Horizon Scanning Series

The Internet of Things

Novel Data and Information Considerations with IoT

This input paper was prepared by David Eyers and Holger Regenbrecht

Suggested Citation

Eyers, D and Regenbrecht, H (2019). Novel Data and Information Considerations with IoT. Input paper for the Horizon Scanning Project "The Internet of Things" on behalf of the Australian Council of Learned Academies, www.acola.org.

Novel Data and Information Considerations with IoT

David Eyers and Holger Regenbrecht

The IoT is likely to produce vast amounts of data: its development will effect one of the first infrastructures needing directly to embrace 'big data' principles, in particular:

- **volume**: billions of devices directly and indirectly reachable; multiple levels and types of network infrastructure; significant amounts of storage, both centralised in the cloud, and decentralised within the IoT devices themselves;
- **velocity**: high bandwidth data transfer offered by widespread fibre-optic infrastructure (FTTP/FTTH), and radio-based communication (5G/6G[1]); and
- **veracity**: the accuracy, precision, trustworthiness and data provenance will be increasingly important, requiring adoption of security-by-design, and risk mitigation against cybercrime.

Beneficiaries of IoT and the smart cities that will become possible:

- computing professionals and technology companies developing and producing new IoT hardware and software solutions;
- telecommunications professionals developing the data communication networks that support the IoT hardware;
- data scientists working on ways to effect data collection, wrangling, and analysis; and
- city planners, commercial organisations and advertisers using location-based services.

Machine Learning (ML) and Artificial Intelligence (AI)

Initial emergence of the IoT focused on large numbers of fairly basic sensor / actuator devices and the networking infrastructure required to provide interconnection of such devices. After that emergence of IoT came the rise of (shallow) AI—more specifically ML processing sensor data, such as in computer vision applications. Convolutional neural networks (CNNs) are increasingly being developed as custom hardware components, e.g., within graphics processing units. Recently, it has become clear that the increased power efficiency of these components, and their miniaturisation, bring them within reach of IoT devices. Thus, we should expect to see a blossoming of decentralised AI, or AI at the edge. The training of models may still be too computationally expensive or reliant on large datasets to run on IoT devices, and in such cases sensor data is likely to travel into cloud/edge computing, and trained models be transferred back into the IoT devices, to allow them to operate autonomously using those models.

Data Provenance[2]

Present day data processing systems mostly apply *ad hoc* means to determine the source of data and the transformations that have been applied to it in transit. When personally-identifiable data is included, existing data processing approaches, and the generally poor degree of automation of accountability available, will increasingly run into regulatory challenges. For example, the European Union's General Data Protection Regulation (GDPR) empowers citizens, as owners of their data, to acquire intelligible records showing the source, destination and transformations applied to information about them. It is expected that the IoT will increase the amount of personally-identifiable data that flows, e.g., within a

[1] Koziol, M. (2018). Now's the time to think about what comes after 5G: We need to make sure the backbone of every network can support future demands for data-[Spectral Lines]. IEEE Spectrum, 55(12), 6-6.

[2] Pasquier, T., Singh, J., Powles, J., Eyers, D., Seltzer, M., & Bacon, J. (2018). Data provenance to audit compliance with privacy policy in the Internet of Things. Personal and Ubiquitous Computing, 22(2), 333-344.s

smart city's transport systems, so accountability-by-design[3] is an emerging priority. Data science techniques will be useful to quantify the likely completeness, precision, and accuracy of data records.

IoT Search Engines (IoTSE); service discovery; named-data networking[4]

Current internet services rely on largely centralised infrastructure, such as search engines run by Google, Microsoft, etc., to discover information. IoT data and service discovery may evolve to be significantly more decentralised: a notion of distributed IoT search engines may emerge. Such a transition might facilitate flexible interactions between IoT and edge devices, such that the primary providers and hosts of information become evenly spread through the internet. In terms of smart cities, government organisations would be able to participate more in information storage and processing.

A lower-level technology pushing in the direction of decentralised information management is named-data networking: the idea is that a description of the information a requestor seeks to find is sufficient to allow a peer-to-peer style coordination to route the requests to appropriate information sources.

Analytics on real-time data streams

When data volumes and velocities grow sufficiently high, it is necessary to apply analytics on data while it is in motion through the network, rather trying to store the data and query it later. Real-time analytics on data in motion—also distributed data stream processing (DSP)[5]—often employs the principle of moving computation and data transformation toward the data sources, rather than the more traditional approach of transferring data to computing infrastructure. Real-time IoT data analytics will lead to a research field in its own right with a plethora of dissemination and exploitation opportunities.

IoT Software (and Firmware) Security[6]

A significant threat to the global viability of IoT, but also a great opportunity for emerging technology is in software and firmware security of IoT devices. It has been sarcastically joked that "The 'S' in IoT stands for security and the 'P' for privacy." The security record of domestic IoT devices so far has been woefully bad: particularly within consumer products such as whitegoods, and internet-enabled toys.

Software update: Critical security vulnerabilities are discovered frequently in operating systems and device firmware, requiring software updates to be installed. However, IoT devices may be inaccessible to software suppliers after they are deployed, leaving them vulnerable. Alternatively, leaving devices open to software suppliers may itself compromise security, given that software update mechanisms are themselves liable to attack. Many IoT devices will be unattended, with no unified means for devices to seek confirmation from their owners about when to deploy updates.

Program correctness and formal verification: Given the difficulties in deploying and updating software and firmware on many IoT devices, there may be growing willingness to investigate mechanically verified software, such as the work done on the seL4[7] operating system microkernel and its derivatives.

---

[3] Crabtree, A., Lodge, T., Colley, J., Greenhalgh, C., & Mortier, R. (2016). Building Accountability into the Internet of Things. SSRN, doi: 10.2139/ssrn.2881876.

[4] Ahlgren, B., Dannewitz, C., Imbrenda, C., Kutscher, D., & Ohlman, B. (2012). A survey of information-centric networking. *IEEE Communications Magazine*, 50(7), 26-36. doi: 10.1109/MCOM.2012.6231276

[5] O'Keeffe, D., Salonidis, T., & Pietzuch, P. (2018). Frontier: Resilient edge processing for the internet of things. *Proceedings of the VLDB Endowment*, 11(10), 1178-1191.

[6] Concerns have led to legislation under review in the US: the IoT Cybersecurity Improvement Act of 2019 https://warner.senate.gov/public/index.cfm/pressreleases?id=88A88A37-AD5E-4C01-932D-A23684AAD7AE

[7] Klein, G., et al. (2009). seL4: formal verification of an OS kernel. In *Proceedings of the ACM SIGOPS 22nd Symposium on Operating Systems Principles*. doi: 10.1145/1629575.1629596

IoT Data Security and information integrity

In addition to software security, the security of data needs to be ensured, both when on the IoT devices, and in transit over networks. Data security and information integrity must be supported by a trusted computing base. Intel and ARM have both released technologies (SGX and TrustZone, respectively) that aim to support trusted computing through hardware-based security support in CPUs.

If IoT devices are designed not to be recovered, data should be deleted over time, or there is a risk that encryption mechanisms remain secure for the remaining lifetime of the IoT device. The latter approach is likely to be unsafe given evidence of the diminishing strength of encryption over time.

Interoperability

Over time, cloud technology has helped homogenise computing infrastructure (with notable exceptions emerging in GPGPU, FPGA and ML accelerated cloud infrastructure). IoT will buck that trend, forcing significant heterogeneity back into deployed devices—a consequence of the many and varied roles and deployment environments of the 'things'. Significant effort will be needed to support the interoperation within the IoT ecosystem at many levels of technology.[8]

At the fundamental technical level, the next version of the internet protocol for sending data packets, IPv6, was touted as a means to allow all IoT devices to be addressable and reachable. It is increasingly clear that interconnection will not provide satisfactory interoperability, and indeed the rate of adoption of IPv6 has been very slow. Privacy concerns that have emerged mean a given IoT device will be unlikely to maintain a single IPv6 address, challenging interconnection by ensuring another layer of dynamic churn that avoids devices being tracked by malicious parties.

Middleware is likely to see a resurgence—middleware is software that exists to assist the data exchange between other software layers, providing a separation of concerns and economies of scale for the stakeholders in the software layers being interconnected. Increasingly, interoperation will be needed at human semantic levels[9] rather than pure technical levels, ensuring that consistency can be checked at more human-relevant levels of interaction, relating to knowledge derived from raw data within the IoT.

The economic benefits from better interoperability will be significant, and would provide a counterpoint to today's IoT technology, which typically only interoperates well within a silo from each manufacturer.

Public perceptions of IoT and related technologies

For successful adoption, the IoT will need to manage public perception, and educate users about fundamental aspects of the technology. In our present "post-truth, anti-science" political world, this may require some creative solutions. IoT technologies can produce rich dashboards depicting environmental conditions. Aside all the other beneficial use cases, such dashboards might provide government or other trusted organisations a means that helps quantify to concerned citizens the likely lack of impact from 5G[10] radio towers, for example.

---

[8] Desai, P., Sheth, A., & Anantharam, P. (2015). Semantic Gateway as a Service Architecture for IoT Interoperability. IEEE International Conference on Mobile Services, New York, NY, 2015, pp. 313-319.

[9] Hert, P.D., et al. (2018). The right to data portability in the GDPR: Towards user-centric interoperability of digital services. *Computer Law & Security Review*, 34(2), pp 193-203. doi: 10.1016/j.clsr.2017.10.003.

[10] For instance, the introduction of 5G in Switzerland is delayed because of massive concerns of citizens and groups: https://www.tagesanzeiger.ch/sonntagszeitung/standard/5G-verspaetet-sich-wegen-Einsprachen-und-Moratorien/story/19753263