IoT Applications

Peter Chapin

CIS-2730, Software Engineering Projects

Vermont State University

What is the “Internet of Things?”

- IoT applications commonly involve the following.
  - Embedded devices
    - Homogeneous. All devices are the same (e.g., a network of identical sensors)
    - Heterogeneous. The devices are different (e.g., sensors, actuators, communication)
  - Central data storage
    - Cloud storage
    - A base station (e.g., phone, laptop, or local server)
  - Data processing
    - Report generation
    - Machine learning for purposes of making predictions
  - Decision making
    - Inform policymakers
    - Control actuators in the environment

Embedded Systems

- Computer systems *embedded* in some other device
  - Automotive, appliances, medical equipment, etc., etc.
  - Devices one does not normally think of as being computers (i.e., not personal computers, servers, or even mobile phones)
- Extremely wide range of capabilities
  - Very small: processors with tiny memories, slow speeds, and either a very simple, specialized OS or even no OS at all. *Very low cost!*
  - Medium scale: “normal” computers, but minimally configured.
  - Powerful: a full-scale system with generous memory and significant processing power. *High cost!*

Our Device

- BeagleBone’s BeaglePlay
  - Quad-core ARM64 processor. It's not extremely fast, but it's not junk, either.
  - Running Debian Linux with most of the bells and whistles
  - Reasonable RAM and reasonable storage
  - Lots of interfacing technologies to facilitate experimentation and prototyping.
  - Flexible, powerful enough to do interesting things, relatively inexpensive (~ $100)

What Will We Do?

- Gather temperature and position data, ideally using an accurate temperature sensor and GPS.
  - Imagine a collection of these devices “in the wild” around an area (e.g., Vermont or perhaps the campus of a university), gathering temperature information.
  - The devices have intermittent Internet connectivity.
- For hardware simplicity, we will use *mock* hardware components: software simulations of physical devices.
  - More flexible during development
  - Introduces the question of how faithfully the mock devices simulate their real-world counterparts.

Data Handling

- We will store (mock) temperature readings and location data in a local database on the BeaglePlay.
  - This takes advantage of the significant capabilities of the BeaglePlay in storage and processing power (to run a full database server locally).
  - Alternatively, SQLite could be used.
- Intermittent Internet connectivity
  - When we connect to the Internet, we will upload all newly accumulated data to a cloud database.
  - Multiple sensor nodes (each with a unique ID) will send data to the same cloud database.

Timing

- The data acquisition program will run continuously (from system boot).
  - It will sample the (mock) sensors periodically, e.g., once every five minutes.
- The cloud upload program will be launched from `cron` once per hour.
  - If there is no Internet, the program will gracefully fail.
  - Otherwise, the program will upload all new readings since the last successful run.
  - Data is *not* removed from the local database.

Time Synchronization?

- The BeaglePlay does not track time while it is powered down.
  - However, if the real-time clock (RTC) battery is installed, it *will* track time while powered off.
- Usually, the BeaglePlay syncs time with the Internet on boot (or shortly after the network connects).
- Our application relies on correct time information.
  - This suggests that we install the RTC battery.
  - Or we must ensure the network is available whenever the system boots.

Measurement Units?

- We agree to the following.
  - Temperatures shall be in Celsius.
  - Times shall be in Coordinated Universal Time (UTC). This is the default for the BeaglePlay unless you override this by configuring a time zone.
  - Locations shall be in degrees of longitude and latitude using decimal values for fractional degrees. For example, (44.452015, -73.112681) are the (latitude, longitude) coordinates for VTSU Williston.

What Does the Cloud Know?

- The cloud database contains the following information.

(*position*, *date/time*, *temperature*, *system_id*)

- We can imagine that an outdoor monitoring station took these readings, and they reflect the air temperature at the given location and time.
- The *system_id* allows us to filter the data by station in case we become suspicious about the validity of a particular station's results.

What Will We Do?

- Proposal
  - Using temperature data gathered over time across a region, predict temperatures at locations not covered by the data (interpolation) and at times later than covered by the data (extrapolation).
  - We'll use a machine learning (ML) algorithm for this, which will be applied to the cloud data.
- In this application, there are no actuators.
  - The cloud does not command the devices.
  - Many IoT applications have such commands.
  - This is subject to change in the future.

Location of Processing

- In many applications, the processing is done in the cloud
  - More computational resources (for a price).
  - Proximity of the data.
- We will do the processing on local workstations (i.e., your personal computers).
  - The data set is not so large as to make downloading all of it prohibitive.
  - Processing the data will not be computationally intensive because of its small size.
  - However, the results might also not be very accurate.

Working Against Us

- Fragmentary Data
  - Only short time intervals have data
  - Many locations have limited data
  - It is common for there to be only one location with data for any particular time.
  - (Mock) measurement error, aka “fuzz.”
- Bad Data
  - The time stamps are particularly susceptible to being wrong (because of the lack of a working RTC on the BeaglePlay by default).
  - Bugs in the software implementing the mock devices.

Working For Us

- Simplistic Mock Devices
  - The nominal temperature variation is a perfect sinusoidal wave.
  - The mock GPS device never fails or “returns” the wrong location.
- Small Data Set
  - Processing is easier with a small data set. In a real application, data sets measured in 100s of GB (or more) would be common.
  - OTOH, our predictions will be inaccurate because the ML model won’t have much to go on.

IoT is Interdisciplinary

- Many disparate technologies are used.
  - Embedded devices and operating systems.
  - System programming (device drivers, hardware control).
  - Databases, including, in general, “big data” handling.
  - Cloud computing.
  - ML/AI and other advanced data processing to guide policy and decision-making.