Internet of Things
Data Acquisition to Machine Learning.

Ciira wa Maina
cwamaina.dekut@gmail.com

12th November, 2018
The Internet of Things

- Connecting things over the internet
- We can measure physical states and use this information to guide actions
Internet of Things

*Adding digital inputs/outputs to physical things* - Jan Jongboom
Use Cases Closer Home

- Agriculture, Environmental Monitoring, ...
Getting the Data

- Before data scientists can work their magic on data, it must be acquired
- Sensor systems provide a rich data source
- But data acquisition is not always easy...
Getting the Data

- It may also require armed escort!
LoRa

- Low power, long range network
- Ideal for low bandwidth situations such as sending sensor data
LoRa Networks

Diagram:
- Devices connected via LoRa to Gateways
- Gateways connected via TCP/IP to LoRaWAN network
- LoRaWAN network includes:
  - Network server
  - Application server
  - Join server (new in LoRaWAN 1.1)
- Your application accesses the network through the LoRaWAN network.
Devices
Mbed OS

- A free, open-source operating system for embedded devices
- Ideal for IoT
- Allows development of applications in C/C++
Mbed OS

- Mbed OS implements
  - Radio drivers
  - Networking
  - Hardware drivers
  - Storage
- Allows rapid prototyping
Mbed OS Code Development

- Online compiler https://os.mbed.com/compiler
- Offline CLI
- Try things out on the online simulator https://labs.mbed.com/simulator

Arm Mbed OS simulator

```c
#include "mbed.h"

DigitalOut led(LED1);

int main() {
    while (1) {
        led = !led;
        printf("Blink! LED is now %d\n", led.read());
        wait_ms(500);
    }
}
```
Babbler: A device to monitor cargo on transit
A device to monitor tilting of power poles
Gateways
Gateways

- Gateway placement is important! The higher the better
Gateways

- Jared and Stephen taking the gateway to new heights.
Gateways
Gateways
Network Servers

- Servers that understand the LoRa protocol
- Companies offer this as a service
Applications

- The network server forwards data to database
- Database could be from any cloud provider or local host
- Applications query data and use it to guide decisions
Demo Application

- Temperature and humidity measurement
- Smart office motion detection
- Data transmitted to the Things Network
- Data stored on Amazon Web Services or Cayenne
Sense the Environment

```c
56 // Send a message over LoRaWAN
57 static void send_message() {
58    CayenneLPP payload(50);
59    int attempt = 0;
60
61    float temperature = 0.0f;
62    float humidity = 0.0f;
63    int error_code;
64
65    while (attempt++ < SENSOR_READ_ATTEMPTS) {
66        error_code = temperature_humidity_sensor.readData();
67        if (error_code != ERROR_NONE) {
68            printf("Error = %d\n", error_code);
69            wait_ms(SENSOR_WAIT_TIME_MS);
70            continue;
71        } else {
72            temperature = temperature_humidity_sensor.ReadTemperature(CELCIUS);
73            humidity = temperature_humidity_sensor.ReadHumidity();
74            break;
75        }
76    }
```
Set Up Devices on The Things Network
Set Up Devices on The Things Network
Amazon Web Services Integration

Thing ARN

A Thing Amazon Resource Name uniquely identifies this thing.


Type

- loesewan

4 Attributes

<table>
<thead>
<tr>
<th>Attribute key</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>app_id</td>
<td>iot-demo-aws</td>
</tr>
<tr>
<td>dev_eui</td>
<td>08F9F09EC668A7</td>
</tr>
<tr>
<td>dev_iai</td>
<td>iot-demo-aws-01</td>
</tr>
<tr>
<td>app_eui</td>
<td>70B3D57E00013F95</td>
</tr>
</tbody>
</table>
Amazon Web Services Integration - Database
Amazon Web Services Integration - Database
Act on Data

- Visualization
- Monitor variables and guide actions
Machine Learning Example

- Temperature data from the coffee farm at DeKUT
- Temperature influences susceptibility to fungal disease
- Current monitoring is manual
Machine Learning Example

- Gaussian Process regression to fill missing values
Gaussian Processes

- A Gaussian process (GP) is a distribution over the space of functions.
- This distribution is completely specified by a mean function $m(t)$ and a covariance function $k(t, t')$.
- $f(t) \sim \mathcal{GP}(m(t), k(t, t'))$
Given observations we would like to infer an interpolant.
Gaussian Processes cont.

- Posterior distribution of functions given observations
Gaussian Processes cont.

- Original temperature data collected at the DeKUT coffee farm between 11th and 22nd September 2018.
Gaussian Processes cont.

- Same data with time axis normalized in terms of days (first 7 days)
- Missing data from the 3rd and 4th day
Gaussian Processes cont.

- Gaussian process fit with radial basis function kernel
- Fit using GPy from SheffieldML
  (sheffieldml.github.io/GPy/).
Gaussian Processes cont.

- Gaussian process fit with periodic kernel
Conclusion

- Data acquisition is an important step in data science
- LoRa is ideal for IoT applications requiring low power and long range
- Rapid prototyping is achievable for proof-of-concept
- Finding the ideal use cases is important
Hands on work

- This repo describes the process of programming the Nucleo boards.
  https://github.com/ciiram/dsa-abuja-mbed-demo

- This repo reproduces the analysis of the coffee data using Gaussian processes.
  https://github.com/ciiram/dsa-abuja-demo
Thank You