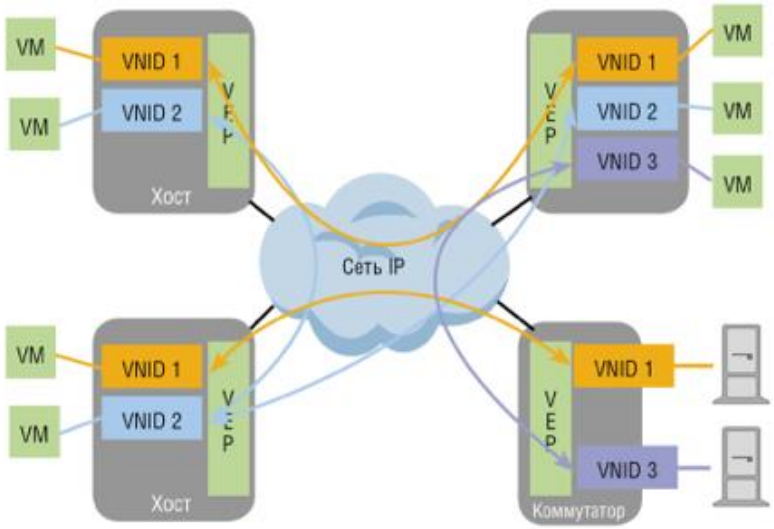
Saint-Petersburg
State University
www.spbu.ru

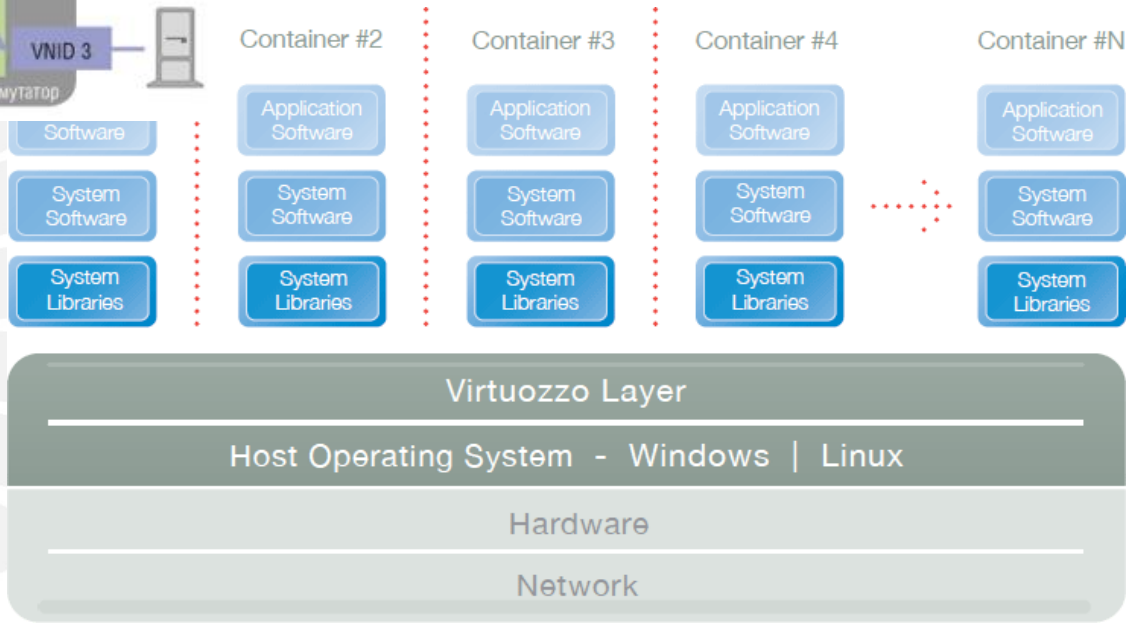
Virtual Blockchain Network: a New Way to Safe Data Exchange

Alexander Bogdanov, Valery Khvatov, , and Nadezhda Shchegoleva

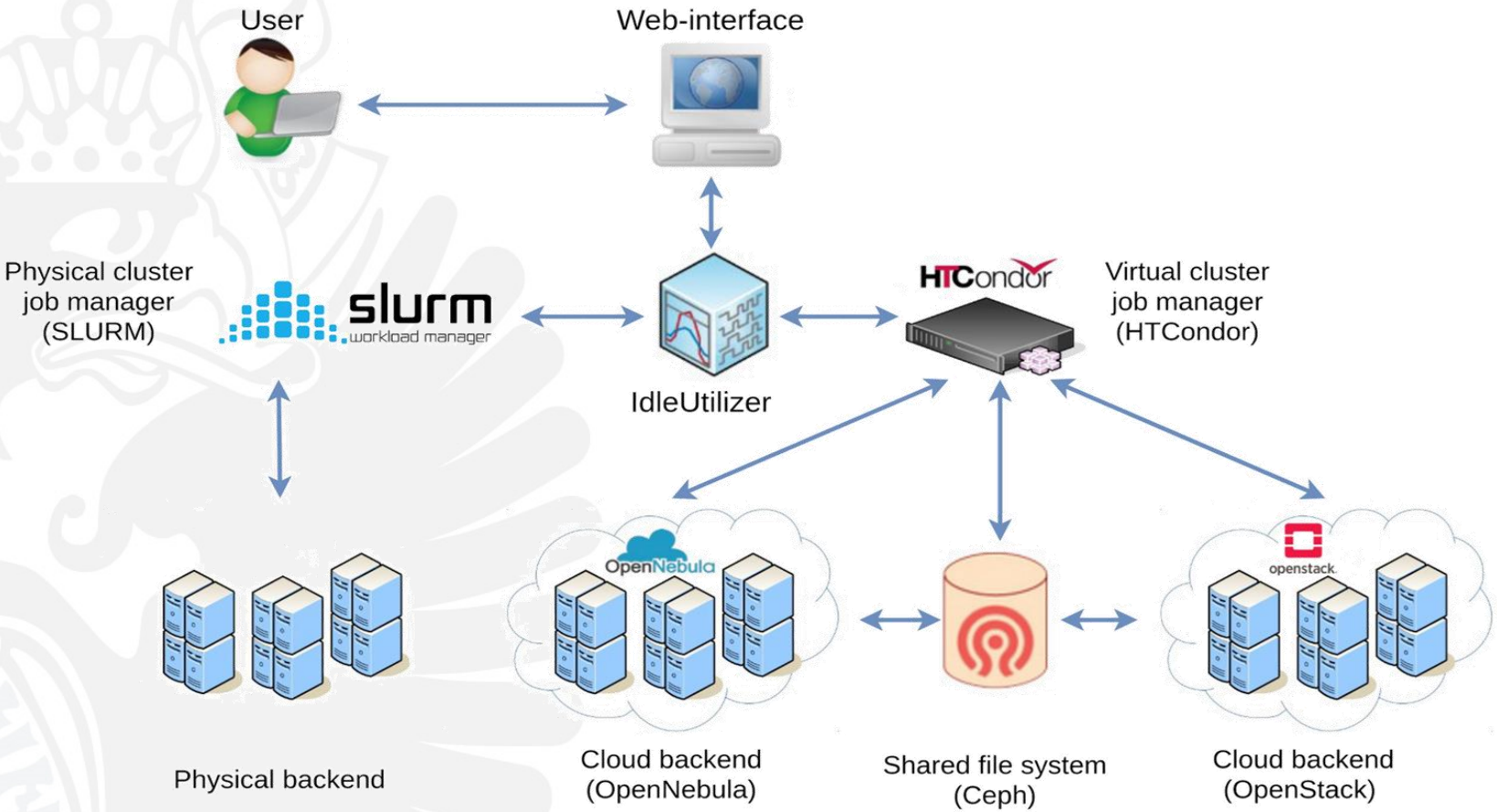
St-Petersburg, 2023

General Purpose VM



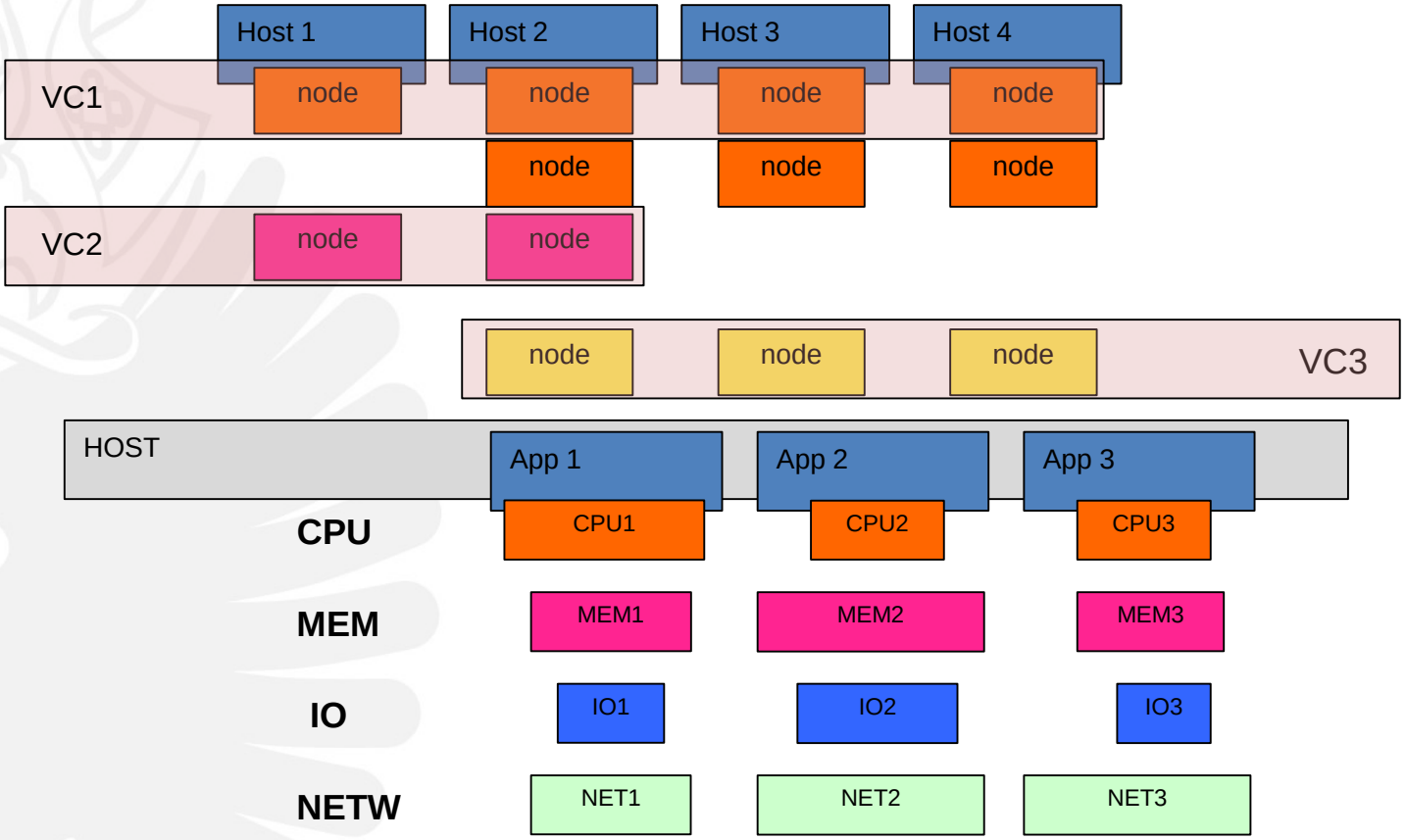


VIRTUAL SUPERCOMPUTER: IDLEUTILIZER SERVICE



DISTRIBUTED VIRTUAL CLUSTER MANAGEMENT SYSTEM

Virtual clusters



DATA API

To be able to work with BIG DATA it is necessary to have a set of tools, that we call Ecosystem, out of which the most important is API:

API is a business capability delivered over the Internet to internal or external consumers

- Network accessible function
- Available using standard web protocols
- With well-defined interfaces
- Designed for access by third-parties

The key features of API are management tools that make it:

- Actively advertised and subscribe-able
- Available with SLAs
- Secured, authenticated, authorized and protected
- Monitored and monetized with analytics

Data API: Unified approach to data integration

- Conventional APIs: Web, Web Services, REST API – not built for analytics;
- Database paradigm: SQL, NoSQL, ODBS and JDBC connectors – familiar to analysts;
- Database Metaphor + API = Data API;
- Specific API for every type of big data (every “V” and their combinations) – under a generic paradigm.

METHODS FOR TESTING BIG DATA APPLICATIONS

Step 1: Data Staging Validation

- Data from various sources like RDBMS, weblogs etc. should be validated to make sure that correct data is pulled into system.
- Comparing source data with the data pushed into the Hadoop system to make sure they match.
- Verify the right data is extracted and loaded into the correct HDFS location

Step 2: "Map Reduce"Validation

- Map Reduce process works correctly
- Data aggregation or segregation rules are implemented on the data
- Key value pairs are generated
- Validating the data after Map Reduce process

Step 3: Output Validation Phase

- To check the transformation rules are correctly applied
- To check the data integrity and successful data load into the target system
- To check that there is no data corruption by comparing the target data with the HDFS file system data

Step 4: Architecture and Performance Testing

- Data ingestion and throughput;
- Data processing sub-component performance.

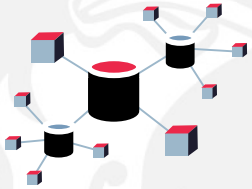
SCALABLE LAYER-1 NETWORK

DECENTRALIZED SOLUTION

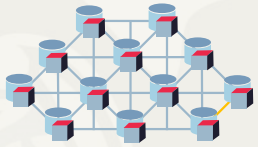
DGT allows for the creation of a decentralized distributed network and for seamless exchange of data, with full security, compliance, and confidentiality.

In principle, there are three possible architectures:

Centralized network: one node responsible for all actions



Distributed network: nodes distribute work through other nodes



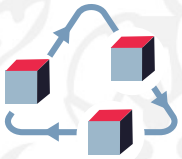
Decentralized network: all nodes participate in exchanging and storing data

DGT

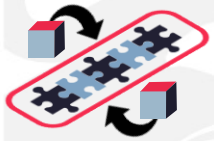
Core functions of the decentralized platform:

- **Collection** of data from distributed sources;
- **Storage** of data in a single immutable registry;
- **Exchange** of confidential data in real time among various information systems;
- **Analysis** of data quality, clearing, standardization, calculation of data metrics
- **Visualization** and data flow analysis

DISTRIBUTED LEDGER TECHNOLOGY



Distributed ledger technologies aim to solve the task of simultaneously processing data by several (or many) nodes. Sub-tasks to be solved include unauthorized data altering, conflict among various inputs, and actualization of data on all nodes.



Web 3.0 is a projected new iteration of the Internet using distributed ledger technologies (blockchain) and token-based economics. Web3 is set to provide increased data security, scalability, and privacy for users, while allowing enterprises to create data-driven economies of value and tokenize both real world and digital assets into new types of financial vehicles.

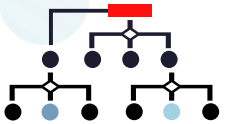
A special **consensus mechanism** is used to synchronize data. The consensus is a mathematically-backed algorithm. The core objective of the consensus is to eliminate conflicts and ensure the integrity of data. There are many different types of Consensus models: Proof-of-Work (PoW), Proof-Of-Stake (PoS), Delegated Proof of Stake (DPos), Practical Byzantine Fault Tolerant (PBFT). DGT uses the stable and rapid **F-BFT Consensus (Federated Byzantine Fault Tolerance) and PoS for Sybil Resistance**.

DGT – is the distributed registry technology of a new generation. IT IS MORE THAN JUST BLOCKCHAIN

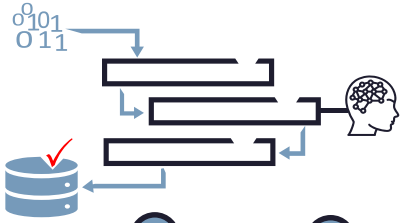
- Support for complex network topologies and dynamic configuration (software defined networking);
- Bleeding edge consensus methods (green, low energy consumption);
- Advanced data storage system (based on DAG – Directed Acyclic Graph);
- Fully secure transactions (guaranteed confidentiality);
- Post-quantum cryptography (quantum resistance)

DGT LAYER-1 FEATURES

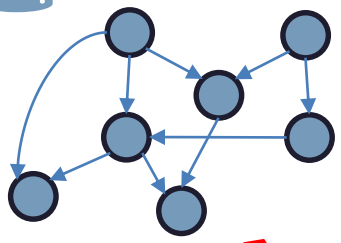
1. **Federative network** formed by nodes, which can be closed, open, a **consortium**. A feature of the network is the special **H-NET architecture**, which allows you to divide the network into separate segments that have different access patterns (public or private segments)



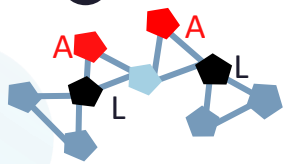
2. Off-chain calculations using **Artificial Intelligence and Confidential Computing methods**



3. Storage in the form of a **directed graph** (DAG), that allows for storing of not only informational objects, but their interconnections, and for supporting a special Heartbeat mechanism for collecting regular information about node metrics



4. Consensus using a secure **arbitration mechanism** protected from Byzantine attacks



H-NET ARCHITECTURE

The platform is designed so that its nodes form a network on top of the Internet. DGT uses its own hybrid architecture and allows you to regulate access to some segments, while leaving free access to others.

01

SEGMENTS

The DGT network consists of segments, each of which differs in access parameters: one is public, and others are private. Private segments define their network access rules, controlling this through the certificate system (X.509 form), independently from the clusters (node groups). In the public segments, clusters are formed automatically. Clusters do not cross segment boundaries.

02

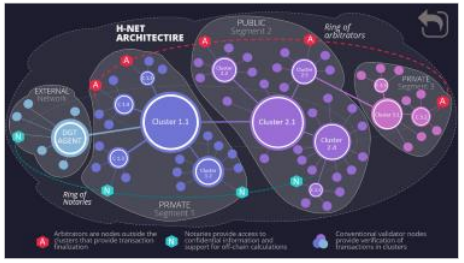
CLUSTERS

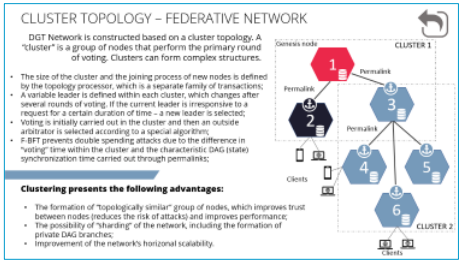
Each node in the DGT is a member of one and only one cluster. Within the cluster, nodes vote for transactions, maintaining F-BFT consensus. Clusters combine several (usually up to 12) nodes and are limited in the number and type of nodes depending on the segment.

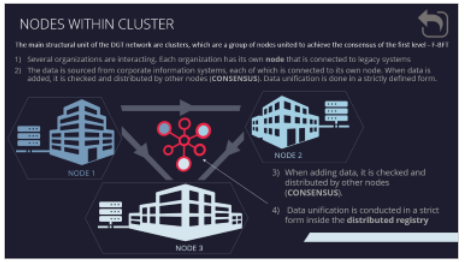
03

NODE ROLES

Nodes in the DGT network can take on different roles that determine their participation in communication. Conventional validators within the cluster verify transactions, constantly changing cluster leaders are responsible for maintaining internal links between clusters (permalinks), and arbitrators ensure network connectivity and ensure the formation of the ledger.



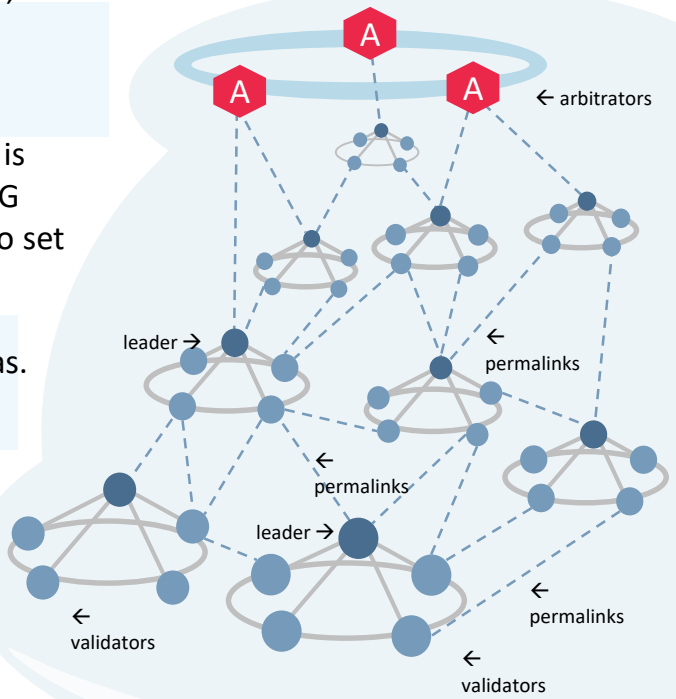




NETWORK DESIGN SUMMARY

The DGT platform has several technical solutions that define the approach to decentralized and distributed computing:

- **Permalinks.** Unlike classical p2p networks, DGT supports communication between nodes through routes called permalinks. Each node can have several connections, one main one, and multiple reserve ones in case of loss of the main channel. Such networking allows you to reduce the cost of communication through the entire network.
- **Ledger.** The ledger has a block structure, on top of which a directed graph (DAG) is built. The block approach allows you to form the classic blockchain structure. DAG supports additional connections between transactions, while also allowing you to set the characteristic network time based on the topological sorting property of the directional graph. Such network time (implemented by a special Heartbeat Mechanism) is the basis of relatively static network configurations – Network Eras.
- **Leader.** Within a cluster era, there is one coordinating node – the leader. The leader dynamically changes when the required volume of transactions within the round is reached (regulated by network settings), as well as when a certain timeout is reached in case the leader fails or does not respond.
- **Arbitrators.** Inside the cluster, nodes "vote" for transactions, collecting the results and transmitting them to special nodes outside the cluster - arbitrators. Arbitrators form a ring and are responsible for the final verification of transactions immediately before inserting them into the registry. Once the insertion is completed and approved by the arbitrators, the transactions are propagated throughout the network using the permalinks.

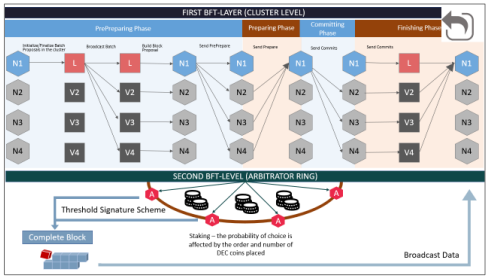


While the cluster and data transfer layer uses a BFT (F-BFT) approach, arbitrators represent a critical part of the DGT infrastructure and defend against Sybil-type attacks with a second layer of a PoS consensus.

F-BFT CONSENSUS

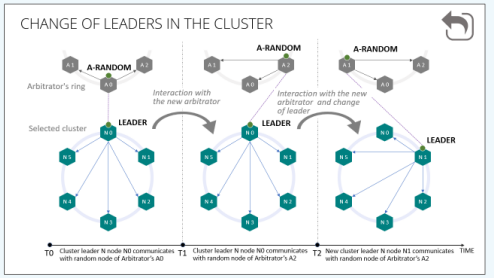
A distributed decentralized environment is characterized by a significant level of complexity. Some nodes can temporarily fail for technical reasons, missing important transactions, others can be captured by intruders and used to distort transactions in order to carry out attacks such as double spending or 51%. A consensus mechanism is responsible for countering these attacks.

- In private networks where access is strictly controlled, an approach based on consensus mechanisms such as CFT (Crash Fault Tolerance), such as RAFT, can be used.



- Public networks such as Bitcoin use the PoW approach, which has proven to be effective in terms of security but does not scale well and is extremely costly from an energy point of view. For small networks with regular participants, it is possible to use consensuses such as PBFT, which allows you to reach consensus through a special communication scheme.
- However, PBFT has great communication complexity, making it inapplicable for networks larger than 50 nodes (communication complexity is $O(n^2)$, где n – total number of nodes).

Under the DGT, F-BFT consensus is applied based on the following assumptions:



- The communication complexity of the network is reduced by the division of the network into clusters, within which a limited number of nodes to achieve consensus is used P-BFT approach with variable leaders who organize interaction;
- Among the nodes are nodes called arbitrators, which form the second level of consensus, their signature is required to insert transactions into the registry;
- Arbitrators are protected by a PoS mechanism that provides Sybil Resistance;
- For the functioning of arbitrators in the public segment, the use of a threshold signature scheme is required

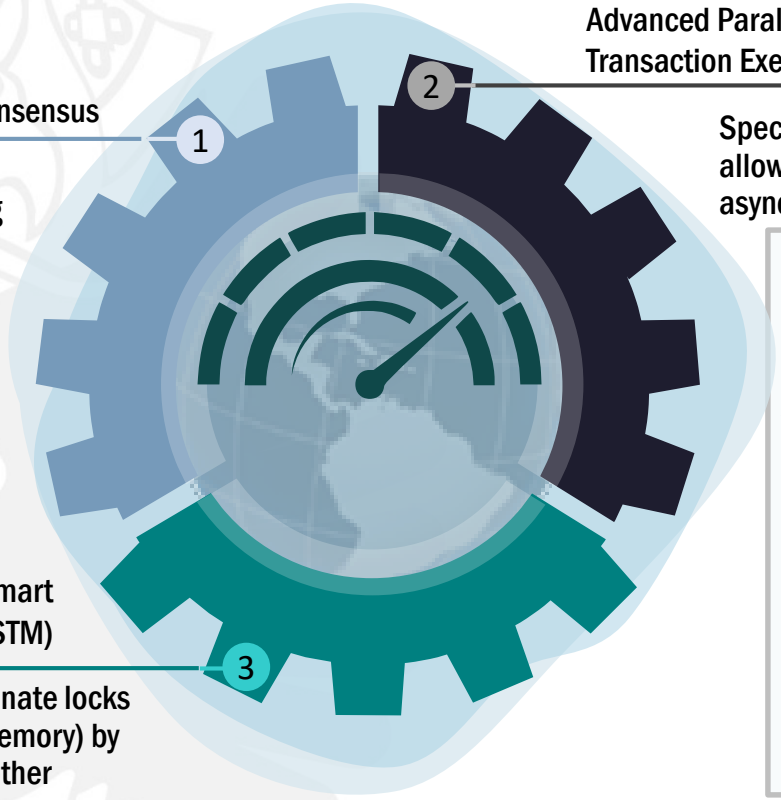
SCALABILITY & PERFORMANCE

Distributed data processing leads to high communication costs, redundant data exchange operations and loss of data processing speed. To overcome these drawbacks, the DGT uses several mechanisms:

Scalable F-BFT consensus

Federated data processing and limiting the number of nodes in a single cluster enables sublinear scaling as the network grows

$$TPS = \Omega\left(\frac{n}{\log n}\right)$$

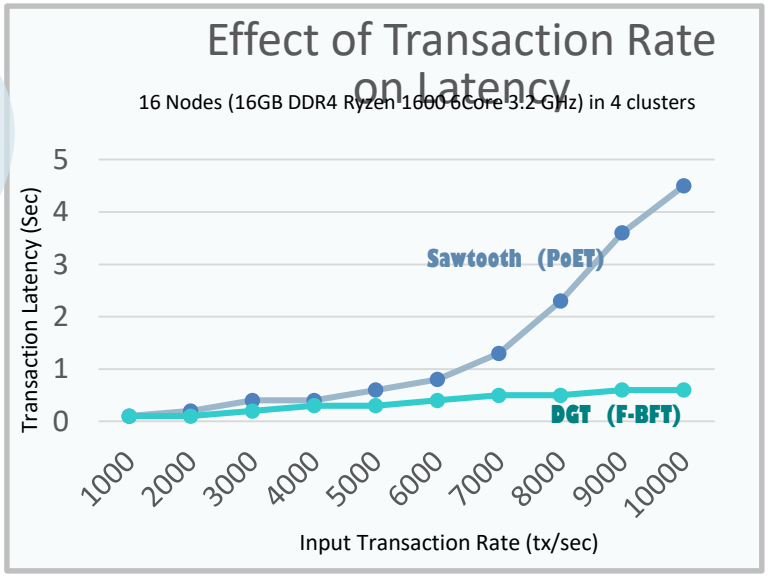


Advanced Parallel Transaction Execution

Special Parallel Schedule mechanism allows you to run transactions asynchronously within a single node

Cluster processing of smart contracts (STM)

Dynamically find and eliminate locks (Software Transactional Memory) by redistributing the load to other clusters



OFF-CHAIN COMPUTATIONS

Off-chain transactions are any transactions processed outside the blockchain. These second-layer protocols aim to circumvent the on-chain's flaws by enabling a cheaper and faster transaction.

Blockchain can be trusted because of its correctness, but not because of its confidentiality. Anyone can view the code and contents of any transaction or smart contract on the network. While this is not a problem for simple payments, there are many use cases that require smart contracts to access personal data. A significant part of web 3 applications requires off-chain computing.

Benefits of off-chain systems

Faster transaction speeds

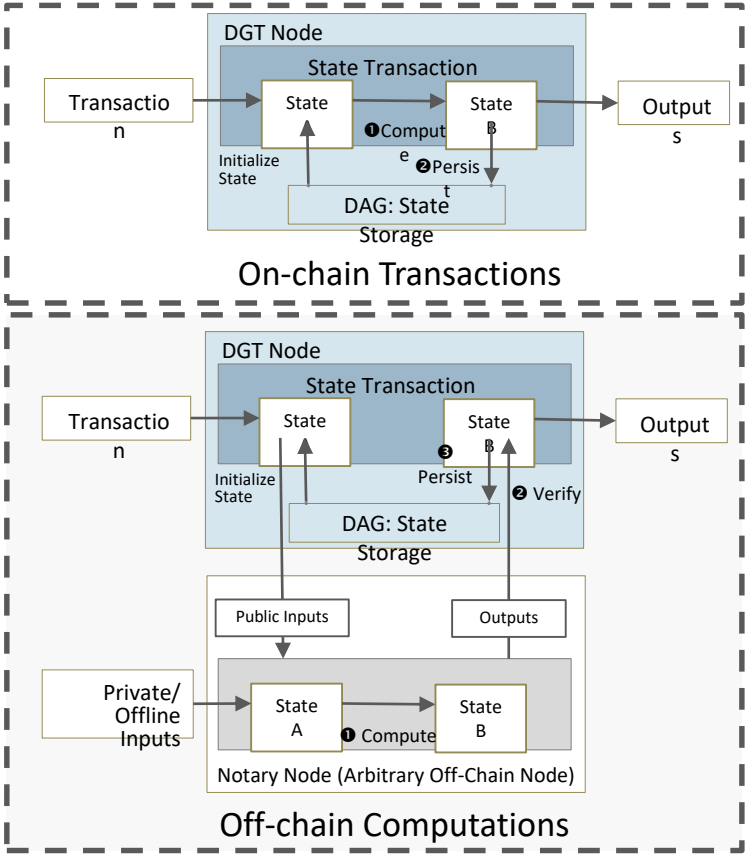
Off-chain transactions don't have to wait for the main blockchain network to confirm a transaction, making it faster or even instantaneous to process.

Lower cost

Transactions confirmed off-chain require little-to-no fees because the process of validation through mining or staking isn't needed. This feature is especially useful in handling large sums of cryptocurrency.

Greater anonymity

Off-chain transactions offer more privacy because the data is not being publicly broadcast to the network



Application scenarios cover private and hybrid transactions, big data storage, dynamic NFT, Real World Assets tokenization verification.

Off-chain calculations are implemented in DGT using nodes of a special kind - notaries that are tied to the network, but have their own secure repositories and an exchange protocol with the DGT network

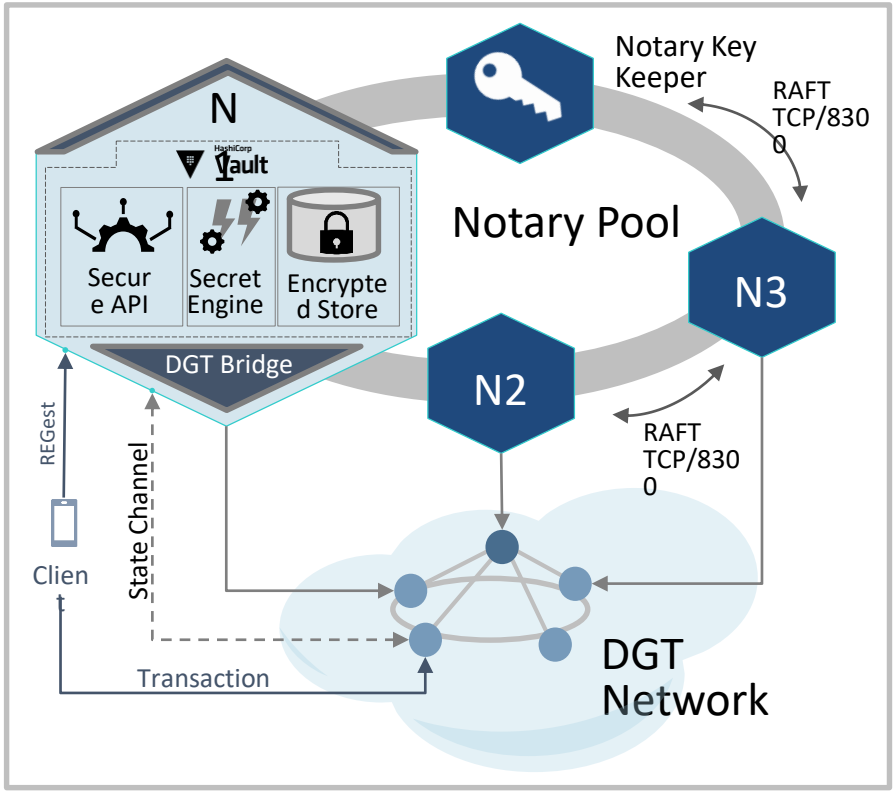
ARCHITECTURE OF NOTARY NODES

Notaries are special nodes that have limited access to the DGT network, as well as equipped with secure storage and support for asynchronous interaction with network clients. Key features of Notary nodes:

Notaries are built on top of the Hashicorp Vault software and have a built-in secure vault, as well as a special set of commands that allow you to save sensitive user attributes (Verifiable Credentials, VC), store them in encrypted form, and carry out verification;

Notaries are united in special pools that allow them to synchronize data with their own consensus based on the RAFT protocol. Each pool has its own key keeper, when registering a notary, as well as objects and their properties, a corresponding entry is recorded in the DAG of the DGT network about the presence of a node with a specified key ("anchor").

For some types of transactions in the DGT, a public parameter is specified that refers to a saved anchor that was previously created by the client when contacting a registered notary. This creates a State Channel that allows you to additionally validate transactions against specified attributes.



Notaries are an auxiliary element of the DGT network, allowing you to perform complex transactions, such as decentralized identification, verification of object attributes during tokenization, etc.

Thank you for attention!

Saint-Petersburg State University

www.spbu.ru