The Internet of Things and Big Data Systems: The International Bazaar

Phillip A. Laplante
Pennsylvania State University
plaplante@psu.edu

Abstract - The Internet of Things will comprise many millions of systems, some being big data. The aggregated information from these systems represent, really big data systems. The problems arising from so many devices, data and processing coming together are likened to an international bazaar, with similar challenges. These big data problems in the IoT are reviewed from a reliability engineering perspective.

Keywords - Big Data, Internet of Things, Reliability

1. INTRODUCTION

The Internet of Things (IoT) is a collective noun for any system consisting of sensors, actuators, computational elements and other devices communicating locally and across the Internet. Many exciting new applications for this technology are envisioned in industry, consumer goods, healthcare, transportation and more. To date, however, very few large-scale, practical systems have been deployed. Perhaps this is because there are a number of reliability problems that have yet to be resolved.

IoT systems can purposely or inadvertently connect, creating, second order effects and unintended interactions. We can think of these problems in IoT ecosystems as similar to those in international bazaar, where planned and unplanned interactions occur, where goods (data) and services (processing) are of unknown quality, and where vendors (data producers) and customers (consumers) can have good or bad intent.

2. IOT – REALLY, REALLY, BIG DATA

Forbes magazine compiled a summary of IoT growth forecasts from several prominent industry groups. The consensus is that IoT growth, in terms of number of deployed ecosystems, number of sensors, and diversity of applications will be explosive. For example, a World Economic Forum survey concluded that by 2025 it is very likely that 1 trillion sensors will be connected to the Internet and that IoT systems will be found in more than half of all homes, connecting enabled appliances, clothing, and even reading glasses [1].

Similarly, Gartner forecasted that 6.4 billion connected things will be in use worldwide in 2016, reaching 20.8 billion by 2020. The International Data Corporation (IDC) predicted that by 2018, there will be 22 billion installed IoT devices and the worldwide wearable device market will reach a total of 111.1 million devices in 2016, with 214.6 million by 2019 [1].

The aggregation of data from a large number of IoT ecosystems, can lead to large data sets for analytic purposes. Consider, for example, the ensemble data from 300 million automobile IoT ecosystems, or 300 million household IoT ecosystems, or the composition of both. Furthermore, IoT applications could connect (deliberately or accidentally) to one or more big data systems outside the ecosystem, thus creating an aggregate of big data system orders of magnitude larger than any of the constituents. In this sense, every IoT system, even a small, local IoT ecosystem, is a potential big data system.

3. ISSUES FOR RELIABILITY ENGINEERS

There are numerous reliability challenges to deploying practical, large scale IoT systems. These challenges include, communications problems (e.g. lost signals, noise), fault-tolerance (e.g. sensor failure) and securing the network. But let's focus on the reliability issues specific to big data IoT systems. In particular, with respect to the data there are three fundamental challenges:

1. Authentication
2. Security
3. Uncertainty

All three challenges relate to the notion of trust, which is a very important principle in the international bazaar.

In the bazaar it helps to know something about vendors and customers. Knowing who you are negotiating with gives clues about motivation and helps to reach consensus. When a negotiating partner disguises their identity it is usually with militant. In the IoT authentication means confirming that the producer and consumer are who they claim to be. For example, data can be misidentified as to its origin (e.g. wrong sensor location) or it can be spoofed by a bad guy. Sending sensitive data to an unauthorized consumer is also problematic. Having authentic data is important for the integrity of the local IoT ecosystem decision making and for the ensemble data analysis across many related IoT systems. Data can be paired with code bits to foster authentication and analytics can be used to indirectly authenticate data. Complex interaction rules can also be used to improve this quality.

The safety of goods being exchanged in the bazaar is clearly important. For example, no one wants to buy poisoned food or clay pots that will explode. Likewise, big data security in the IoT means that the data that is being processed is uncompromised by attackers, and that the system is not leaking information or admitting unwanted information from adversaries. Security is a real problem in an IoT system -- IDC predicts that by 2018 66% of networks will have an IoT security breach [1]. The unfortunate consequences of adversaries inserting corrupted data or leaking sensitive data is the subject of many sensational news stories (e.g. hacking a vice president's pacemaker). IoT security for devices, communications and data is one of the most active research areas.

Finally, uncertainty in the bazaar refers to the quality of the goods and services. For example, customers want to know if the food is fresh. In the IoT uncertainty refers to the problem of data usability. Consider a sensor that is producing data at some discrete interval and sharing it within the IoT. The sensor could malfunction during one or more of the time intervals. The data could be corrupted or lost due to noise or as a defect in the communications mechanism. If we know the data is corrupted somehow, should the data (or its absence) be ignored or zeroed out? What information, if any can be deduced from the missing data? How does the missing information affect the actions prescribed by the IoT decision making algorithms or by offline analytics? Many mathematical frameworks are available for handling uncertainty, for example, expert systems, fuzzy theory, neural networks, possibility theory, probabilistic reasoning, neural networks and rough sets. But the proper selection of the correct approach(es) to handling uncertain information is an important one.

4. THE ROLE OF STANDARDIZATION

Returning to the metaphor of an IoT as an international bazaar, we have many different languages (protocols) spoken, different currencies (data formats) being used and different bargaining rules (algorithms) being applied. To smooth over these differences in a real IoT, extensive standardization will be required.

The IEEE and the Reliability Society are at the forefront of standardization efforts for the IoT. The first IEEE Big Data Initiative Standards Workshop was held in collaboration with the IEEE Reliability Society's International Symposium on Software Reliability Engineering (ISSRE) conference in November of 2015. Workshop participants identified more than a dozen new big data standards needed included those for Mobile Computing, Wireless and Analytics, Big Data for 5G Networks, Mobile Cloud Computing and Wireless Sensor Networks [2].

5. FINAL OBSERVATIONS

Many IoT applications will be complex, big data systems, both planned and inadvertent, and plenty of data for offline analytics. The possibilities are exciting but the challenges for trustworthy data alone are great. But whatever set of solutions emerge the IEEE Reliability Society will be helping to set the rules and observe the proceedings in the IoT bazaar.

REFERENCES

[1] Gil Press, “Internet of Things (IoT) Predictions from Forrester, Machina Research, WEF, Gartner, IDC”, January 16, Forbes.com, 2016 http://www forbes.com/sites/gilpress/2016/01/27/internet-of-things-iot-predictions-from-forrester-machina-research-wef-gartner-idc/#4b1601546be6

[2] IEEE Big Data Standards, http://bigdata.ieee.org/standards, accessed 3/23/16

AUTHOR BIOGRAPHY

Phillip A. Laplante is Professor of Software Engineering at The Pennsylvania State University. He received his B.S., M.Eng., and Ph.D. from Stevens Institute of Technology and an MBA from the University of Colorado. He is a Fellow of the IEEE and SPIE and has won international awards for his teaching, research and service. Since 2010 he has led the effort to develop a national licensing exam for software engineers. He has worked in avionics, CAD, and software testing systems and he has published 27 books and more than 200 scholarly papers. He is a licensed professional engineer in the Commonwealth of Pennsylvania and a Certified Software Development Professional. He is also a frequent technology advisor to senior executives, investors, entrepreneurs and attorneys. His research interests are in software testing, requirements engineering and software quality and management. Prior to his appointment at Penn State he was a software development professional, technology executive, college president and entrepreneur. More information can be found at www.personal.psu.edu/pal11.