Sensor Edge Processing for the IoT

Class 2: Data Considerations

February 28, 2017

Louis W. Giokas

DesignNews

Presented by:

CEC CONTINUING EDUCATION CENTER

Digi-Key ELECTRONICS

This Week's Agenda

Monday	Architecture
Tuesday	Data Considerations
Wednesday	Devices – Sensor Level
Thursday	Devices – Aggregation/ Communication Level
Friday	Algorithms

Course Description

- The cloud is the central gathering point for all data these days. The Internet of Things (IoT) generates a lot of data. Getting that data to the cloud introduces latency which is a problem for IoT analytics and real time decision making.
- The answer to this is processing data at the "edge".
- New devices, and not a few older devices, make this possible.
- We will look at architectures and algorithms useful for edge processing of sensor data in the IoT.

Today's Agenda

- Types of Data
- Data at the Sensor
- Data at the Edge
- Data in the Cloud
- Conclusion/Next Class

Types of Data

- There are lots of different types
  - Obvious, I know
  - But, this is not transactional data
    - We don't need a transactional database system
- There are many types of sources
  - We normally class them as Industrial and Consumer
  - There are other classes we will also discuss

Types of Data

- Sensor Data
  - Most of the time we are talking about sensors
    - There are lots of types of sensors
      - Simple sensors: temperature, switch (open/closed)
      - Status sensors: take a number of measurements and consolidate (e.g., engine monitor sensors)
      - Video streams (e.g., security cameras, inspection systems)
      - Other continuous streasms (e.g., radar, Lidar)
  - We define sensor data as always flowing "up" the chain of devices

Types of Data

- Command data
  - The IoT is not a one way street
    - Command and control architecture
  - Commands sent back down from the processing or analytics level tend to be time-critical
    - This is not necessarily “real-time” command and control
  - May require more communication
    - Ack/nack
  - Broadcast
    - Parameter updates

Data at the Sensor

- Assume all sensors in the IoT have some communication capability
  - This may be to a local communication aggregator
  - Typically we store a minimum amount of data at the sensor
    - Offload to higher levels as soon as possible
  - Buffering limited

Data at the Sensor

- Sensors vary widely in amount of processing power and memory available
- Power (electrical) is another potential limiting factor
  - Ultra low power such as energy harvesting devices
    - Also RFID powered where signal provides the power
  - Battery powered
    - Internal or external
  - Line powered

Data at the Sensor

- Communications channels may dictate the type and amount of data that can be stored
  - Intermittent communications requires some form of buffering
    - Typically store only raw sensor data
  - Cellular networks allow (periodic) connectivity that is generally highly reliable
  - Wi-Fi connectivity allows high speed access to the Internet, such as used in Consumer IoT
  - Direct IP connectivity, such as used in Industrial IoT settings, allows high speed continuous communication

Data at the Sensor

- Sensors can store some history
- Parameter data
  - Allow the sensor to make decisions/detect anomolies
  - We will discuss further under algorithms
- Commands
  - Formats, tables and history

Data at the Edge

- Edge devices have a lot more storage and processing power, potentially, than the sensors connected to them
- It is here we can store all the necessary data for processing more complex algorithms
- In some cases we can even have mass storage devices
  - Currently likely to be solid state disk

Data at the Edge

- Edge processing scenarios may involve locally meaningful computations and actions
  - Work cell control in a factory
  - Home control and assisted living applications
  - Local traffic control
- While these applications and others will also interact with higher level analyses, meaningful work can be done with local data

Data at the Edge

- Example:
  - Sewer flow meter data
    - Collected every five minutes
      - 12 times per hour, 288 times per day, 2,016 time per week, 104,832 times per year
    - If there are several sensors in a particular area, say 10, we get 1,048,320 sensor readings per year. If each is 100 bytes long, that translates to 100 MB per year.
      - This can be stored in a SoC device in flash memory
    - If there is a control function, then long term trends can be observed and used to make decisions
  - Similar data rates are used for systems like large diesel engines

Data at the Edge

- Data stored at the edge, as with most IoT data does not need to be stored in a transactional (read relational) DBMS
- Non-SQL and other non-traditional DBMS technologies are more appropriate
  - They are also easier to implement on various types of equipment
- A RDBMS can be used, but is generally overkill

Data in the Cloud

- In general, most data will eventually be moved to the cloud
- This allows long term archiving of data
  - Sometime called the Data Lake
- This also allows for exploratory data analysis on a large scale
  - Necessary to find out what patterns there are in the data as it develops over time
    - In many cases such data has not been available in such detail

Data in the Cloud

- The cloud layer can act as a assistive resource to edge resources
  - Users and devices may move from one edge region to another
  - The cloud assists in tracking and in providing continuity between edge devices
  - Especially useful for vehicular systems and user based applications such as Ambient Assisted Living (AAL)

Data in the Cloud

- Distributed IT resources make up the edge, sometimes called the fog (re: yesterday's lecture)
  - Replication of data as it is needed
    - E.g., moving between systems
  - Broadcast of common data
  - Periodic update of parameters and control information
  - Mobile is a driver

Data in the Cloud

- Database types
  - One of the true Big Data applications
  - Offered by Cloud vendors as standard services
  - Google
    - Google Cloud Storage: a unified object store targeted at analytics
    - Cloud Big Table: key/value datastore
    - Cloud Datastore: NoSQL DBMS with ACID transactions

Data in the Cloud

- Database types, continued
  - Amazon
    - DynamoDB: NoSQL DBMS
    - Amazon Redshift: BI Data Warehouse
      - Works with existing BI tools
  - Others
    - Rackspace: ObjectRocket (NoSQL)
    - Oracle: Oracle NoSQL
    - IBM: Cloudant

Data in the Cloud

- Analytics approaches that work best in a cloud environment
  - Statistical Learning (or Machine Learning)
    - Supervised and unsupervised learning techniques
    - Regression
    - Classification
    - Neural Networks
    - Trees/Random forests
    - Support Vector Machines

Conclusion/Next Class

- Today we looked at data, where it resides and how it might be used at each level
- There are clear differences in the capabilities at each layer and different types of data we might store there
- The layers cooperate to allow responsiveness in the dynamic IoT environment
- Tomorrow we will look at the sensor devices and some of their capabilities