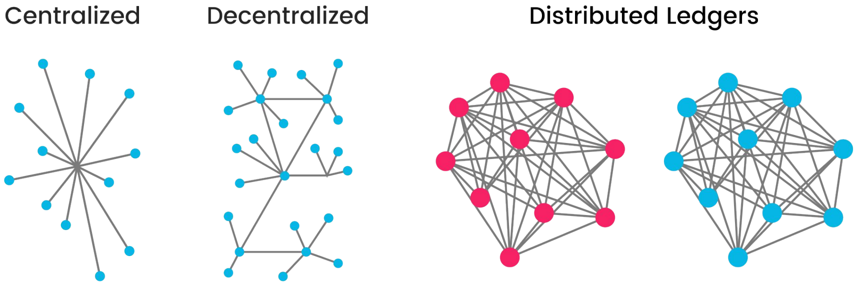
Serverless Messaging System

Synchronizing data on distributed systems with low network connectivity

By **Pranith Hengavalli**
1NT15IS074

Literature Survey: Paper 1

Distributed Databases - Database System Concepts, 5th edition [http://codex.cs.yale.edu/avi/db-book/db5/slide-dir/ch22.ppt]



We implement a homogenous distributed database, where each node is running an identical copy of software.

A replication relational data model is used so each node has a copy of the data.

Advantages of Replication:
Availability: failure of site containing relation *r* does not result in unavailability of *r* as replicas exist.
Reduced data transfer: relation *r* is available locally at each site containing a replica of *r*.

Literature Survey: Paper 2

Multicast DNS (mDNS)

[https://tools.ietf.org/html/rfc6762]

We use the Bonjour protocol for peer discovery on the local network, so we can discover similar services without hard-coding their addresses.

The DHT is based on the Kademlia implementation to create a list of available nodes.

Literature Survey: Paper 3

Public-Private Key Cryptography

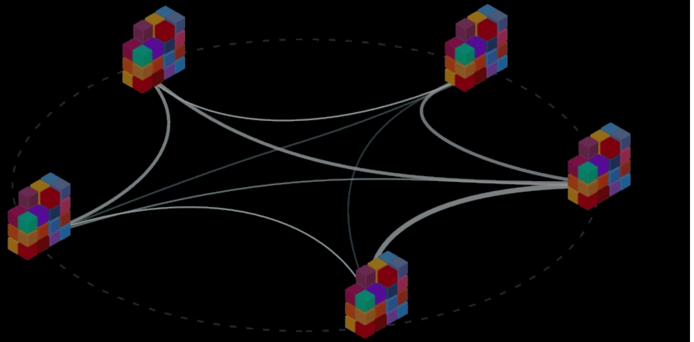
Stallings, William (1990-05-03). Cryptography and Network Security: Principles and Practice. Prentice Hall. p. 165. ISBN 9780138690175.

Using asymmetric cryptography, only the intended receiver can access content signed using their public key.

The message can also be signed by the sender to help the receiver verify authenticity of the source.

The keypair is used to prevent identity fraud in the messaging chain, so the origin of each message is known.

Overview



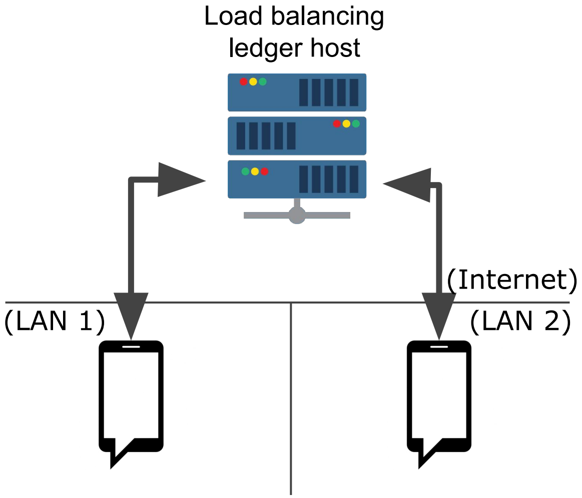
Distributed: An application with no central server to fetch data from.

Redundant: Messages are propagated through a distributed database that is available to all users.

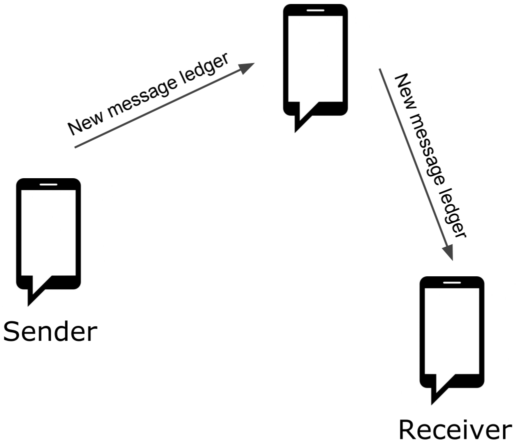
Peer Discovery: Clients communicate directly over local networks using peer discovery.

Self-Propagating: Each client updates to the newest available version of the database.

The message ledger is a distributed database containing all the most recent messages that have been sent.



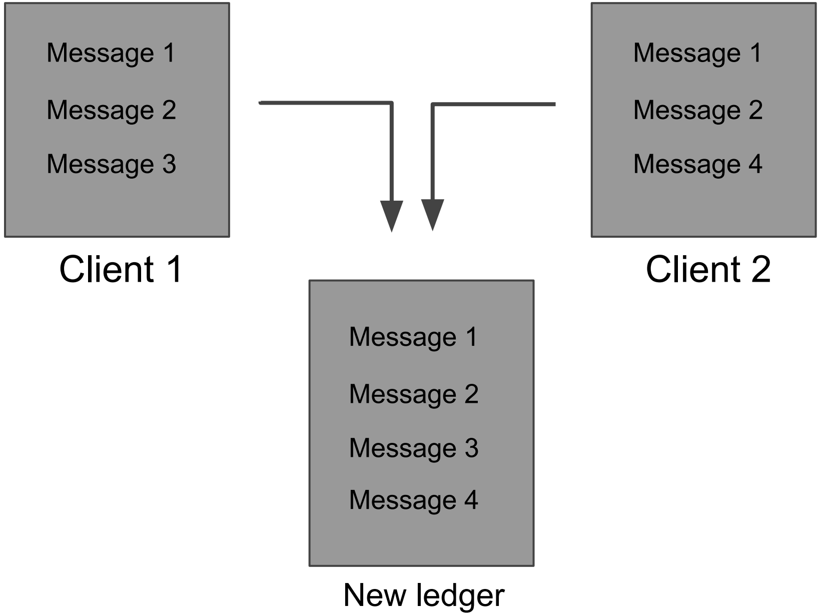
Communication over Internet



Communication over Ad-hoc Network

The Handshake

When two clients discover each other, they compare their ledgers and exchange any missing information.



Message 1
Message 2
Message 3

Client 1

Message 1
Message 2
Message 4

Client 2

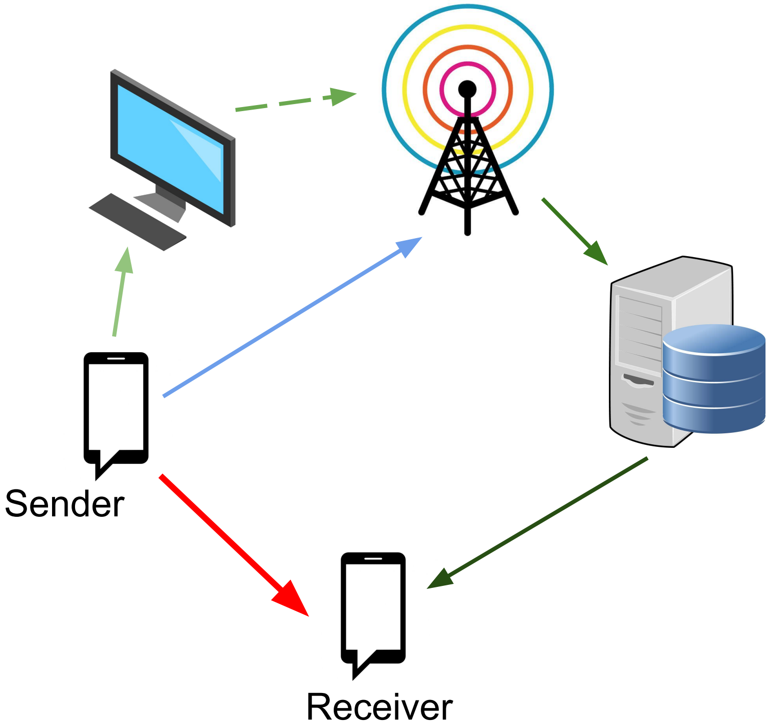
Message 1
Message 2
Message 3
Message 4

New ledger

Use Case: Unstoppable Chat Application

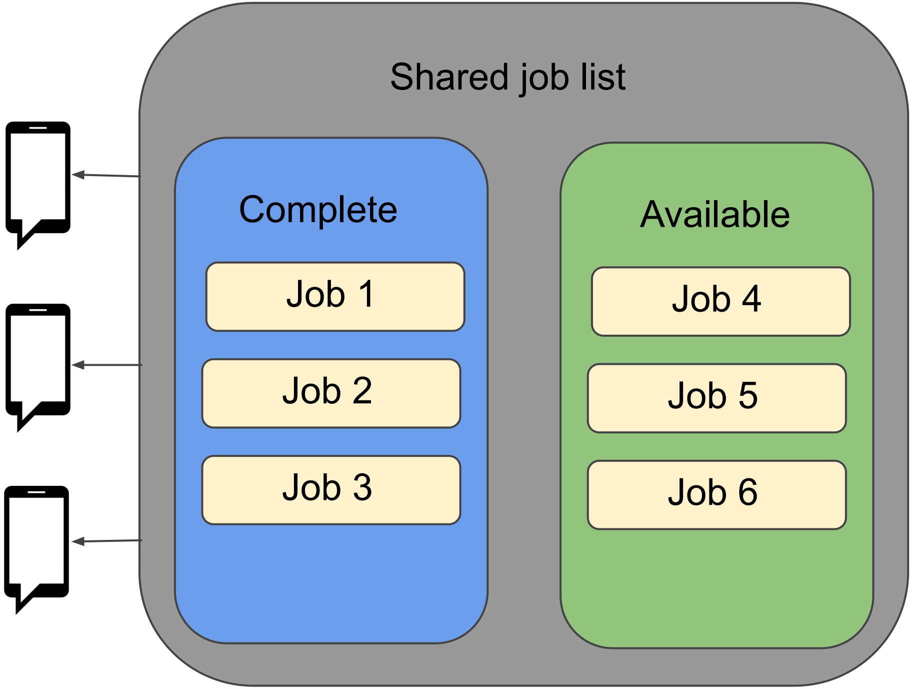
A chat application where messages can be propagated between users on completely separate networks, as long as an eventual indirect connection is made between them.

The message hops from node to node, eventually reaching the intended receiver.



Use Case: Parallel Processing

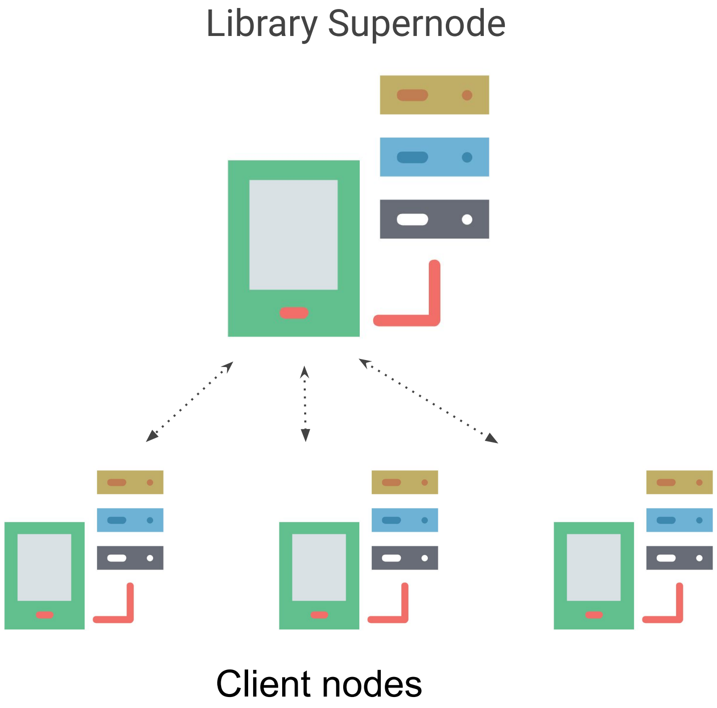
The distributed database can be used to assign tasks concurrently to each node, and to show a pooled list of available and completed jobs.



Use Case: Library Catalog

If a library runs a node with an up-to-date catalog of books, a client can have this database immediately synchronised once it reaches the library network without incurring data charges.

The SMS protocol can then be used to check out a book by updating the record locally, which then gets synced to the main node.



Advantages

- Public message ledger means that false records cannot easily be inserted.
- Messages can reach even air-gapped clients through indirect propagation.
- Multiple redundant backups of data while maintaining privacy.
- Works even without traditional data connection.
- Can be used in isolated environments without configuration.

Network usage pricing vs Local storage pricing

Cloud data storage providers sell storage at ~Rs. 5 per GB.

Network prices are at ~ Rs. 13 per GB.

By having local caches of data, we can reduce overall costs by routing traffic smartly.

Potential issues

Asynchronous messaging

If two gapped clients are passing messages through a single middleman, the conversation is limited by the rate at which the middleman can switch networks.

Over the public network, this rate limitation is removed as messages are propagated over load balancing servers.

Scalability

As user base increases, the ledger is too large to constantly update across devices.

However, implementations like the bitcoin block-chain suggest that the size will be manageable.

Conclusion

By using a homogenous distributed database, we build a network of self healing, self updating nodes that provide a high level of data redundancy for various applications.

This can help us overcome problems related to high network traffic