nodes, but where some of the nodes are corrupt and disseminating false information or

¹<https://www.w3.org/People/Berners-Lee/>

²<https://home.cern/>

³<https://developer.mozilla.org/en-US/docs/Web/HTML>

⁴<https://tools.ietf.org/html/rfc3986>

⁵<https://tools.ietf.org/html/rfc2616>

⁶<https://tools.ietf.org/html/rfc2818>

⁷https://en.wikipedia.org/wiki/Application_server

⁸<https://webfoundation.org/about/vision/history-of-the-web/>

are unreliable. This agreement is called as **consensus**. With Bitcoin[8], decentralized consensus became possible. Agreement is achieved in the Bitcoin network by way of *proof-of-work*⁹ consensus mechanism which is resistant to Sybil Attack[16]. Proof-of work is both computationally and energy expensive; other consensus mechanism such as *proof-of-stake*¹⁰ relies on stake in the system instead of computational power.

3.3 Application Concepts

There are five concepts in a web application that have traditionally been implemented in way that puts control with a centralized entity: data, identity, value, computing and bandwidth[9]. Each of these require trust in a 3rd party - a trust which can be betrayed. Recent advancements in distributed-system technology can put users in control of these things. Below sections describes each concept in detail and shows how one can build applications in a way such that centralized control is not required.

3.4 Data

Data is the most important concept in any web application. First, let's look at how traditional web applications interact with data. Whenever, a user logs into an application, the application connects to a remote server and sends the authentication details. These details lets the server know which user is interacting with the application. Once authenticated, the user data is fetched from the remote storage and displayed to the user. All complex computations and data storage occurs on dedicated servers maintained in the cloud.

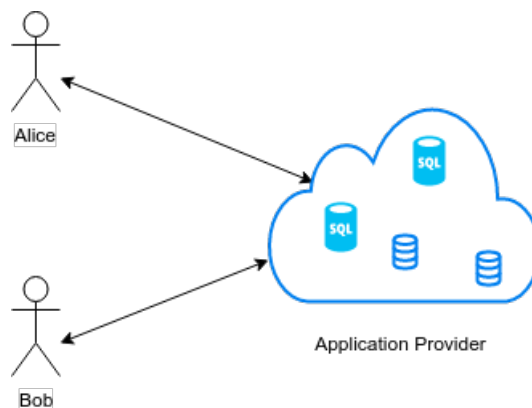


Figure 3.2: How traditional web applications work

Figure 3.2 shows how users interact with a traditional web application. Whenever, Alice wants to interact with Bob, she sends a message to the application server which then delivers the message to Bob. There is no direct path between Alice and Bob. This results in centralization where the provider acts

⁹https://en.bitcoin.it/wiki/Proof_of_work

¹⁰https://en.bitcoin.it/wiki/Proof_of_Stake

as an intermediary between Alice and Bob's interaction; effectively governing how data is stored and shared among them.

This model of application interaction requires that we trust the providers with our data and hope that they won't misuse or sell our data without our permission. Since revelations by Edward Snowden[17], we now know that trust can, has and will be broken as long as we entrust our data to a central entity[9]. Centralized stores of data also serve as a tool for surveillance, allowing big entities to monitor our internet behaviour without our knowledge. Cloud providers, despite having a distributed backend, are centrally owned.

Additionally, as we move from a labor-based economy to an information-based economy, data will become the primary form of value. Therefore, it's important that we not only possess our data, but own it as the world evolves. An ideal solution that enables this should provide a way of storing data in a decentralized way that is robust and as trustless as possible[9].

3.4.1 Storing Data Directly in a Blockchain

This method does solve the decentralization of data as everyone who has a copy of the Blockchain is storing the data but cannot alter it. The data can be encrypted such that it can only be accessed by someone having the private key. However, Blockchains were not meant for storing large amounts of data. It was designed for storing simple transactional logs, as is evident from looking at the Bitcoin Blockchain where individual records are in bytes¹¹ and therefore cannot hold much data. Moreover, the Blockchain architecture implies that each node in the network must store a complete copy of the Blockchain. Therefore, storing a significant amount of data on a Blockchain is both expensive and impractical.

3.4.2 Storing Data in a Distributed Hash Table

DHTs ensure data resiliency by enabling easy distribution and indexing of data. Early peer-to-peer (P2P) filesharing applications like KaZaA[18], Napster[19] and Gnutella[20] used their own versions of DHTs with a varying degree of decentralization. Some had centralized trackers to monitor the movement of data, while some had central sources from which all data must pass, leaving them with a single point of failure[9].

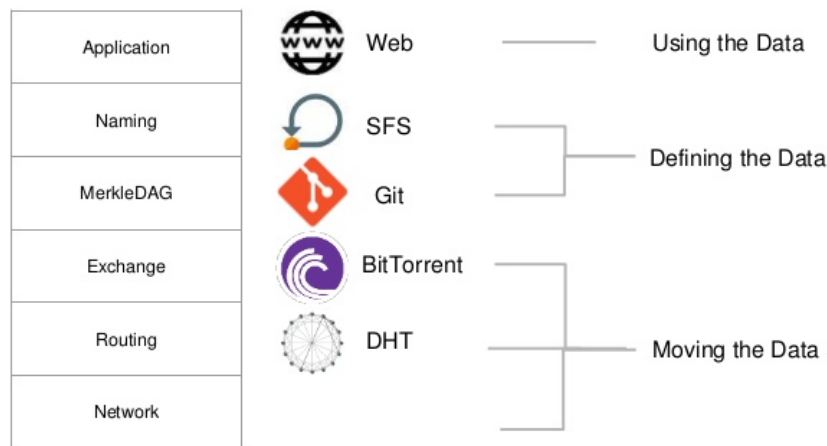
BitTorrent[21] was the first protocol to popularize DHT for sharing of large files. As of 2013, BitTorrent has 1527 million concurrent users at any given time[22]. The BitTorrent Mainline DHT serves as a decentralized data store with centralized trackers to monitor the network. It uses a tit-for-tat strategy between seeders and leechers to maximize the bandwidth of data distribution. This makes the BitTorrent protocol the de facto method for transfer of large datasets over the web. However, using BitTorrent as a data store is not feasible as there are no incentives for nodes to store data leading to *data impermanence*.

Along with the decentralized storage capabilities of DHT and the speed of BitTorrent's protocol, we also want data permanence. Therefore, it's necessary to incentivize the nodes storing the data in some way. Moreover, we also need to ensure that the links to the data don't die, an idea which was first proposed in Project Xanadu[23].

¹¹<https://en.bitcoin.it/wiki/Transaction>

InterPlanetary File System (IPFS)

IPFS[24] is a peer-to-peer file transfer protocol that implements these features, enabling a more permanent, decentralized Web, where links don't die and no single entity controls the data. It achieves this by combining previous peer-to-peer technologies such as DHT, BitTorrent and Git[25]. Data in the IPFS network is modeled as a *merkleDAG*¹², a simple data structure that can be conceptualized as a series of nodes connected to each other[9]. This makes IPFS a *content-addressed*¹³ system allowing efficient data lookup and retrieval as it doesn't rely on a single server to access the data.



12

Figure 3.3: IPFS Stack

Filecoin[26], a *decentralized storage network*, combines IPFS (shown in Figure 3.3¹⁴) with an incentive structure that turns cloud storage into an algorithmic market effectively overcoming the limitations of BitTorrent. This market runs on a blockchain with a native protocol token, which miners earn by providing storage to clients.

3.4.3 Storing Data in a Cloud in Encrypted Containers

IPFS enables a way for decentralized storage such that it overcomes the single point of failure risk, however, the user does not control where the data is being stored. Because of this, once a data is added to the network, and is picked up by other nodes for re-hosting, it cannot be taken back as there is no way to verify information deletion¹⁵.

¹²Similar to a Merkle tree data structure, however, they do not need to be balanced and it's non-leaf nodes can contain some data.

¹³A content-addressed storage is way to store information such that it can retrieved based on its content and not its location.

¹⁴Adopted from: <https://image.slidesharecdn.com/ipfs-171229085327/95/ipfs-12-638.jpg?cb=1514537643>

¹⁵<https://github.com/ipfs/faq/issues/9>

Gaia: User-Controlled Storage

Gaia[27] is a decentralized storage system that enables user-controlled private data lockers. It works by hosting data in one or more existing storage systems of user's choice. Data on Gaia is encrypted and signed by user-controlled cryptographic keys before being uploaded to a provider. Storing data using existing cloud infrastructure ensures high data availability without compromising on application performance. Further, since each data is signed by keys which a user controls, it ensures that they don't need to trust the underlying cloud providers.

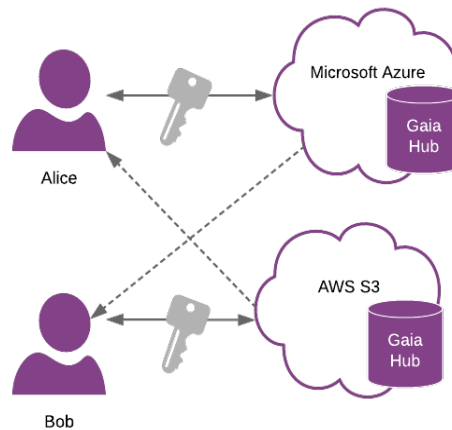


Figure 3.4: How users interact with Gaia storage¹⁶

Writing data to a Gaia hub involves a `POST` request along with a signed *authentication* token. This token is signed by the private key which controls the particular bucket being written to. Separate buckets are used for each application, this ensures that a given private key grants access only to a specific bucket on the Gaia Server.

On GAIA writes

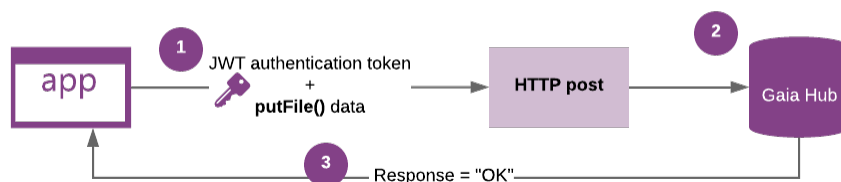


Figure 3.5: The token ensures that a write is authorized¹⁷

¹⁶<https://docs.blockstack.org/storage/overview.html>

Reading data from a user's Gaia hub involves a *zonefile* lookup. This zonefile is a signed JSON object containing the URLs pointing to the user's Gaia data locker. Once verified that the zonefile is signed by user's key, a standard HTTP request is made to fetch the requested data.

Figure 3.6 shows an overview of Gaia. Looking up data for a name works as follows:

1. Lookup the *name* in the Virtualchain to get the *(name, hash)* pair.
2. Lookup the *hash(name)* in Peer Network to get respective zonefile.
3. Get the user's Gaia URL from the zonefile and lookup the URL to connect to storage backend.
4. Fetch the requested data and verify the respective signature or hash.

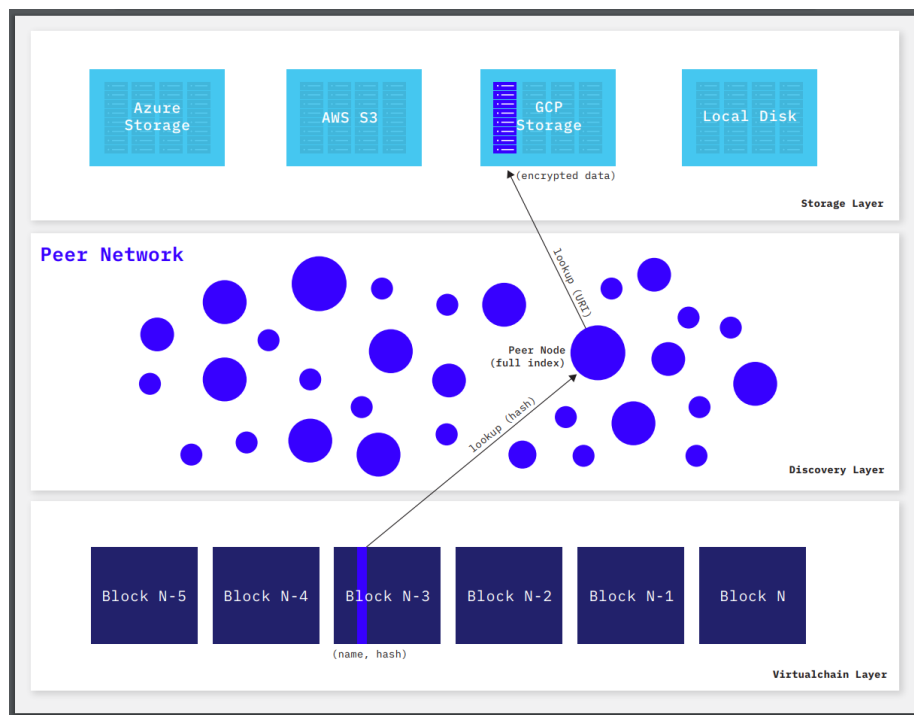


Figure 3.6: Overview of Gaia and steps for looking up data.¹⁸

3.5 Identity

User identities are essential to using internet applications. In order to confirm their identity, users must provide some information. This information depends on the application's authentication process. If an application maintains a user

¹⁷<https://docs.blockstack.org/storage/authentication.html>

¹⁸<https://blockstack.org/whitepaper.pdf>

database, it requires a username and a password and sometimes a second factor. If an application relies on a third-party, like Google¹⁹ or Facebook²⁰ for identity management, it uses the OAuth 2.0 authentication[28] protocol, which identifies a user by generating an assertion from the identity service.

Third party identity providers implementing the OAuth²¹ protocol often have their own version of implementation which leads to identity fragmentation across the web. OpenID[29] is a decentralized identity protocol that allows users to create one identity which can be carried across multiple providers. However, the user still needs to trust one of the service providers with their identity information[9].

Another way of authenticating users with an application is by using digital certificates[30]. These certificates provide a proof of ownership of a public key. Authentication and management of public keys is facilitated using a Public Key Infrastructure (PKI)[31] system. Currently, the most common approaches to PKIs are: Certificate Authorities (CAs) and PGP Web of Trust[32].

A Certificate Authority (CA) acts a trusted third party responsible for management and distribution of digital certificates for a network of users[33]. These trusted third parties introduces central control in a PKI system making them prone to single point of failure risk[34].

In PGP Web of trust, authentication is entirely decentralized; users are able to designate others as trustworthy by signing their public keys. This process generates a digital certificate containing the user's public key and signatures from entities that have deemed him trustworthy. This system does benefit from its distributed nature as there is no central control. However, PGP does not offer identity retention. There is no guarantee of consistency and nothing prevents multiple users from creating public keys for the same identity[33].

3.5.1 Decentralized Public Key Infrastructure (DPKI)

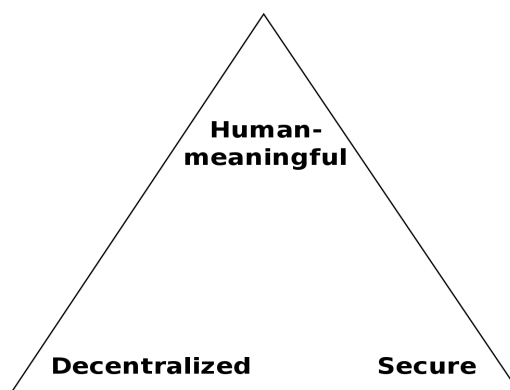


Figure 3.7: Zooko's Triangle²²

¹⁹<https://developers.google.com/identity/>

²⁰<https://developers.facebook.com/docs/facebook-login/>

²¹<https://oauth.net/>

²²https://commons.wikimedia.org/wiki/File:Zooko%27s_Triangle.svg

The foundational precept of a DPKI is that *identities belong to the entities they represent*. This requires designing a *decentralized* infrastructure where every identity is controlled by its *principal owner* and not by some trusted third party[4].

DPKI is essentially a system to give unique names to participants in a network protocol. These names should have three desirable traits (Zooko's triangle) namely: *Human-meaningful*, *Decentralized* and *Secure*. OpenID solved security and human-meaningfulness.

Namecoin²³ was the first working solution that satisfied all three traits of Zooko's triangle by adding decentralization. It was the first fork of Bitcoin[8] protocol designed to act as a decentralized domain name server (DNS) for *.bit* addresses. Namecoin's blockchain essentially could be used as an intermediary between a user and the service requesting user's identity[9]. It allowed a user to *register* a name by creating a blockchain transaction. This transaction included the desired name which got embedded under the */id* namespace.

3.5.2 Decentralized Identifiers (DIDs)

A DID²⁴ is a new type of identifier that is globally unique, resolvable with high availability and cryptographically verifiable. DIDs typically contain a cryptographic key pair which enables the controller of a DID to prove control over it.

The concept of global unique decentralized identifiers is not new; Universally Unique Identifiers²⁵ (UUIDs), also called Globally Unique Identifiers (GUIDs) were first such identifiers developed in the 1980s and formally specified in RFC4122²⁶. Another class of identifiers known as persistent identifiers was standardized as Uniform Resource Names (URNs) by RFC8141²⁷.

However, UUIDs are not globally resolvable and URNs, if resolvable, require a central registry. Further, neither UUIDs or URNs have the ability to cryptographically verify ownership of the identifier. DIDs, on the other hand, fulfill all four requirements: persistence, global resolvability, cryptographically verifiability, and decentralization required for a self-sovereign identity[35].



Figure 3.8: The URN Specification

The DID (Figure 3.9²⁸) specification follows the same pattern as the URN (Figure 3.8²⁹) specification. The key difference is that with DIDs, the names-

²³<https://www.namecoin.org/>

²⁴<https://w3c-ccg.github.io/did-spec/>

²⁵https://en.wikipedia.org/wiki/Universally_unique_identifier

²⁶<https://tools.ietf.org/html/rfc4122>

²⁷<https://tools.ietf.org/html/rfc8141>

²⁸<https://w3c-ccg.github.io/did-primer/did-primer-diagrams/did-format.png>

²⁹<https://w3c-ccg.github.io/did-primer/did-primer-diagrams/urn-format.png>



Figure 3.9: The DID Specification

pace component identifies a *DID method* which defines how DIDs work with a specific blockchain. The *DID method* specifications must define the format and generation of the method-specific identifier. The DID methods can also be developed for identifiers registered in federated or centralized identity management systems, thus creating a interoperability bridge between the worlds of centralized, federated and decentralized identifiers.

3.5.3 Related Work

Blockstack Authentication

Blockstack³⁰ is a decentralized computing network and app ecosystem that uses public-key cryptography and blockchain technology to provide its users with a DPKI system thereby making it possible to have a universal username that works across applications without the need for any passwords.

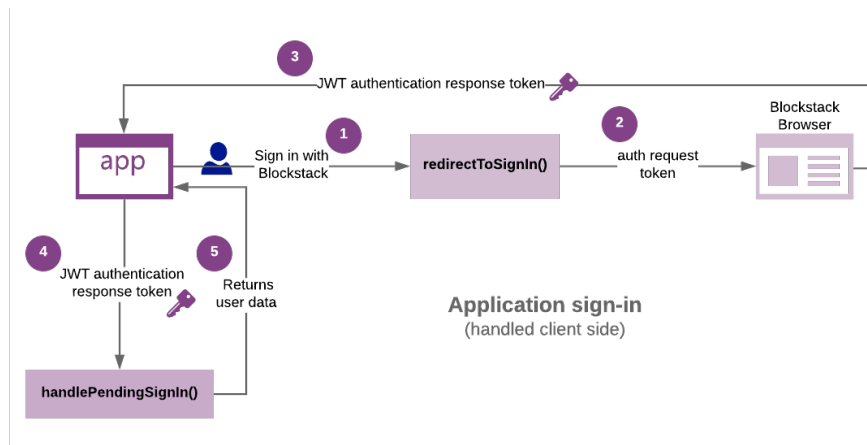


Figure 3.10: Blockstack Authentication Flow³¹

Blockstack Auth³² is a token-based authentication³³ system. When a user signs into an application, an `authRequest` token is sent to the Blockstack Browser. Once the sign-in is approved, the Blockstack Browser responds with

³⁰<https://blockstack.org/>

³¹<https://docs.blockstack.org/storage/images/app-sign-in.png>

³²https://docs.blockstack.org/develop/overview_auth.html

³³https://www.w3.org/2001/sw/Europe/events/foaf-galway/papers/fp/token_based_authentication/

an `authResponse` token. This response provides the application with enough information to generate and store authentication data. Since all usernames are registered on the blockchain, every application have an up-to-date view of all usernames and corresponding public keys. This eliminates the need for a server-side identity provider.

Figure 3.10 visualizes the authentication flow for an application using Blockstack Auth. When a user chooses to **Sign in with Blockstack**, it calls the `redirectToSignIn()` method which sends an `authRequest` token to the Blockstack Browser³⁴. Upon receiving the request, the Blockstack Browser presents the user with a choice of IDs to use to sign in, as well as a list of permissions the application needs. Upon selecting an ID, the Blockstack Browser redirects back to the application with a `authResponse` token containing three pieces of information³⁵:

- The user's username (or the hash of the public key if no username is set).
- An *application-specific* private key for encrypting and signing user's data. This key is deterministically generated from the user's master private key, the ID used to sign in, and the application's HTTP Origin.
- The URLs to the user private data locker (Gaia Hub³⁶).

Above pieces of information instructs the application on how to find and store a user's data on their behalf.

3.6 Value

Value help us to quantify quantities of an asset in order to make them measurable³⁷. It can mean money, intellectual property or securities. In traditional internet applications trust is required in a third-party (a bank or a financial institution) when transferring value between two entities. While this system works well enough for most transactions, it still suffers from the inherent weaknesses of the trust based model[8].

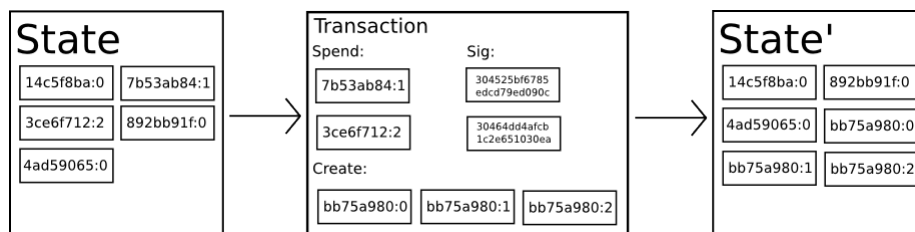


Figure 3.11: Bitcoin as a state transition system³⁸

³⁴<https://github.com/blockstack/blockstack-browser>

³⁵<https://blockstack.org/whitepaper.pdf>

³⁶<https://github.com/blockstack/gaia/tree/master/hub>

³⁷<https://medium.com/swlh/what-is-the-internet-of-values-3f14b5d35a90>

³⁸<https://raw.githubusercontent.com/ethereumbuilders/GitBook/master/en/vitalik-diagrams/statetransition.png>

Bitcoin[8] was the first peer-to-peer electronic payment system that's based on cryptographic proof instead of trust, allowing any two entities to transfer value over the internet without the need for any trusted third-party. A transaction in the Bitcoin network involves digitally signing a hash of the previous transaction and the public key of the recipient and adding these to the end of the transaction. Multiple transactions are combined to form a block and each block is linked with the previous block by it's hash and a *nonce* generated by a proof-of-work[36] algorithm. This chain of blocks forms a distributed data structure called the *blockchain*. It can be thought of as a state transition system, where a *state* consists of the ownership status of all existing bitcoins and a *state transition function* that takes a state and a transaction and outputs a new state[37].

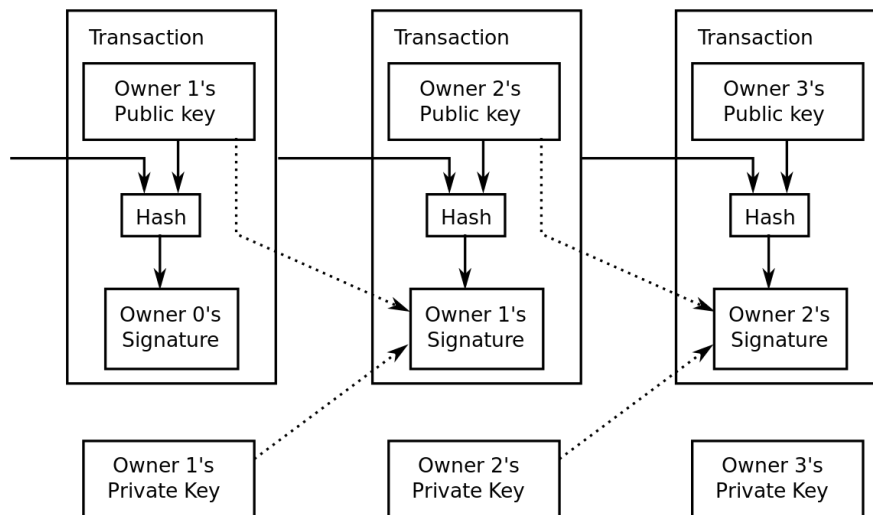


Figure 3.12: Transactions in the Bitcoin Network³⁹

Blockchain, the underlying technology of Bitcoin can also be applied to build systems which can represent assets in the physical world. For example, *Secure property titles with owner authority*⁴⁰, a document written by Nick Szabo⁴¹ describes how a blockchain-based system can be used to build a land registry. Ethereum[37], an alternative protocol with a built-in Turing-complete programming language enables such systems by allowing anyone to write *smart contracts*[38] where they can create their own arbitrary rules of ownership⁴².

Ethereum also enables the creation of *tokens* which allows us to represent something specific in an application, for example, economic value, a dividend, a stake or even a voting right⁴³. The two most common tokens standards in

³⁹https://cdn-images-1.medium.com/max/1200/1*d9AIfrD_DOLtXfKVnadN6w.png

⁴⁰<https://nakamotoinstitute.org/secure-property-titles/>

⁴¹https://en.wikipedia.org/wiki/Nick_Szabo

⁴²<https://github.com/ethereum/wiki/wiki/White-Paper>

⁴³<https://education.district0x.io/general-topics/understanding-ethereum/the-role-of-tokens/>

the Ethereum ecosystem are *ERC-20*⁴⁴: A class of identical tokens, and *ERC-721*⁴⁵: A class of unique tokens. These token standards can be used to tokenize ownership of any arbitrary data or even real world assets thereby enabling an *Internet of Value*⁴⁶.

3.7 Computing

Any web application needs computing resources to work. Traditional internet relied on a *client-server* model where server is usually a more powerful, better connected machine and thus provided most of the computing resources. This, however, also made server as a single bottleneck for performance and reliability[39]. As internet connectivity improved and web applications became more complex, the demand for computing resources increased.

Virtualization⁴⁷, a concept pioneered by the VM[40] operating system led to the emergence of *cloud computing*. Cloud computing satisfied the growing demand of web applications by providing an on-demand delivery of computing resources over the internet. This greatly defined the current internet architecture where applications relies on cloud services providers who provides computing resources based on different service models (Figure 3.13).

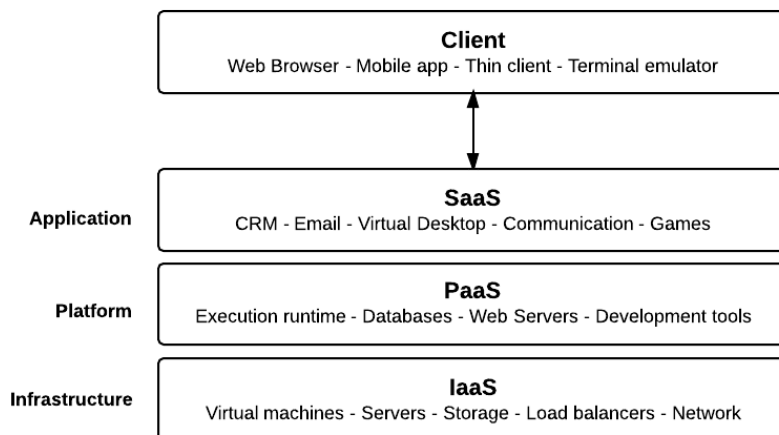


Figure 3.13: Cloud computing service models⁴⁸

Cloud computing, however, poses privacy concerns as it requires entrusting data to information systems managed by third-party cloud service providers who has access to all of user's data at any time and could accidentally or deliberately use it for unauthorized purposes[41]. Because data from multiple entities can be stored on a large cloud servers, it poses security concerns as well. Hackers

⁴⁴<https://github.com/ethereum/EIPs/blob/master/EIPS/eip-20.md>

⁴⁵<https://github.com/ethereum/EIPs/blob/master/EIPS/eip-721.md>

⁴⁶<https://hackernoon.com/a-vision-of-the-internet-of-value-ad187abf5826>

⁴⁷<http://www.kernelthread.com/publications/virtualization/>

⁴⁸<https://static.bluepiit.com/blog/wp-content/uploads/sites/2/2015/12/types-of-cloud-computing-models.png>

can theoretically gain control of huge amounts of data through a single attack - a process called *hyperjacking*⁴⁹.

Blockchain can be used to create a decentralized global market where infrastructure providers, application owners and individual users can come together to trade their unused computing resources. This will also allow us to harness the computing resources of millions of personal computers and workstations connected to the public internet which have hundreds of compute cycles going unused every second[39]. Coupled with end-to-end encryption, this network of connected computers can overcome the security and privacy concerns posed by Cloud computing.

3.7.1 Related Work

Golem Network

Golem⁵⁰ aims to create a decentralized supercomputer by creating a global market for computing power. It connects computers in a peer-to-peer network, enabling every participant to share their unused resources. It uses the Ethereum blockchain for accounting, which enables direct payments between providers, application owners and individual users. Finally, it employs an *Application Registry* which is an Ethereum smart contract, to which anyone can publish their application.

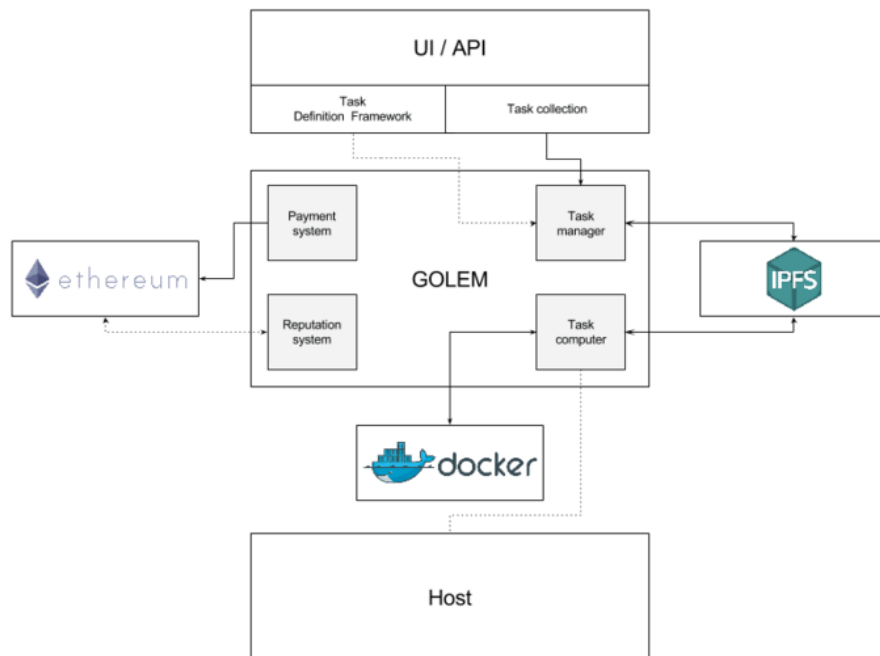


Figure 3.14: Overview of Golem Network Architecture⁵¹

⁴⁹<https://www.securityweek.com/deep-dive-hyperjacking>

⁵⁰<https://golem.network/crowdfunding/Golemwhitepaper.pdf>

3.8 Bandwidth

ISPs act as gateways between end users and the internet (Figure 3.15). They solve the "last mile" problem by connecting users to the high-speed internet and acting as a central hub for communication. However, these central gateways are also central points of failure. Governments can shut them down or ask them to blacklist certain IPs[9].

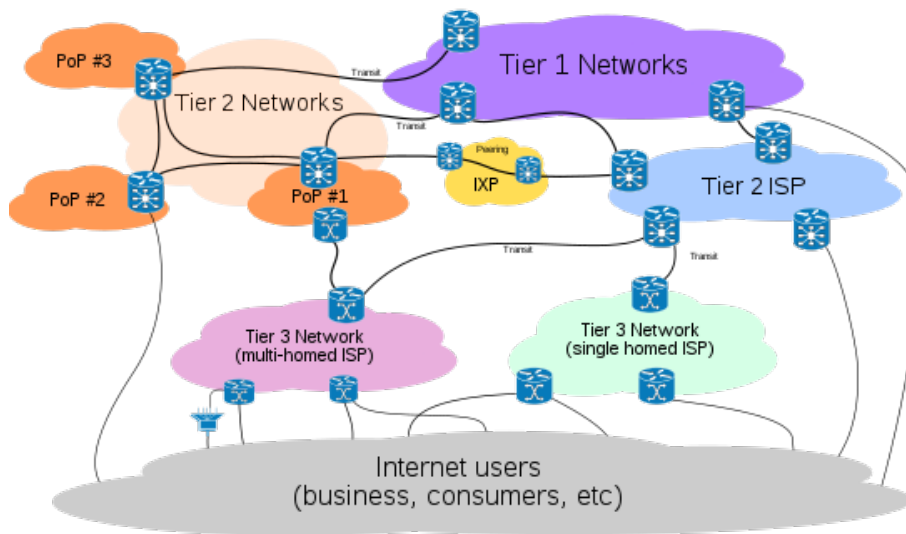


Figure 3.15: The Internet⁵²

Mesh networks are the decentralized version of standards internet where users don't need to go through a central gateway. In a mesh network, nodes connect to as many adjacent nodes for efficient routing of data. However, a decentralized consensus system is required to incentivize all participants to share their bandwidth. Blockchain can be used to serve as an incentive layer effectively creating a decentralized marketplace for bandwidth. Mark Nadal and Dr. Amber Cazzell describes one such protocol in their paper *A Trustless Decentralized Bandwidth Incentive*⁵³.

3.8.1 Related Work

NOIA Network

NOIA⁵⁴ network is a Distributed peer-to-peer content delivery network, governed by the blockchain where nodes are incentivized for providing storage and unused bandwidth to the network.

⁵¹Adapted from https://en.bitcoinwiki.org/wiki/Golem_Worldwide_Supercomputer

⁵²https://commons.wikimedia.org/wiki/File:Internet_Connectivity_Distribution_%26_Core.svg

⁵³<https://web.stanford.edu/~nadal/A-Decentralized-Data-Synchronization-Protocol.pdf>

⁵⁴<https://noia.network/>

⁵⁵<https://docs.noia.network/noia/segment-routing--srv6->

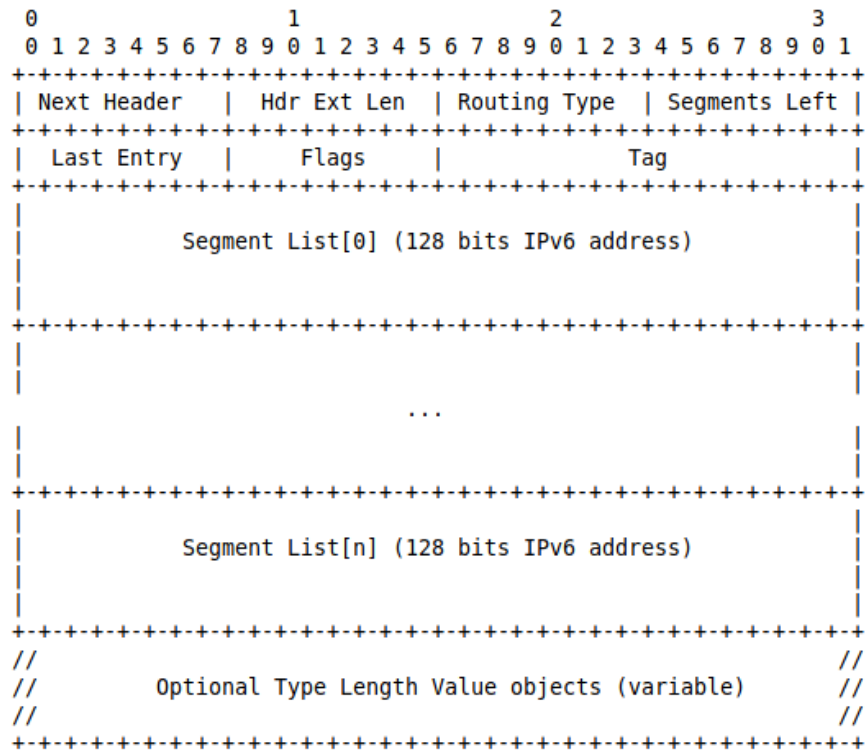


Figure 3.16: IPv6 Segment Routing Header⁵⁵

NOIA enables a *programmable internet*⁵⁶ by leveraging IPv6 and Segment Routing (SR). IPv6 increases the address size from 32 bits to 128 bits. This increase in packet size allows NOIA to encode custom information in packet headers thus providing an identification system for nodes connected to the network. These *custom headers* (Figure 3.16) enables SR by leveraging the source routing paradigm. SR is then used to create a Software Defined Network (SDN) where nodes are identified by their segment identifiers (SIDs). SIDs are registered on the NOIA Ledger to create a network topology of SR enabled access points. A Decentralized Internet Transit Exchange (DITEX) is used for trading of access points as a Transit.

Althea Network

Althea⁵⁷ Network aims to solve the "last mile" problem by connecting a source of internet connectivity to the end users using a mesh network topology where packets are forwarded using the Babel routing protocol[42]. It creates a decentralized ISP where routers pay each other for bandwidth with cryptocurrencies. Governance in Althea network is done using a *proof-of-stake* blockchain built using the Cosmos⁵⁸ protocol.

⁵⁶<https://docs.noia.network/noia/programmable-internet>

⁵⁷<https://althea.net/>

⁵⁸<https://cosmos.network/>

⁵⁹Adapted from: <https://althea.net/whitepaper>

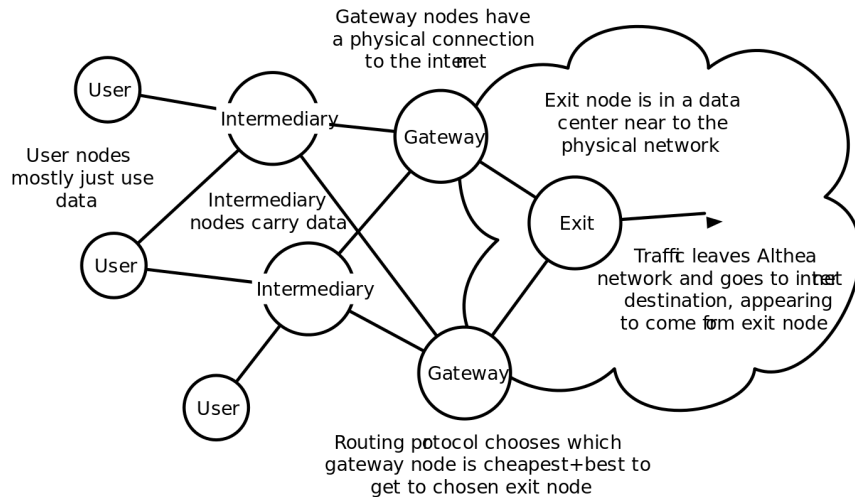


Figure 3.17: The Althea Network⁵⁹

3.9 Application Design

Designing any web application requires understanding of the underlying architecture. This section combines all the application concepts to visualize the underlying architecture which enables decentralized applications with data ownership.

3.9.1 Decentralized Application Architecture

Self-sovereign identity is one of the most important element in decentralized applications with data ownership. The DIDs⁶⁰ specifications enables the creation of a DPKI system on top of any blockchain. A user can create a *decentralized identity* by registering their public key with a desired username by creating a blockchain transaction.

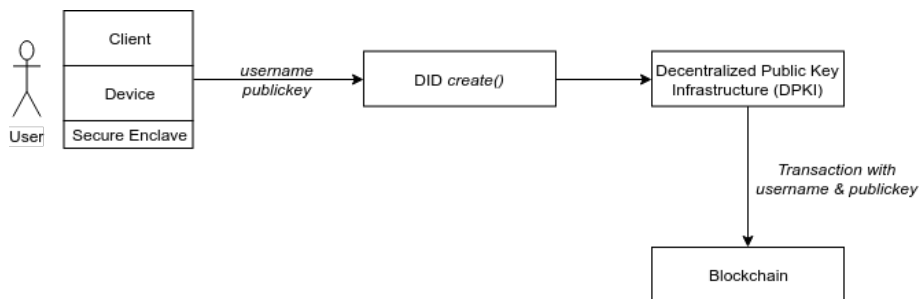


Figure 3.18: Creating a Decentralized Identity

A decentralized identity anchored on a blockchain serves two purposes. First,

⁶⁰<https://w3c-ccg.github.io/did-spec/>

it gives us a cryptographic keypair which can be used by an application for encryption/decryption of data. Second, it serves as a digital wallet for secure exchange of value across the Internet using cryptocurrencies.

Decentralized applications follow a *serverless* architecture where backend services such as database, storage and authentication are provided as APIs that enable client applications to connect directly to these services. The frontend components such as business logic and application code are deployed inside a hosting environment such that the servers that run the code is completely abstracted.

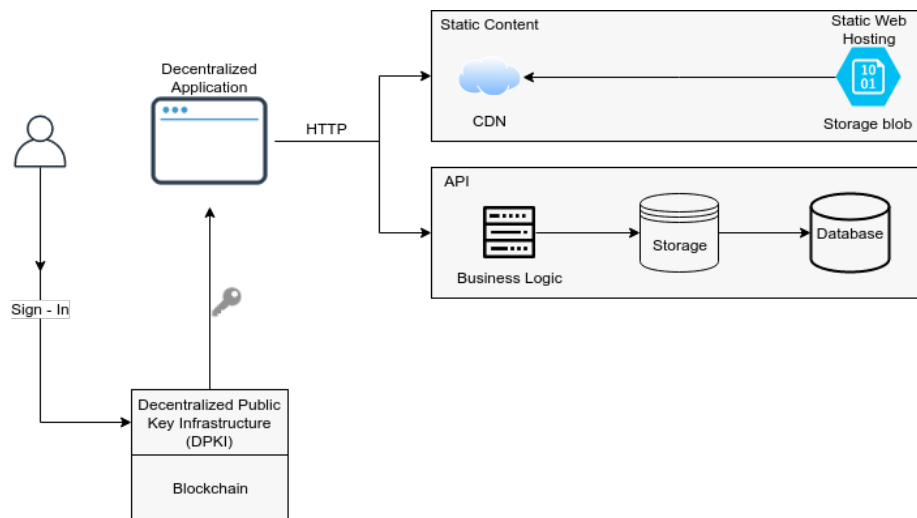


Figure 3.19: Decentralized Application Architecture

Figure 3.19 visualizes the architecture of a decentralized web application. Authentication is handled using a DPKI system. The application connects to various services like storage and database using APIs. The client application encrypts all private data using user's public key before sending it to a decentralized storage network. A database service can be used for easy indexing of data.

3.10 Summary

This chapter outlined the different concepts of a web application and showed how Blockchain enables decentralization for each of the concepts. The key concept which allows us to build decentralized applications with data ownership is *Decentralized Identity*. It enables a Web where we don't have to use username/-passwords anymore and where our data have cryptographic security.

Chapter 4

Building a Decentralized File Sharing Application

4.1 Introduction

Chapter 5

Results

Chapter 6

Discussion

Chapter 7

Conclusion

Appendix A

Acknowledgements

Bibliography

- [1] W. Stallings, *Cryptography and Network Security: Principles and Practice*, ser. The William Stallings books on computer and data communications technology. Prentice Hall, 1999. [Online]. Available: <https://books.google.de/books?id=Dam9zrViJjEC>
- [2] D. Wilson and G. Ateniese, “From pretty good to great: Enhancing pgp using bitcoin and the blockchain,” in *International conference on network and system security*. Springer, 2015, pp. 368–375.
- [3] J. Weise, “Public key infrastructure overview,” *Sun BluePrints OnLine*, August, pp. 1–27, 2001.
- [4] C. Allen, A. Brock, V. Buterin, J. Callas, D. Dorje, C. Lundkvist, P. Kravchenko, J. Nelson, D. Reed, M. Sabadello *et al.*, “Decentralized public key infrastructure. a white paper from rebooting the web of trust,” 2015.
- [5] W. Galuba and S. Girdzijauskas, “Distributed hash table,” *Encyclopedia of database systems*, pp. 903–904, 2009.
- [6] P. Maymounkov and D. Mazieres, “Kademlia: A peer-to-peer information system based on the xor metric,” in *International Workshop on Peer-to-Peer Systems*. Springer, 2002, pp. 53–65.
- [7] I. Baumgart and S. Mies, “S/kademlia: A practicable approach towards secure key-based routing,” in *2007 International Conference on Parallel and Distributed Systems*. IEEE, 2007, pp. 1–8.
- [8] S. Nakamoto *et al.*, “Bitcoin: A peer-to-peer electronic cash system,” 2008.
- [9] S. Raval, *Decentralized applications: harnessing Bitcoin’s blockchain technology*. ” O’Reilly Media, Inc.”, 2016.
- [10] L. Lamport, R. Shostak, and M. Pease, “The byzantine generals problem,” *ACM Transactions on Programming Languages and Systems (TOPLAS)*, vol. 4, no. 3, pp. 382–401, 1982.
- [11] U. W. Chohan, “The double spending problem and cryptocurrencies,” *Available at SSRN 3090174*, 2017.
- [12] J. Bonneau, A. Miller, J. Clark, A. Narayanan, J. A. Kroll, and E. W. Felten, “Research perspectives and challenges for bitcoin and cryptocurrencies (extended version),” *Cryptology ePrint Archive, Report 2015/452*, 2015.

-
- [13] J. Nelson, M. Ali, R. Shea, and M. J. Freedman, “Extending existing blockchains with virtualchain,” in *Workshop on Distributed Cryptocurrencies and Consensus Ledgers*, 2016.
 - [14] P. Baran, “On distributed communications,” 1964.
 - [15] T. J. Berners-Lee, “Information management: A proposal,” Tech. Rep., 1989.
 - [16] J. R. Douceur, “The sybil attack,” in *International workshop on peer-to-peer systems*. Springer, 2002, pp. 251–260.
 - [17] G. Greenwald, *No place to hide: Edward Snowden, the NSA, and the US surveillance state*. Macmillan, 2014.
 - [18] N. S. Good and A. Krekelberg, “Usability and privacy: a study of kazaa p2p file-sharing,” in *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 2003, pp. 137–144.
 - [19] R. Ku and R. Shih, “The creative destruction of copyright: Napster and the new economics of digital technology,” *U. Chi. L. Rev.*, vol. 69, p. 263, 2002.
 - [20] M. Ripeanu, A. Iamnitchi, and I. Foster, “Mapping the gnutella network,” *IEEE Internet Computing*, no. 1, pp. 50–57, 2002.
 - [21] B. Cohen, “The bittorrent protocol specification,” 2008.
 - [22] L. Wang and J. Kangasharju, “Measuring large-scale distributed systems: case of bittorrent mainline dht,” in *IEEE P2P 2013 Proceedings*. IEEE, 2013, pp. 1–10.
 - [23] W. B. Rayward, “Visions of xanadu: Paul otlet (1868–1944) and hypertext,” *Journal of the American Society for information science*, vol. 45, no. 4, pp. 235–250, 1994.
 - [24] J. Benet, “Ipfs-content addressed, versioned, p2p file system,” *arXiv preprint arXiv:1407.3561*, 2014.
 - [25] J. Loeliger and M. McCullough, *Version Control with Git: Powerful tools and techniques for collaborative software development*. ” O’Reilly Media, Inc.”, 2012.
 - [26] J. Benet and N. Greco, “Filecoin: A decentralized storage network,” *Protoc. Labs*, 2018.
 - [27] M. Ali, J. Nelson, R. Shea, and M. J. Freedman, “Blockstack: A global naming and storage system secured by blockchains,” in *2016 {USENIX} Annual Technical Conference ({USENIX}{ATC} 16)*, 2016, pp. 181–194.
 - [28] D. Hardt, “The oauth 2.0 authorization framework,” 2012.
 - [29] D. Recordon and D. Reed, “Openid 2.0: a platform for user-centric identity management,” in *Proceedings of the second ACM workshop on Digital identity management*. ACM, 2006, pp. 11–16.
-

-
- [30] F. A. Tycksen Jr and C. W. Jennings, “Digital certificate,” Feb. 13 2001, uS Patent 6,189,097.
- [31] C. Adams and S. Lloyd, *Understanding PKI: concepts, standards, and deployment considerations*. Addison-Wesley Professional, 2003.
- [32] A. Abdul-Rahman, “The pgp trust model,” in *EDI-Forum: the Journal of Electronic Commerce*, vol. 10, no. 3, 1997, pp. 27–31.
- [33] C. Fromknecht, D. Velicanu, and S. Yakoubov, “A decentralized public key infrastructure with identity retention.” *IACR Cryptology ePrint Archive*, vol. 2014, p. 803, 2014.
- [34] K. Dooley, *Designing Large Scale Lans: Help for Network Designers*. ” O’Reilly Media, Inc.”, 2001.
- [35] S. M. Smith and D. Khovratovich, “Identity system essentials,” *Evemyrn*, Mar, vol. 29, p. 16, 2016.
- [36] A. Back *et al.*, “Hashcash-a denial of service counter-measure,” 2002.
- [37] V. Buterin *et al.*, “Ethereum: A next-generation smart contract and decentralized application platform,” URL <https://github.com/ethereum/wiki/wiki/%5BEnglish%5D-White-Paper>, vol. 7, 2014.
- [38] N. Szabo, “Formalizing and securing relationships on public networks,” *First Monday*, vol. 2, no. 9, 1997.
- [39] J. Crowcroft, T. Moreton, I. Pratt, and A. Twigg, “Peer-to-peer systems and the grid,” *draft, University of Cambridge Computer Laboratory*, 2003.
- [40] R. J. Creasy, “The origin of the vm/370 time-sharing system,” *IBM Journal of Research and Development*, vol. 25, no. 5, pp. 483–490, 1981.
- [41] M. D. Ryan, “Cloud computing privacy concerns on our doorstep,” *Communications of the ACM*, vol. 54, no. 1, pp. 36–38, 2011.
- [42] J. Chroboczek, “The babel routing protocol,” 2011.
-