UDC 004.62

CHALLENGES OF PROCESSING INTERNET OF THINGS DATA

Kurdecha V.V., Trokhymenko D.V.
Institute of Telecommunication Systems,
Igor Sikorsky Kyiv Polytechnic Institute, Ukraine
E-mail: theelico@gmail.com

ПРОБЛЕМИ ОБРОБКИ ДАНИХ В МЕРЕЖАХ «ІНТЕРНЕТУ РЕЧЕЙ»

Неоднорідність "речей" робить обсяг, швидкість та мінливість даних Інтернету речей такими, що створює значні труднощі існуючим інформаційним системам. Динамічність та обмежені ресурси створюють деякі проблеми, які слід вирішити для ефективного оброблення даних реального світу. В роботі було досліджено основні проблеми в обробці даних Інтернету речей.

Identifying challenges. The Internet of Things core idea can be summarized in one sentence: ‘A worldwide network of interconnected entities’. In most cases, these entities, ‘things’ have a locatable, addressable, and readable counterpart on the Internet. They can open a communication channel with any other entity, providing and receiving services at any time, any place, and in any way. Many technologies serve as the building blocks of this paradigm, such as wireless sensor networks, cloud services, machine-to-machine interfaces, and so on. Also, this paradigm has a multitude of application domains, such as automotive, healthcare, logistics, environmental monitoring, and many others. And this heterogenous and dynamic nature is what makes developing efficient processing system a challenge. The main difficulties encountered in implementing stable and cost-effective IoT solutions can be divided into several groups: structure of IoT data, determining the proper frequency of sensor readings, accounting for errors and missing readings, security and privacy.

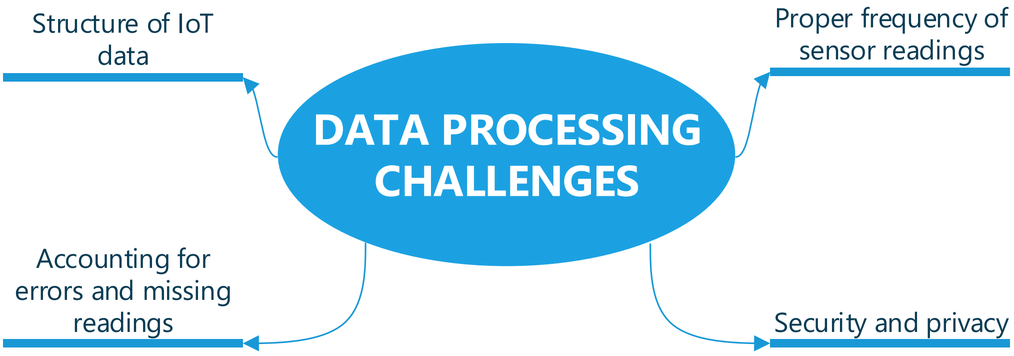
Structure of IoT data
Proper frequency of sensor readings
DATA PROCESSING CHALLENGES
Accounting for errors and missing readings
Security and privacy

Fig. 1. IoT data processing challenges

Deceptive simplicity of IoT data. Real world data is transient, subject to environment changes and it is mostly time and location dependent. In many currently used data sources it is often necessary to identify what information was available, how it was formatted. Ironically, this is one area where IoT sensor data can seem deceptively simple compared to these sources. Usually most sensors provide simple-formatted data. The good news is that this makes ingesting raw sensor data fairly straight forward to quickly go from a raw feed to a dataset or table that is ready for further work. Main problem arises after the ingestion of the raw data, due to it being heterogenous, most importantly in fields of how this data should be processed and stored.

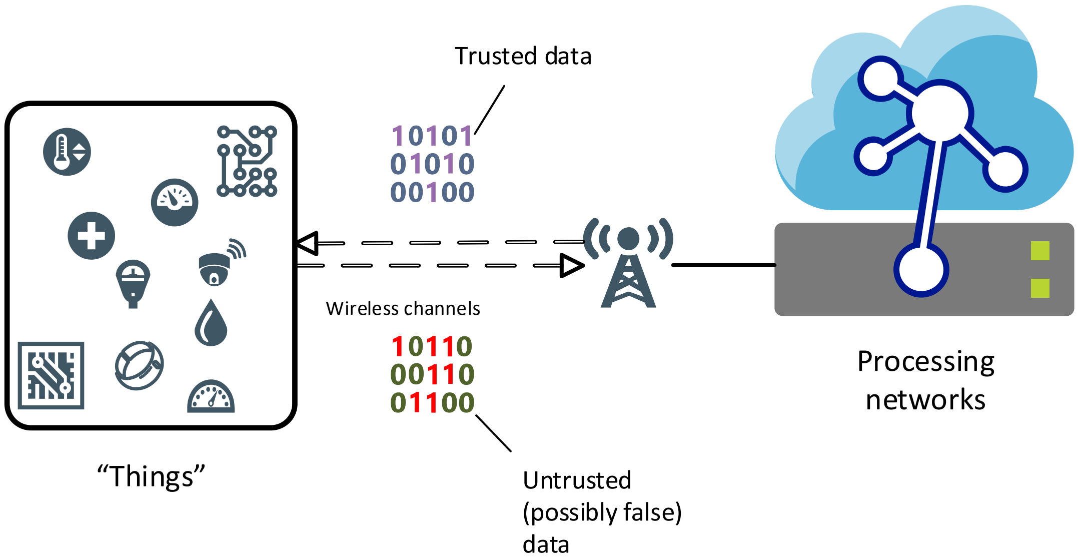

Fig.2 Sample IoT network layout

Determining the proper frequency of sensor readings. With large volumes of data, efficiency of storage and data handling mechanisms become a key challenge, especially considering heterogenous nature and dynamicity of the data sources. Depending on the application, a streamed sensory data could be stored either temporarily or for greater spans of time. So, designing and implementing repositories that enable publishing and accessing the needed data in large distributed and dynamic environments while also providing efficient indexing and discovery mechanism is important issue for IoT. More efficient mechanism on information search and retrieval, indexing, query, and information access will be required to address the issues in this field. However, the solutions for handling, maintaining, and processing the data also need to be scalable and efficient. As a result, due to these limitations it is necessary to determine what data should be gathered and actually has value for the application it is used in. Cloud and fog computing clearly are promising technical approaches that address some of these challenges. It requires the assess each metric to prevent acquiring useless and acquire useful data.

Accounting for errors and missing readings. Inaccuracy and varying qualities in the IoT data are unavoidable. If the next readings are normal, it is easy to correct the error. But, what if a sensor gets moved into an improper position or malfunctions or fails to transmit at all? Detecting and filtering anomalies and false readings from the devices, could be done with quality related attributes of the IoT data, that can help detect errors, and retrieve and process the data according to different quality requirements. But current methods may suffer from over-relying on assumptions about data. Using correlation can often impose non-existent correlations between sensors that leads to worse estimation as a result. In general, on one hand, the exactness of assumption-based models directly affects the accuracy of prediction results. On the other hand, such assumptions may not hold for various datasets. So, it is necessary to find a new way to learn structures from sensor data that doesn't rely on such knowledge. Also, when data is provided by different resources, such as public sensors, trust arises as another key issue. Trustworthiness of resources, identification of the source providing the data, and an understanding of accuracy and reliability of the data, while semantics can play an important role for defining trust and reliability attributes, trust model development and its feedback and verification mechanisms are major issues that need to be addressed.

Security and privacy. IoT data can often contain our environment, the status of our homes and cities, or our personal health and activities. That's why mechanisms to provide and guarantee the security and privacy of data are crucial issues in IoT. However protecting the Internet of Things is a complex and difficult task. The number of attack vectors available, such as attacks that target communication channels, physical threats, denial of service, identity fabrication, is staggering as global connectivity and accessibility are key points of the IoT. And, the inherent complexity of the IoT, where entities can easily exchange information with each other, further complicates the design and deployment of efficient, inter-operable and scalable security mechanisms. As data is communicated over the Internet and can be shared with different parties and users, it is also important to define appropriate access control mechanisms, e.g., who can use the data, what part of the data they can use, when and where they can use the data. Further development of IoT will also be highly dependent on developing reliable and efficient solutions that can support and maintain security and privacy requirements in resource-constrained environments with various types of devices and communication networks as a part of the IoT design.

Conclusion. To sum everything up, in a near future the creation of a system that will address all of the issues is rather unlikely. Because humanity still needs cheap, fast and reliable data gathering and processing systems. So, for now we have to have tradeoffs, like in choosing sensors and data ingestions method to save either storage space or prolong sensor lifetime. And for determining which tradeoffs should be taken due to nature of IoT data many factors should be taken into consideration considering, but not limited to type, size, location or other.

References

1. Barnaghi, P., Wang, W., Henson, C., & Taylor, K. Semantics for the Internet of Things: early progress and back to the future. *International Journal on Semantic Web and Information Systems (IJSWIS)*, *8*(1), 1-21. (2012)
2. Shao, Yongshuai, and Zhe Chen. "Reconstruction of Missing Big Sensor Data." *arXiv preprint arXiv:1705.01402* (2017).
3. Roman, Rodrigo, Jianying Zhou, and Javier Lopez. "On the features and challenges of security and privacy in distributed internet of things." *Computer Networks* 57.10 (2013): 2266-2279.