Thoughts on Reliability in the Internet of Things

James Kempf, Jari Arkko, Neda Beheshti, Kiran Yedavalli
(firstname dot lastname at ericsson.com)
Ericsson Research

Abstract

This paper discusses the role of reliability in networks dedicated to the observation and management of objects in the physical world, the "Internet of Things". The Internet of Things differs from today's global Internet in a number of ways. For instance, the networks are typically unmanaged, many applications are safety critical, the link layers are optimized for low power usage, and most nodes have to be implementable in a lightweight manner. The authors draw on their experiences of designing and deploying such networks, and provide some architectural guidelines to deal with reliability issues from packet transmission to network lifetime and application behavior.

1. Introduction

Much of the focus in the area of "Internet of Things" has been directed to specific link layer designs, routing protocols, IP layer operation, and applications – often discussed in isolation from the other components. This paper focuses on a system level aspect, namely reliability, and what implications it has for each of the protocol stack layers. We use the term reliability in a broad manner, not merely to talk about transport protocol behavior but also about:

- Self-configuration and ability to withstand changing environmental conditions,
- Long-term usability,
- Application robustness in the face of uncertain information,
- Resistance to security problems,
- Overall system reliability.

The rest of this paper is organized as follows. Section 2 outlines the reliability requirements. Sections 3 through 6 look at the implications for link layer, transport layer, application layer, and overall system architecture design. Where possible, we try to provide guidelines that help provide a reliable system. Finally, Section 7 suggests some potential areas for further standardization.

2. Requirements

A network and application must not merely be able to pass information in a reliable manner. Most Internet of Things applications for buildings, factories, hospitals, or the power grid are long-term investments that must also be operable for a long time, perhaps decades in some cases. The networks are also typically unmanaged (at least in home, human, and transport applications). This implies that the network must be able to configure itself as environmental conditions or components in the network itself change.

Even when the network itself is operating reliably, the nature of the applications running on the network may imply unreliable operation. For instance, the ability to observe the physical world may be limited. Sensor technology may be imperfect, a bit error may appear, or the nature of the physical processes may involve some variation. As an example, a building humidity measurement system deployed by the authors has had a reliability rate of roughly one unexplained alarm per year in a small building – perhaps due to a bit error somewhere or a drop of condensation water. While such an error rate sounds small, it can be confusing for the end users, and become an unmanageable problem in configurations that are hundred or thousand times larger.

Many applications are also safety critical. For instance, a "building health" application tracks an expensive asset. Fire alarms and human health tracking applications have direct consequences for human life. As a result, these applications have higher availability requirements than traditional networking services.

One particular problem with these types of applications is that they may be invisible, and their failure is not easy to detect. For instance, does the system provide an alarm if it cannot reach a particular fire sensor? Again, these applications differ from traditional network applications in that there may not be direct end-user interaction, at least not until the system must act.

The applications must also be appropriately secured so that outside (or even inside) threats cannot compromise them.

3. Link Layer Considerations

For most Internet applications, reliability primarily requires reliability on the end host stacks and the use of TCP between the two corresponding applications. Retransmission compensates for dropped packets, and, with the exception of some wireless networks, packet drops due to flaws in the underlying physical and MAC layer substrate are rare in today's Internet. Most packet drops are due to congestion, and congestion drops within reason are well handled by retransmission. In wireless access networks for mobile Internet service, where handovers can lead to packet loss of a type statistically different from wired networks, mechanisms are in place for the mobility management protocols to compensate.

In sensor networks, the situation is not as clear cut. The physical links used in most sensor networks today are hosted in the ISM portions of the spectrum (primarily 900 MHz in the US and 2.5 and 5 GHz worldwide). The uses of these bands are completely unregulated and, as a result, it is very easy for interference to render the links unreliable. A neighbor's WLAN access point tuned to the same channel as your sensor network application can disable critical sensor data reporting. In addition, much networking gear for popular ISM band-based protocols such as WLAN assumes an office environment where people can easily become aware of problems and quickly fix them. For example, one of the authors was advised by a sensor network monitoring company not to use a WLAN bridge on a sensor network application because bridges often failed, requiring a reboot. In an office environment, this would not cause a problem, but for a sensor network, the sensor network monitoring company needs to make a phone call, or, worse, arrange for a technician to fix the problem. Both actions are expensive. While cellular hardware costs are still somewhat high for use directly on sensors, cellular network protocols such as 3G and LTE may be a better solution for sensor gateway uplinks. The cellular network operator can act as an intermediary and provide a service that takes care of insuring connectivity, perhaps through SLAs that specify varying degrees of reliability at differing price points (i.e. for x dollars you get one truck roll per month, etc.). Interference problems should be far less since cellular wireless links run in dedicated spectrum. Extensions to successful AAA protocols such as Radius and other work for use between sensor gateways and sensor monitoring applications, and between gateways and cellular network operator networks, may be necessary.

4. Transport Layer and Routing Considerations

Transport layer protocols also have an important role to play in ensuring reliability. Because of constrained code size and processing power, most sensor network programmers want to use UDP as the transport layer protocol. The sensor transmits a packet to the gateway then goes back to sleep. Since network transmission, especially for wireless sensor networks, is one of the largest consumers of power, this pattern results in significantly larger power saving than if the sensor were to use TCP, staying awake to process the acknowledgement. However, use of UDP without retransmission at the transport layer risks a significant decrease in reliability, despite the MAC layer retransmissions provided by protocols such as IEEE 802.15.4 (Zigbee). The MAC layer retransmissions are limited and even in moderately dense sensor networks retransmissions can cause congestion, so the packets could still be dropped in the MAC layer.

For reasons discussed above, links in sensor networks may be more unreliable than in the Internet, and so some form of reliable transport protocol that has the "fire and forget" power-saving property of UDP but the reliability of TCP is needed. One way to accommodate this is to take it into account in the overall

system architecture design. For instance, UDP-based request/response protocol exchanges may be fine, if a central server component is in charge of remembering the reliability state and initiates all transactions. However, this is easily possible only if a polling model is suitable for the application in question; there are some applications where this arrangement is not practical.

The reliability requirements of the implemented transport protocol depend on the sensor network application. For example, different mechanisms could be implemented for packet-based applications (which require all packets be received reliably at the destination) versus event-based applications (which require events, and not necessarily all individual packets, be reliably reported to the destination). An event-based application might call for less-complex transport mechanisms.

Another concern relates to potential bandwidth/capability mismatch between a small sensor/actuator and devices that need information from it. For instance, a small temperature sensor may be unable to respond to a large number of queries, may employ simple protocols (UDP/CORE), and may exist in an address domain not reachable from all Internet hosts (private IPv4 address or in IPv6 address space). This type of a mismatch can be solved by introducing a gateway component in the network. The gateway is likely to have more power, computational capabilities, and high bandwidth Internet access to respond to all queries, replicate information streams to all potential users of the information from the sensor, and so on. Upstream and downstream routing mechanisms could also impose different requirements, as the former is usually unicast, while in many scenarios, the latter is multicast or broadcast.

Reliable routing in sensor networks appends the constraint of energy-efficiency to the traditional problem of reliable routing in fixed and mobile networks. There have been several proposals on routing protocols for sensor networks, which suggest mechanisms like replacing the hop-count measure of traditional routing algorithms by some energy-cost measures in sensor networks. In sensor networks, dedicated simple mechanisms need to be developed and implemented in order to reconstruct new paths when established paths break.

5. Application Considerations

It is hard to provide specific guidance for all possible applications. However, in our experience the following guidelines appear to be important for many of the applications that the authors have worked with:

- The application needs to provide an indication of its correct operational state. This helps provide observability of problems that would otherwise go undetected. Fire alarm systems need to indicate if they have reachability to all the fire sensors and alarm horns, battery power to continue operation in case of power outage, and so on.
- In many applications there are large numbers of sensors, some of which may become faulty over time, or the measurements may be imprecise. The system needs to be designed with usability in mind, so that end users can perceive what the real situation is, filter out unnecessary details and repetitive information, and shut down or replace incorrectly operating parts of the system.

6. System Architecture Considerations

In general, the architecture should employ different components for different tasks. A small sensor should be as simple as possible, whereas higher-layer processing, CPU, bandwidth, or storage intensive tasks should be delegated to other components in the network. For instance, a server component can more easily store historical data, serve a large number of users of that information, and convert information to other formats.

Similarly, in order to ensure that the system retains its usefulness over a long period of time, it makes sense to break functionality into different components. For instance, a simple sensor and actuator network may communicate with a server which in turn may communicate with a user interface (e.g., an alarm system). It is desirable that the sensors have a very long lifetime, particularly when they are embedded in the environment. Technical evolution in the user interfaces may be much faster, however. What may today

exist as a local alarm siren or an SMS facility may tomorrow be a browser component in a phone, and day after tomorrow something completely different.

7. Standardization

Based on the above discussion, it is our belief that standardization efforts should be made in the following areas:

- Reliability considerations during system architecture
- Reliability considerations during system deployment
- Support for sensor gateway communication with mobile networks
- Transport layer reliability.

8. Conclusion

In this paper we identified the reliability requirements for wireless sensor networks used in many high value and highly critical applications. We discussed the reliability problems present in link layer, transport layer, and application layer aspects of the sensor networks and propose possible solutions. We argue that system and data reliability should be an important design consideration corresponding to the sensor network application during its system architecture and deployment phase itself. As a conclusion we identify that system architecture reliability, system deployment reliability, and transport layer reliability are important areas for standardization for IETF.