FOG-engine: Towards Big Data Analytics in the Fog

Farhad Mehdipour

Tech Futures Lab & Unitec Institute of Technology
Auckland, New Zealand
farhad@techfutureslab.com

Bahman Javadi
Western Sydney University,
Sydney, Australia

Aniket Mahanti

University of Auckland,
Auckland, New Zealand

The 2nd International Conference on Big Data Intelligence and Computing, Auckland, New Zealand, August 8-10, 2016
Introduction

• Challenges of the current cloud-based platforms
  – The cloud physically located in a distant datacenter
    → Latency
  – Vertically fragmented
  – Real-time processing large quantities of IoT data
    → more security, capacity, and analytics challenges
  – Incapability of current cloud for efficient Big Data Analytic

Our solution

• An on-premise and real-time data analytic engine (FOG-Engine) located near where data is generated
• Collaboration and proximity interaction between IoT devices in a distributed and dynamic manner
Current platforms issues: Not Fully Integrated, No low-latency, and might be Expensive

- **Service Ecosystem:** Fragmented IoT infrastructure → designers need to interact with **many services** from sensors/actuators to data analytic

- **Cloud:** latency
  - The cloud uses **virtual machines** → unnecessary data movement

- **Network:** latency, bandwidth and cost
  - Large geographical distance → **Higher Latency**
  - The aggregated b/w of sensors >> network b/w
  - Big Data is heavy to move → **Higher Cost and Latency**

- **Raw data:** the size can be huge (e.g. camera)

- **Network & Internet Infrastructure:**
  - Big Data Flow
  - Feedback

- **IoTs**
  - Network of Sensors and Actuators (Physical World)
Fog Computing

• **The Fog**
  – extends the cloud computing paradigm to the edge of the network,
  – enables a new breed of applications and services
  – an appropriate solution for the applications and services that fold under the umbrella of the IoTs.

• **Benefits**
  – low latency
  – location awareness
  – widespread geographical distribution
  – mobility support
  – the strong presence of streaming and real-time applications
  – heterogeneity
# Related Works

<table>
<thead>
<tr>
<th>Related Works</th>
<th>AWS</th>
<th>Microsoft</th>
<th>IBM</th>
<th>Google</th>
<th>Alibaba</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Service</strong></td>
<td>AWS IoT</td>
<td>Azure IoT Hub</td>
<td>IBM Watson IoT</td>
<td>Google IoT</td>
<td>AliCloud IoT</td>
</tr>
<tr>
<td><strong>Data Collection</strong></td>
<td>HTTP, WebSockets, MQTT</td>
<td>HTTP, AMQP, MQTT and custom protocols (using protocol gateway project)</td>
<td>MQTT, HTTP</td>
<td>HTTP</td>
<td>HTTP</td>
</tr>
<tr>
<td><strong>Security</strong></td>
<td>Link Encryption (TLS), Authentication (SigV4, X.509)</td>
<td>Link Encryption (TLS), Authentication (Per-device with SAS token)</td>
<td>Link Encryption (TLS), Authentication (IBM Cloud SSO), Identity management (LDAP)</td>
<td>Link Encryption (TLS)</td>
<td>Link Encryption (TLS)</td>
</tr>
<tr>
<td><strong>Integration</strong></td>
<td>REST APIs</td>
<td>REST APIs</td>
<td>REST and Real-time APIs</td>
<td>REST APIs, gRPC</td>
<td>REST APIs</td>
</tr>
<tr>
<td><strong>Data Analytics</strong></td>
<td>Amazon Machine Learning model (Amazon QuickSight)</td>
<td>Stream Analytics, Machine Learning</td>
<td>IBM Bluemix Data Analytics</td>
<td>Cloud Dataflow, BigQuery, Datalab, Dataproc</td>
<td>MaxCompute</td>
</tr>
</tbody>
</table>
Our View: Decentralized Hierarchical Big Data Processing on the Edge

**Centralized:** The FOG-Engine supports all the functionalities of the system.

**Data Preprocessing & Analytic**

**FOG-Engine:** Data Analytic micro-Engine

**FOG:**

**Network Access**

**CLOUD**

- Global data analytics on semi-structured data
- **WAN:** low b/w (x10MB/s)

**Semi-structured data**

**LAN:** very high b/w (x1GB/s)

**MAN:** high b/w (x100MB/s)

**Network Access**

**IoTs**

**Network of Sensors and Actuators (Physical World)**
How to Realize Big Data’s Vs: Velocity, Volume, ...

Our Solution

Current Cloud

Our Solution

Current platform
Towards Real-Timeness, with Added Values

Current Cloud

Our Solution

Real-Timeness

Higher Speed
Higher bandwidth
Flexibility
Heterogeneity
Extensibility, etc.
More Engineerable
## FOG-Engine vs. Cloud

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>FOG-engine</th>
<th>Cloud platform</th>
</tr>
</thead>
<tbody>
<tr>
<td>Processing hierarchy</td>
<td>Local data analytics</td>
<td>Global data analytics</td>
</tr>
<tr>
<td>Processing fashion</td>
<td>In-stream processing</td>
<td>Batch processing</td>
</tr>
<tr>
<td>Computing power</td>
<td>GFLOPS</td>
<td>TFLOPS</td>
</tr>
<tr>
<td>Network Latency</td>
<td>Miliseconds</td>
<td>Seconds</td>
</tr>
<tr>
<td>Data storage</td>
<td>Gigabytes</td>
<td>Infinite</td>
</tr>
<tr>
<td>Data lifetime</td>
<td>Hours/Days</td>
<td>Infinite</td>
</tr>
<tr>
<td>Fault-tolerance</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>Processing resources</td>
<td>Heterogeneous (e.g. CPU, FPGA)</td>
<td>Homogeneous (Data center)</td>
</tr>
<tr>
<td>Versatility</td>
<td>Only exists on demand</td>
<td>Intangible servers</td>
</tr>
<tr>
<td>Provisioning</td>
<td>Limited by the number of FOG-engines in the vicinity</td>
<td>Infinite, with latency</td>
</tr>
<tr>
<td>Mobility of nodes</td>
<td>May be mobile (e.g. in the car)</td>
<td>None</td>
</tr>
</tbody>
</table>
A Typical Data Analytic Flow

1. **Raw data**
2. **Data collection/integration**
3. **Data cleaning**
4. **Feature extraction & transformation**
5. **DB storage**
6. **Interpretation and presentation**
7. **Data analytics**

**Data Analysts**
A Modified Data Analytic Flow

1. Raw Data from IoTs

2. Data collection/integration

3. Data cleaning

4. Feature extraction & transformation

5. DB Storage

6. Interpretation & presentation

7. Data analytics

8. Data from other FOG-engines/IoT nodes

9. Cloud

10. FOG-engines/Users

11. Interpretation & presentation

12. Data Analytics

13. Data integration

14. DB Storage
CLOUD

Centralized data analytics and storage

Fog

Fog engine

Network Access

Raw data stream

Smart City
General Architecture of FOG-Engine

Data Analytics & Storage Unit
- Data Analytic Engine
- Data Cleaning, Aggregation & Visualization
- Data Collection and Import
- Data Storage System

Orchestration Unit

Communication Unit
- Network Interface to Physical World
- Peer-to-Peer Networking API
- Network Interface to Cloud (Gateway)

IoTs

Collaborating peers in the FOG

Cloud
Detailed Architecture of FOG-engine

USB: Universal serial bus
BT: Bluetooth
UART: Universal Asynchronous Receiver/Transmitter
SPI: Serial Peripheral Interface Bus
GPIO: General-purpose input/output pins
Preliminary Results

- Implementation platform: Raspberry Pi 2.0 and 3.0

- Scenarios
  1) Multiple receivers, multiple analysers, and multiple transmitters scenario
  2) Multiple receivers, multiple analysers, and single transmitter scenario
  3) Multiple receivers, single analyser, and single transmitter scenario
Scenario II

Multiple receivers, multiple analysers, and single transmitter scenario:

- Multiple FOG-engines receive and analyse data individually,
- FEs data is transmitted to the cloud via one of them which acts as a cluster head
The transmission time (ms) for various data sizes

**IoT-FE communication**

FE to Cloud communication time is significant compared to IoT-FE and FE-FE communication times (as expected).

The transmission time (ms) for various data size

**FE-Cloud communication**

The transmission time (ms) for various data sizes

**FE-FE communication**
The transmission time of 1KB data for various communication types

Data transmission time increases by increasing the size of data for all communication types.
Thank you

Any Questions/Comments?