2017 13th International Conference on Distributed Computing in Sensor Systems

Big Sensed Data challenges in the Internet of Things

Hossam S. Hassanein and Sharief M. A. Oteafy
School of Computing
Queen's University
Kingston, Ontario, Canada
{ hossam | oteafy }@cs.queensu.ca

***Abstract*— Internet of Things (IoT) systems are inherently built on data gathered from heterogeneous sources. In the quest to gather more data for better analytics, many IoT systems are instigating significant challenges. First, the sheer volume and velocity of data generated by IoT systems are burdening our networking infrastructure, especially at the edge. The mobility and intermittent connectivity of edge IoT nodes are further hampering real-time access and reporting of IoT data. As we attempt to synergize IoT systems to leverage resource discovery and remedy some of these challenges, the rising challenges of Quality of Information (QoI) and Quality of Resource (QoR) calibration, render many IoT interoperability attempts far-fetched. We survey a number of challenges in realizing IoT interoperability, and advocate for a uniform view of data management in IoT systems. We delve into three planes that encompass Big Sensed Data (BSD) research directions, presenting a building block for future research efforts in IoT data management.**

Keywords—Internet of Things; Big Sensed Data; Next Generation Networks; Quality of Data; Quality of Information

I. INTRODUCTION

The Internet of Things (IoT) is proliferating on reliable and scalable collection of (mostly sensed) data. Meanwhile, the growing domain of sensing over smart devices and smart vehicles, are all generating an exponentially increasing amount of data. The ensuing advent of Big Sensed Data (BSD) is generating a number of critical challenges. First, collected data is mainly insightful to each deployed network, any "sense-making" processes to be built upon heterogeneously collected data faces significant interoperability problems, exposing challenges with varying quality, data-labelling inconsistencies, inaccuracies, time-sensitivities and different reporting granularities. Second, sensing systems inherently adopt a collect-and-report model, whereby collected data is indiscriminately pushed onto the networking infrastructure, regardless of the Quality of Information (QoI) or its value (VoI).

Not only do we face scalability issues, but establishing reliable IoT Services on top of BSD is not attainable over inconsistently collected, validated and reported data. Thus, the future of Big Data is hampered by the sheer volume of reported data, its uncalibrated discrepancies, and worse by the flood of redundant and lower quality data. Real-time decision making is inherently built on the efficacy of ubiquitous sensing systems, not on the aggregation of devices that are isolated in operation and management. In a time when important IoT applications such as real-time road monitoring, health Informatics and emergency services require rapid and scalable access to contextual information about patients, mobile crowds and the general public, the status quo falls significantly short.

In Figure 1 we present three planes that encompass current developments in IoT data management, namely Resources, Data and Information planes. Specifically, resources in the bottom plane are probed via a number of access schemes (e.g. NFC for very short range communication, Bluetooth and ZigBee for short range and WiFi for long-range), which then provide crude data to the Data plane. At this stage, hierarchical fusion takes place in light of reported resource attributes, and across different data repositories that collect data from individual sensing systems and heterogeneous data sources. Finally, a more broad context-based analysis of Quality of Data and Quality of Information will aid in fine-tuned data fusion to feed into the Information plane. We envision that services will run on top of the Information plane having mandated specific thresholds for requested QoI indicators. We emphasize the intermediate infrastructures required to communicate data (access networks), and carry out data fusion and pruning.

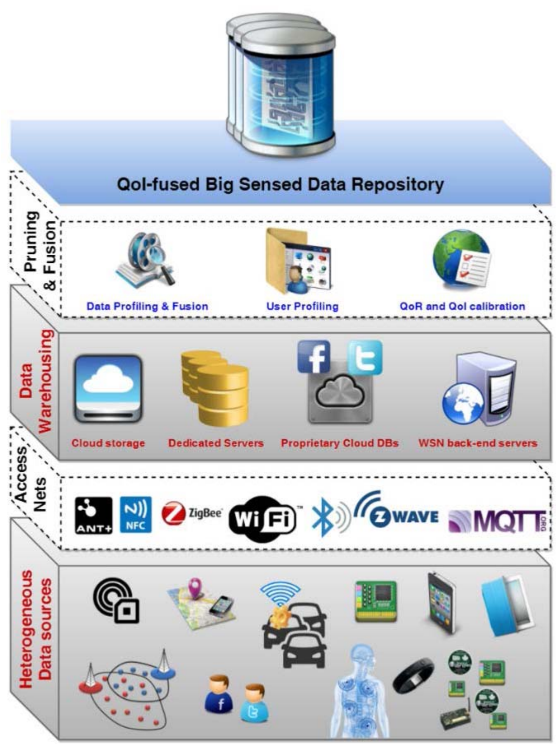

Fig. 1. Big Sensed Data planes for managing IoT data

2325-2944/17 $31.00 © 2017 IEEE
DOI 10.1109/DCOSS.2017.35

II. Challenges in BSD management

Recent efforts have been directed at understanding QoI and QoR in IoT systems, attempting to encompass a number of metrics to calibrate QoI, QoR and QoData. In parallel efforts, gateway integration approaches have been presented to circumvent device interoperability.

A. Uniform QoR measurement

Data is generated and reported by sensing and communication resources from heterogeneous devices in the IoT. The interplay of "perceived" quality in the context of services requesting the data, versus actual QoR metrics that are established at the resources plane, is an important yet challenging research direction, with many voids to address. Ultimately, the goal of a ubiquitous BSD architecture would be to aid long-term sensor deployment and soliciting efforts to reduce the footprint of new sense-making applications. That is, if an Information Service requires a pre-determined set of functional requirements, we must be able to dynamically assess whether the current region of interest has the resources required to perform the task, and if they can deliver the required QoD mandated by Service Level Agreements (SLAs).

B. Quality of Data (QoD) calibration

In a perfect world, all data would be obtained from a single global-scale homogenous network, and hence QoD calibration would be a (rather) straightforward problem. In fact, this is the advocacy of major players in the IoT (such as HP and IBM) as they each propose to include only their sensors in "intelligent" sense making processes. The scale of ubiquity in BSD, and its potential contributors, mandate a uniform framework for QoD calibration despite the heterogeneity of underlying architectures or resources. Moreover, in adopting QoD calibration and assessment, we must explore the impact on duty cycles and contributions of heterogeneous nodes, i.e. actively silence some nodes with low QoD output, or simply yield to pruning and fusion decisions. Specifically, this research direction must target a framework that will prune data as it is pushed forward towards data repositories, and aid pruning efforts to send feed-back meta-data to the resources plane to de-list or relief resources that are no longer contributing data that meets current QoD standards.

We envision that enlisting heterogeneous data sources is at the heart of improving information systems that provide services using such data. Instead of opting for conforming views of what a sensor (or data source) should be, and how to confine data sources to less proprietaries, we should aim to capitalize on the heterogeneity of these resources as an important tool for verification and validation in sensing systems, especially as they scale in functionality and coverage.

C. Convergence with Information Centric Networks

The end product of data in IoT is information; one that we hope to push on the Internet to enable further services. The rise of Information Centric Networks (ICNs) offer an excellent sink for data from IoT systems. We argue that BSD will eventually converge with ICNs in their operational mandates, not only as an underlying infrastructure. That is, a critical challenge is introducing standardized naming schemes for data, regardless of the data source, to enable hierarchical data fusion, and uniform calibration across the spectrum of reported events. Then, IoT services will no longer need to search for data-generating IoT systems, but merely depend on data survivability and abundance on the ICN infrastructure.

D. Ontology based data fusion

We must also engage parallel work in different literatures to aid the development of BSD frameworks. An untapped area of exploration lies in Ontology based data fusion and integration. This area has been heavily investigated in Web systems and distributed queries (and many other sciences) to categorize and describe entities, their interactions and pertinent information. Applying Ontology-based fusion in BSD holds great promise as it inherently addresses the diversity of heterogeneous data sources, and will elicit the variability and properties of data in relation to its reporting resource.

Biographies

Hossam Hassanein (S'86-M'90-SM'06-F'17) received the Ph.D. degree in computing science from the University of Alberta, Canada, in 1990. He is a leading authority in the areas of broadband, wireless and mobile networks architecture, protocols, control, and performance evaluation. His record spans more than 500 publications in journals, conferences, and book chapters, in addition to numerous keynotes and plenary talks at flagship venues. He is also the Founder and Director of the Telecommunications Research Lab at Queen's University, Canada, with extensive international academic and industrial collaborations. Dr. Hassanein is an IEEE Communications Society Distinguished Speaker and a past Chair of the IEEE Communication Society Technical Committee on AHSN. Dr. Hassanein is a Fellow of the IEEE and has received several recognitions and best papers awards.

Sharief Oteafy (S'08–M'13) is an Adjunct Assistant Professor at the School of Computing, Queen's University. His current research focuses on dynamic architectures for enabling large scale synergy in the Internet of Things; encompassing dynamic resource management across IoT platforms, in addition to managing the proliferation of Big Sensed Data. He is an active member of the IEEE ComSoc Standards Association and is currently the Ad Hoc and Sensor Networks standards Liaison, and a member in the IEEE Tactile Internet standard WG. Dr. Oteafy co-authored a book on "Dynamic Wireless Sensor Networks". He co-chaired a number of IEEE workshops, in conjunction with IEEE ICC and IEEE LCN conferences, and served on the TPC of numerous IEEE and ACM symposia. He delivered tutorials on Big Sensed Data management in IEEE ICC, IEEE CAMAD and IEEE Globecom conferences, and serves as an Associate Editor in IEEE Access.

References

[1] S. Oteafy and H. Hassanein, "Resilient IoT architectures over Dynamic Sensor Networks with adaptive components", IEEE Internet of Things Journal, 2016.

[2] S. Oteafy and H. Hassanein, "Organic Wireless Sensor Networks: A resilient paradigm for ubiquitous sensing", *ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems* (MSWiM), 21-25 September, 2014.