A Blockchain-Based Adaptive Middleware for Optimal Data Storage Selection for Internet of Things

Syed Muhammad Danish, Prof. Kaiwen Zhang, Prof. Hans-Arno Jacobsen
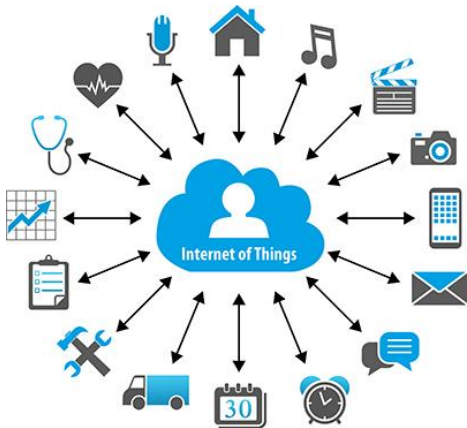

Le génie pour l'industrie

MIDDLEWARE SYSTEMS RESEARCH GROUP MSRG.ORG

1

Motivation

- **Massive** and heterogeneous IoT data.
- Different service requirements i.e. performance, security, privacy, availability and price.
- Single cloud storage results in sub-optimal solution.
- Multi-cloud storage architecture has limitations.
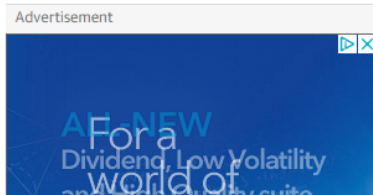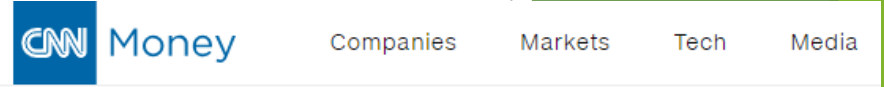- Cloud does not provide adequate privacy.

Motivation

The Guardian

Google to shut down Google+ after failing to disclose user data leak

Company didn't disclose leak for months to avoid a public relations headache and potential regulatory enforcement

CNN Money — Companies · Markets · Tech · Media

Google nukes thousands of Gmail accounts

By Laurie Segall, staff reporter February 28, 2011: 2:29 PM ET

NEW YORK (CNNMoney) -- Imagine opening up your e-mail and finding years of correspondence gone.

As many as 150,000 Gmail users have been confronting that scary scenario throughout the past day. Around 3:00 pm ET Sunday, Google began "investigating reports of an issue" with its popular e-mail service. Over the next few hours, it **confirmed** that a small fraction of Gmail users were experiencing disruptions.
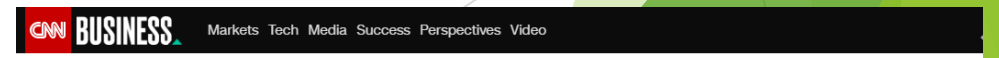
CNBC

TECH

Israeli security company reportedly has tool that spies on Apple, Google and Facebook cloud data

The Guardian

Facebook fined for data breaches in Cambridge Analytica scandal

Firm fined £500,000 for lack of transparency and failing to protect users' information

Support The Guardian's model for open independent journalism

CNN BUSINESS — Markets · Tech · Media · Success · Perspectives · Video

Google still lets third-party apps scan your Gmail data

3

Leading Question

What can be the alternative for Cloud storage and Multi-cloud?

Problem Statement

- Performance constraints of blockchain.
- Decentralized storage technologies.
  - Partially decentralized
  - Fully decentralized
- Better privacy and security.
- Cheap
  - Storj offers 0.015$/GB
  - Google cloud offers 0.026$/GB
- Middleware for intelligent storage technology selection.
- Service requirement of IoT devices and storage technologies parameters.

Blockchain Motivation

▶ Data Integrity

▶ Accountability and auditability of Middleware.
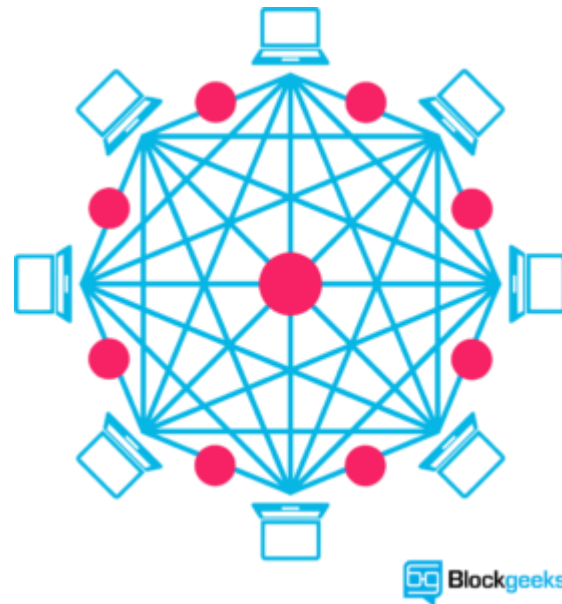
Proposed Architecture
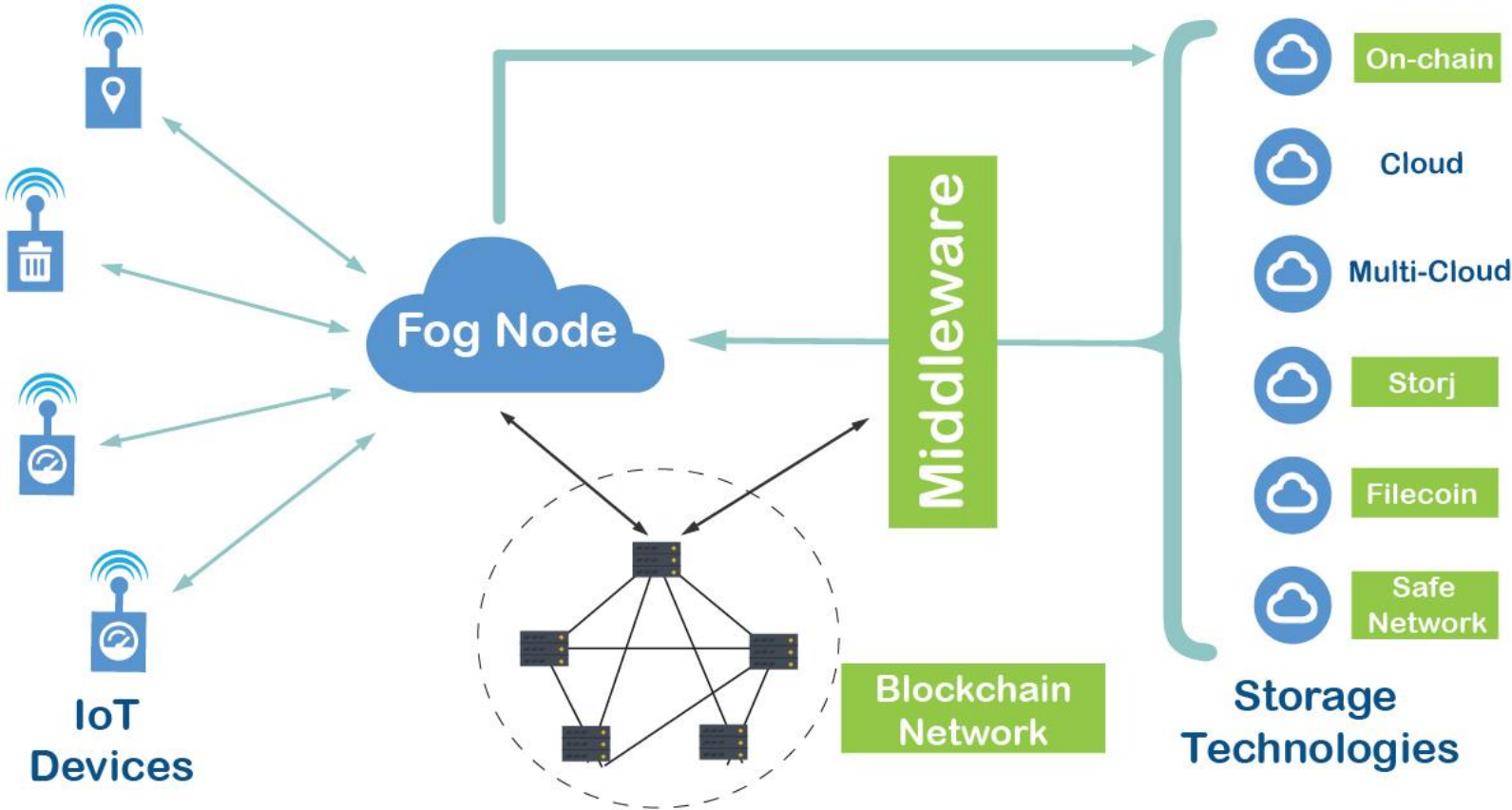
Proposed Solution

- 0-1 Integer programming multi-objective decision optimization problem.

- 1 Decision variable.

- The minimized objective function is composed of five different costs i.e.
  - Bandwidth Cost
  - Storage and Computation cost
  - Latency cost
  - Availability cost
  - Privacy cost

- Similarity to Uncapacitated Facility Location (UFL) optimization problem.

- Our problem is NP-Hard (Proof by reduction).

- Online (Dynamic Phase) and offline (Static Phase) heuristics and meta-heuristics.

8

Problem Formulation

$$\mathbb{C}(\overline{y}) = C_{BW}(\overline{y}) + C_{CS}(\overline{y}) + C_{LT}(\overline{y}) - C_{AV}(\overline{y}) - C_{PR}(\overline{y})$$

$$\text{minimize} \quad \mathbb{C}(\overline{y})$$

$$\text{subject to} \quad \overline{y} \in \mathcal{Y},$$

$$\sum_{j \in J} \sum_{i \in I} P_j F_{d,i} \le \mathcal{B}, \forall i \in I, \forall j \in J,$$

$$\sum_{j \in J} \sum_{i \in I} (Gj + Hj) F_{d,i} \le \mathcal{M}, \forall i \in I, \forall j \in J,$$
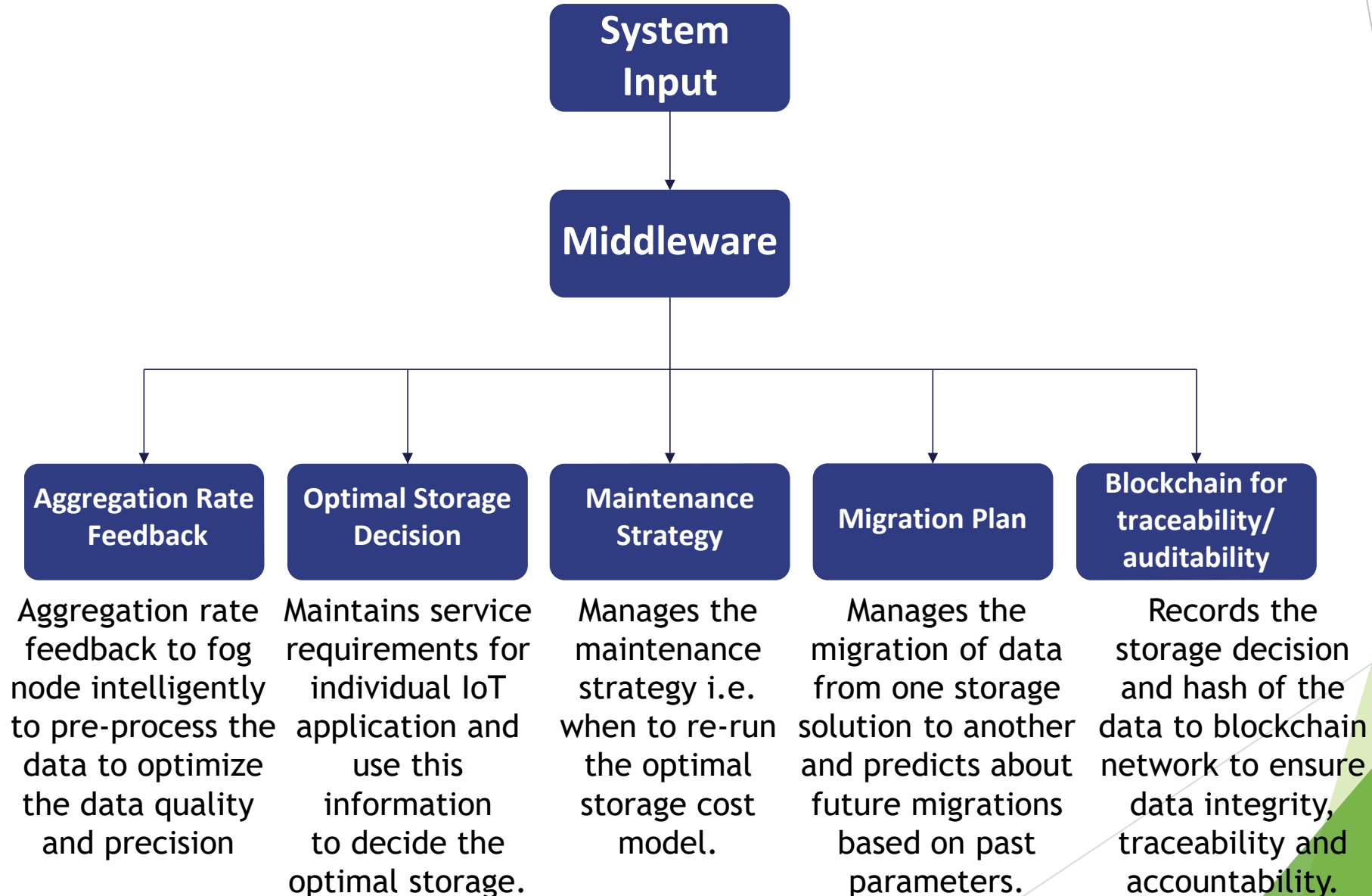
$$l_j \le \mathcal{L}, \forall j \in J,$$

$$\alpha_j \le \mathcal{A}, \forall j \in J,$$

$$\beta_j \le \mathcal{B}, \forall j \in J$$

Survey of Decentralized Storage Technologies

Storage	Decentralization	Smart Contract	Blockchain	Price Decision	Price	Data Location	Anonymity
Cloud	Centralized	No	No	Cloud	0.026 $ /GB/Month	Cloud server	No
Storj	Partially Decentralized	No	No	Storj	$0.015 /GB /MONTH	Multiple Storage nodes worldwide	No
Filecoin	Fully Decentralized	Yes	Yes	Storage Nodes	Not decided yet	Multiple Storage nodes worldwide	No
Safe Network	Fully Decentralized	No	No	Safe Network	Not decided yet	Multiple Storage nodes worldwide	Yes

Proposed Middleware Functionalities

System Input

Middleware

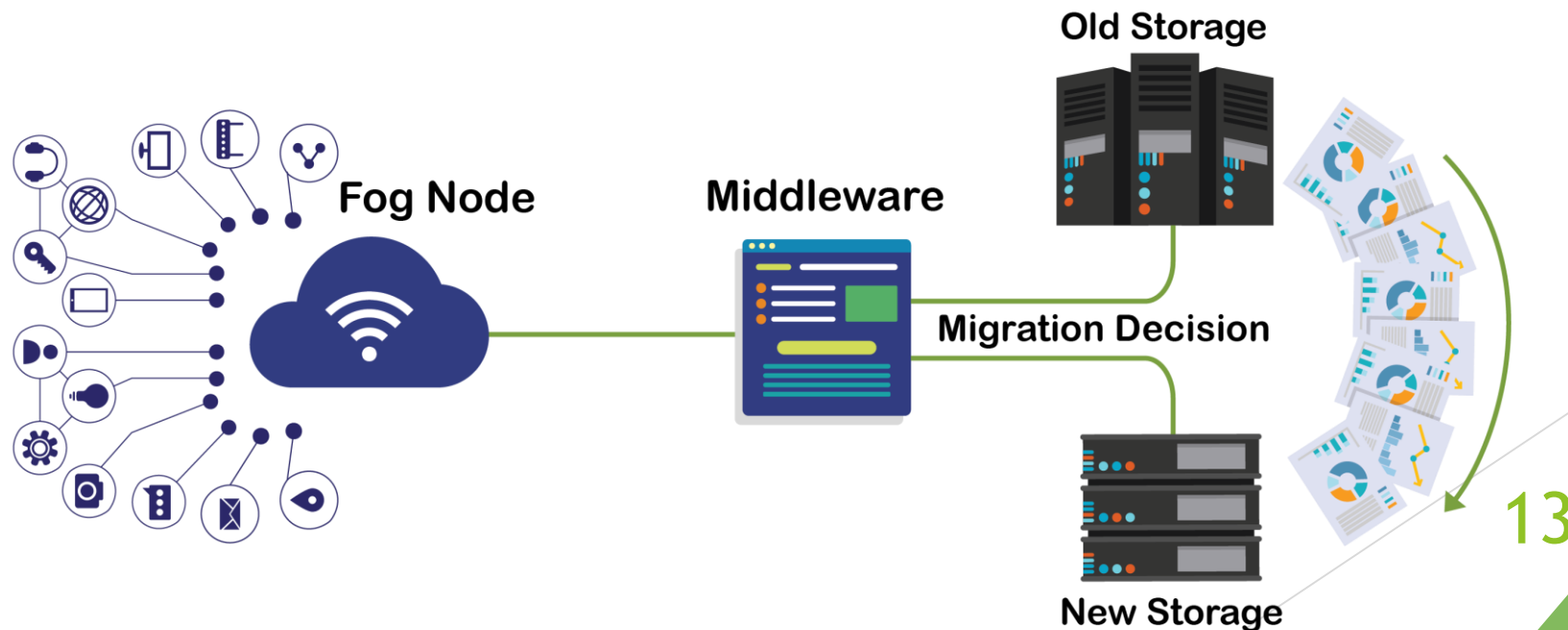| **Aggregation Rate Feedback**	**Optimal Storage Decision**	**Maintenance Strategy**	**Migration Plan**	**Blockchain for traceability/ auditability**
Aggregation rate feedback to fog node intelligently to pre-process the data to optimize the data quality and precision	Maintains service requirements for individual IoT application and use this information to decide the optimal storage.	Manages the maintenance strategy i.e. when to re-run the optimal storage cost model.	Manages the migration of data from one storage solution to another and predicts about future migrations based on past parameters.	Records the storage decision and hash of the data to blockchain network to ensure data integrity, traceability and accountability.

Maintenance Strategy

▶ Two phases of operation

  ▶ Static phase

  ▶ Dynamic phase

▶ Anomaly detection mechanism

  ▶ Training data (statistical methods)

  ▶ Decision in real-time

▶ Machine Learning

▶ Computational Efficiency

▶ Comparison with fixed time slot

Fog Node

Dynamic Parameters of IoT Devices

Middleware

Storage 1

Dynamic Parameters of Storage Technologies
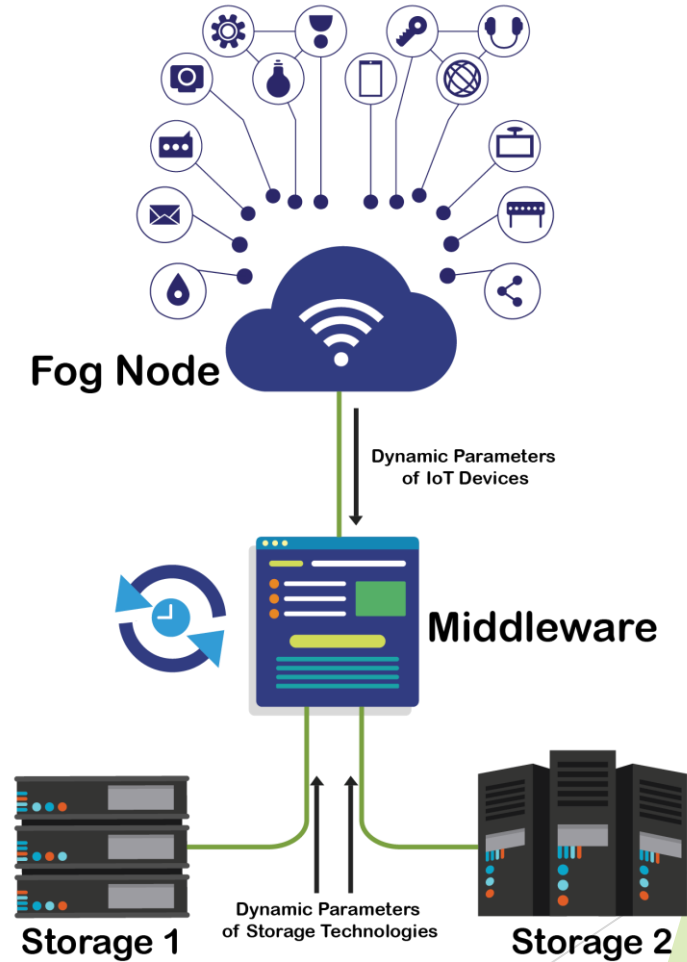
Storage 2

Migration Plan
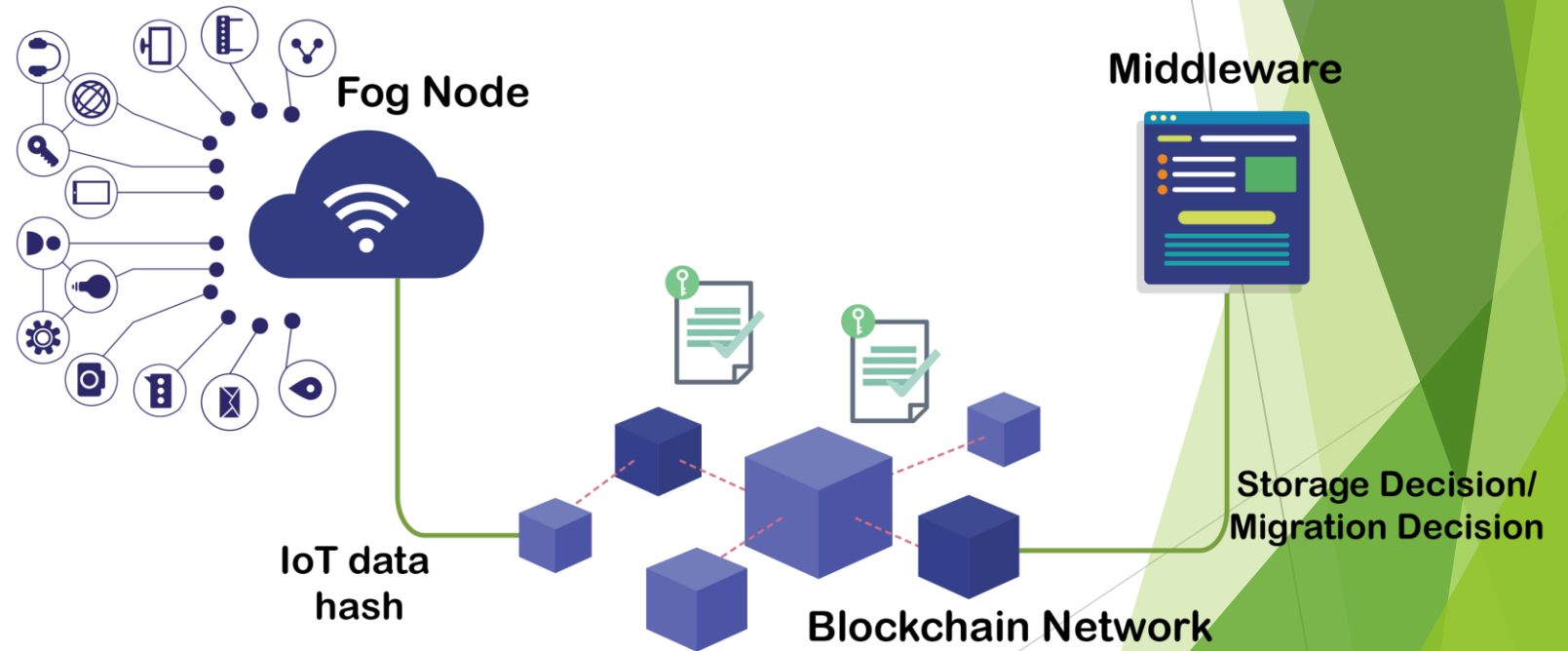
▶ Migrate old data.

▶ Defining importance of data by weights.

▶ Condition to migrate data.

Price of Current storage > Price of new storage + Price of migration

13

Blockchain for Traceability and Auditability

- Traceability, auditability and accountability.
- Optimal Storage decision
- Migration Decision
- Important data on-chain
- Data integrity

14

Summary

- Massive and Heterogeneous IoT data generation

- Decentralized Storage Technologies

- Middleware for Optimal Storage Selection

- Multi-objective IP optimization problem (NP-Hard problem)

- Maintenance Strategy

- Migration Plan

- Blockchain for traceability and auditability.

Thank you