

The Internet of Things and Big Data

Murali Cheruvu
Indiana University
3209 E 10th St
Bloomington, Indiana 47408
mcheruvu@iu.edu

ABSTRACT

The Internet of Things, or IoT, is all about data from connected devices. Millions of consumer and industrial devices drive IoT growth and challenge with data volume and variety. Big Data analytics helps combing through these high volumes of complex IoT data into meaningful business insights.

KEYWORDS

i523, hid306, Data Science, Internet of Things, IoT, Smart Devices, Sensors, Actuators, Big Data Analytics, Cloud Computing

Paper has too many - in it

1 INTRODUCTION

The Internet of Things (*IoT*) is the network of physical devices, vehicles, and other items embedded with electronics, software, sensors, actuators, and network connectivity which enable these objects to collect and exchange data[6]. Devices of all types - cars, thermostats, implants for radio-frequency identification (RFID), pacemakers and more - have become smarter, opening up the need for their connectivity with the internet. Today, over 50% of IoT activity is centered in manufacturing, transportation, smart environments and consumer applications like wearable gadgets, but within five years all industries will have rolled out IoT initiatives. *Gartner, Inc.* forecasts that 8.4 billion connected things will be in use worldwide by end of 2017 and will reach 20.4 billion by 2020[2].

2 IOT INTUITION

The rise of IoT changes everything by enabling *smart* things. Products and environments are becoming smarter. Broadly speaking, two kinds of IoT are emerging: *Consumer IoT* and *Industrial IoT*. Products such as Apple Watch, Fitbit, Smart TV, etc. are considered Consumer IoT. Examples of Industrial IoT are: manufacturing equipment and medical devices. A few more examples of IoT include:

2.1 Smartphones

With smartphone's range of sensors (accelerometer, gyro, video, proximity, compass, GPS, etc.) and connectivity options (cell, wi-fi, bluetooth, etc.) user has well equipped IoT device that can automatically monitor movements, location and workouts throughout the day.

2.2 Smart Homes

An example of smart home enabled by IoT devices: The user arrives home and his car communicates with the garage to open the door.

The thermostat is automatically adjusted to his preferred temperature due to sensing his proximity. He walks through his door as it unlocks in response to his smartphone or RFID implant. The home lighting is smartly turned on at dark.

2.3 Smart Cities

Smart surveillance, safer and automated transportation, smarter energy management systems and environmental monitoring are all examples of IoT applications for smart cities.

2.4 Smart Medical Alerts

The Proteus ingestible pill sensor is powered by contact with stomach fluid and communicates a signal that determines the timing of when patient took her medication and the identity of the pill. This information is transferred to a patch worn on the skin to be logged for the patient and her doctor's reference. Heart rate, body position and activity can also be detected accordingly.

2.5 Smart Aircrafts

Rolls-Royce is using Azure Cloud Stream Analytics and Power Business Intelligence (BI) to link up sensor data from its engines with more contextual information like air traffic control, route data, weather and fuel usage to get a complete report of the health of its aircraft engines.

3 NEED OF BIG DATA

The true value of IoT is not in the internet connected devices themselves; the value lies in making context-aware relevant data and converting the result into enterprise-grade, tangible and *actionable* business insights. The IoT and Big Data are intimately connected: billions of internet-connected things will, by definition, generate massive amounts of data. As the *things* turn more digital, IoT will analyze complex data structures and respond intelligently in real time.

Big Data, meanwhile, is characterized by *four Vs* - volume, variety, velocity and veracity[5]. That is, data come in large amounts (*volume*), with a combination of structured and unstructured data (*variety*), arrive at often real-time speed (*velocity*) and can be of uncertain source (*veracity*). Such information is unsuitable for processing using traditional SQL-queried relational database management systems (RDBMSs), which is why a cluster of alternative tools - notably Apache's open-source *Hadoop* distributed data processing system, various *NoSQL* databases and a range of business intelligence (*BI*) platforms - have evolved to serve such a complex data process.

4 IOT BUILDING BLOCKS

To scale the needs of IoT, the strategy should include infrastructure and applications that process and leverage machine and sensor data accordingly. At the moment, IoT platforms are often custom-built functional architecture. Enterprises that take the first step into this new market should look for interoperability between existing systems and a new IoT operating environment. The building blocks of an ideal IoT platform include:

4.1 Sensors and actuators

A major part of the IoT is not so much about smart things (devices), but about sensors and actuators. Smartphone would not have been smarter if it does not have an array of sensors embedded in it. A typical smartphone is equipped with five to nine sensors, depending on the model. *Sensors* measure physical inputs and transform them into raw data; *actuators* act on the signal from the sensors and convert it into output, which is then digitally storable for access and analysis. These tiny innovations can measure anything ranging from temperature, force, flow, position to even light intensity then can be attached to everything from smartphones to medical devices and then record & send data onto the cloud[3].

4.2 Network Connectivity

Network Connectivity in the devices is achieved through: wireless/wired, wi-fi, bluetooth, zigbee, VPN and cellular - 2G/3G/LTE/4G. Thread technology is emerging as an alternative for home automation applications and Whitespace TV technologies being implemented in major cities for wider area IoT-based use cases. Depending on the application, factors such as range, data requirements, security, power demands and battery life will dictate the choice of one or some form of combination of the technologies. In March 2015, the Internet Architecture Board - a group within the Internet Society that oversees the technical evolution of the internet - released a guide to IoT networking. This outlined four common communication models used by IoT smart devices: Device-to-Device, Device-to-Cloud, Device-to-Gateway, and Back-End Data-Sharing[4].

4.3 Collaboration and Security

Human and organizational behavior is critical in realizing the value of IoT approaches, and it is particularly important in shifting an organization to demonstrate clearly what will change, how it affects people, and what they stand to gain from IoT applications. Tons of collected IoT data could easily contain sensitive information about people and operations, and can even lose the control of critical systems. Beyond protecting personal privacy and business secrets, as more systems become automated, the risk of attacks becomes both more likely and more impactful.

Devices themselves should be secured, as should operating systems, networks and every other exposed piece of technology along the way. The roles of users, administrators and managers should be individually defined with appropriate access and strong authentication embedded in the design. A multi-layered approach to security is essential, and it should have checks and balances to reinforce protection and, if necessary, diagnose any breaches. For the IoT to work effectively, all the challenges around regulatory, legal, privacy

and cybersecurity must be addressed; there needs to be a framework within which devices and applications can exchange data securely over wired or wireless networks. To address these challenges and for better IoT interoperability, one key player, *OneM2M* issued Release 1, a set of 10 specifications covering requirements, architecture, Application Programming Interface (API) specifications, security solutions and mapping them to common industry protocols[1].

4.4 Cloud Computing

The cloud computing brings needed agility, scalability, storage, processing, global reach and reliability to an IoT platform. Flexible scalability can be achieved by using (a) Cloud Centric IoT - Good choice for low-cost things where data can easily be moved, with few ramifications (b) Edge Analytics - Ideal for things producing large volumes of data that are difficult, costly or sensitive to move, and (c) Distributed Mesh Computing - *Future-ready* multi-party devices automatically collaborate with privacy intact.

4.5 Big Data Analytics

Data Analytics involves statistical tools and techniques with business acumen to bring out hidden information from the data. Advanced types of data analytics include data mining, which involves sorting through large data sets to identify trends, patterns and relationships; predictive analytics, which seeks to predict customer behavior, equipment failures and other future events; and machine learning, an artificial intelligence technique that uses automated algorithms to churn through data sets more quickly than data scientists can do via conventional analytical modeling. Text mining provides a means of analyzing documents, emails and other text-based content. Big Data analytics applies data mining, predictive analytics and machine learning tools to volume of data coming from various sources with various types of data formats.

Big Data analytics, in the context of the IoT, refers to sensor analog inputs being converted to digital data, analyzed, and resulting in a response going back to the device. Much of this data is in an unstructured form, making it difficult to put into structured tables with rows and columns. To extract valuable information from this complex data, Big Data applications often rely on cutting edge analytics involving data science. Distributed computers in the cloud running sophisticated algorithms can help enhance the veracity of information by data mining through the noise created by the massive volume, variety, and velocity. Some analytics may need to be performed using edge or mesh computing, some in the data center and some in a cloud environment, depending on the trade-off of speed versus depth. IoT analytics applications can help companies understand the IoT data at their disposal, with an eye toward reducing maintenance costs, avoiding equipment failures and improving business operations.

5 CONCLUSION

Internet of Things shaping human life with greater connectivity and ultimate functionality, and all this is happening through ubiquitous networking to the Internet. There is seemingly no limit to what can be connected to the Internet. IoT will become more personal and predictive. The goal of a connected IoT ecosystem is to get the

most out of the internet of your things in your context. Industrial IoT side, it is becoming disruptive yet inevitable for companies to welcome it. Creating a connected IoT ecosystem that maximizes business value, collaboration is need with technologies, data, process, insight, action and people. The *T* of IoT is clearly important, but too often, it is the only area of focus when examining IoT in business. Rest of the systems need to be instrumented to leverage the data: communicating it to the right place for action - whether the cloud, data center, or edge - and then using analytics to understand data patterns and craft a response to fix or optimize. However, security and privacy will be the top considerations for companies developing IoT devices. Innovative organizations are starting to put this to use today.

REFERENCES

- [1] 2015. IoT Interoperability. (Jan. 2015). <http://www.onem2m.org/images/files/oneM2M-whitepaper-January-2015.pdf>
- [2] 2017. Garner Press Release. (Feb. 2017). <http://www.gartner.com/newsroom/id/3598917>
- [3] Hakim Cassimally Adrian McEwen. 2014. *Designing the Internet of Things*. Wiley.
- [4] Karen Rose, Scott Eldridge, and Lyman Chapin. 2015. *The Internet of Things: An Overview*. Technical Report. <https://www.internetsociety.org/resources/doc/2015/iot-overview>
- [5] Wikipedia. 2017. Big Data. (2017). https://en.wikipedia.org/wiki/Big_data [Online; accessed 23-Sept-2017].
- [6] Wikipedia. 2017. Internet of things. (2017). https://en.wikipedia.org/wiki/Internet_of_things [Online; accessed 23-Sept-2017].

6 BIBTEX ISSUES

- Warning-no key, author in 2Gartner
- Warning-no author, editor, organization, or key in 2Gartner
- Warning-to sort, need author or key in 2Gartner
- Warning-no key, author in 6OneM2M
- Warning-no author, editor, organization, or key in 6OneM2M
- Warning-to sort, need author or key in 6OneM2M
- Warning-no key, author in 2Gartner
- Warning-no key, author in 2Gartner
- Warning-no key, author in 6OneM2M
- Warning-no key, author in 6OneM2M
- Warning-no key, author in 6OneM2M
- Warning-no author, editor, organization, or key in 6OneM2M
- Warning-empty author in 6OneM2M
- Warning-no key, author in 2Gartner
- Warning-no author, editor, organization, or key in 2Gartner
- Warning-empty author in 2Gartner
- Warning-empty address in 4WileyBook
- Warning-empty institution in 5InternetSociety
- (There were 18 warnings)

7 ISSUES

DONE:

Example of done item: Once you fix an item, change TODO to DONE

7.1 Assignment Submission Issues

Do not make changes to your paper during grading, when your repository should be frozen.

7.2 Uncaught Bibliography Errors

Missing bibliography ffile generated by JabRef

Bibtex labels cannot have any spaces, _ or & in it

Citations in text showing as [?]: this means either your report.bib is not up-to-date or there is a spelling error in the label of the item you want to cite, either in report.bib or in report.tex

7.3 Formatting

Incorrect number of keywords or HID and i523 not included in the keywords

Other formatting issues

7.4 Writing Errors

Errors in title, e.g. capitalization

Spelling errors

Are you using *a* and *the* properly?

Do not use phrases such as *shown in the Figure below*. Instead, use *as shown in Figure 3*, when referring to the 3rd ffigure

Do not use the word *I* instead use *we* even if you are the sole author

Do not use the phrase *In this paper/report we show* instead use *We show*. It is not important if this is a paper or a report and does not need to be mentioned

If you want to say *and* do not use *&* but use the word *and*

Use a space after . , :

When using a section command, the section title is not written in all-caps as format does this for you

\section{Introduction} and NOT \section{INTRODUCTION}

7.5 Citation Issues and Plagiarism

It is your responsibility to make sure no plagiarism occurs. The instructions and resources were given in the class

Claims made without citations provided

Need to paraphrase long quotations (whole sentences or longer)

Need to quote directly cited material

7.6 Latex Errors

Erroneous use of quotation marks, i.e. use “quotes” , instead of ” ”

To emphasize a word, use *emphasize* and not “quote”

When using the characters & # % _ put a backslash before them so that they show up correctly

Pasting and copying from the Web often results in non-ASCII characters to be used in your text, please remove them and replace accordingly. This is the case for quotes, dashes and all the other special characters.

In case you copied a figure from another paper you need to ask for copyright permission. In case of a class paper, you must include a reference to the original in the caption

Remove any figure that is not referred to explicitly in the text (As shown in Figure ..)

Do not use `textwidth` as a parameter for `includegraphics`

Figures should be reasonably sized and often you just need to add `columnwidth`

e.g.

```
/includegraphics[width=\columnwidth]{images/myimage.pdf}
```

7.7 Structural Issues

Acknowledgement section missing

Incorrect README file

In case of a class and if you do a multi-author paper, you need to add an appendix describing who did what in the paper

The paper has less than 2 pages of text, i.e. excluding images, tables and figures

The paper has more than 6 pages of text, i.e. excluding images, tables and figures

Do not artificially inflate your paper if you are below the page limit

7.8 Details about the Figures and Tables

Capitalization errors in referring to captions, e.g. Figure 1, Table 2

Do use *label* and *ref* to automatically create figure numbers

Wrong placement of figure caption. They should be on the bottom of the figure

Wrong placement of table caption. They should be on the top of the table

Images submitted incorrectly. They should be in native format, e.g. .graffle, .pptx, .png, .jpg

Do not submit eps images. Instead, convert them to PDF

The image files must be in a single directory named “images”

In case there is a powerpoint in the submission, the image must be exported as PDF

Make the figures large enough so we can read the details. If needed make the figure over two columns

Do not worry about the figure placement if they are at a different location than you think. Figures are allowed to float. For this class, you should place all figures at the end of the report.