

End-to-End QoS Specification and Monitoring in the Internet of Things

Awatif Alqahtani, Ellis Solaiman, Rajkumar Buyya¹, Rajiv Ranjan

School of Computing Science, Newcastle University, United Kingdom

¹Department of Xomputing and Information Systems, University of Melbourne, Australia

The Internet of Things (IoT) is a new computing paradigm in which uniquely addressable objects such as Radio-Frequency IDentification (RFID) tags, sensors, actuators, and mobile phones, become part of the Internet environment, and cooperate with each other in order along with other resources to achieve common goals [1][4]. This paradigm opens the doors to new innovations that will build novel type of interactions among things and humans, and enables the realization of smart cities, infrastructures, and services for enhancing the quality of life and utilization of resources [5]. The number of connected smart objects is estimated to reach 212 billion by the end of 2020 [2, 3]. Such large numbers of connected smart objects will generate huge volumes of data that needs to be analysed and stored [6]. According to a study conducted by IBM, 2.5 quintillion bytes of data is generated every day [7]. Storing and processing such large volumes of Big Data is non-trivial, and requires the flexibility offered by Cloud computing [8]. Cloud computing offers a pool of configurable resources (hardware/software) that are available on demand [9], allowing users to submit jobs to service providers on the basis of pay-per-use. While the IoT provides smart devices with the ability to sense and generate large amount of data that reflect the physical world in different forms and different data speeds, Cloud Computing offers advanced technologies for ingesting, analysing and storing data [10]. The number of applications based on IoT and Cloud Computing is projected to increase rapidly over next few years. To this end, Gascon and Asin [11] predicted that in near future there will be about 54 types of IoT applications for addressing different domain [5] specific problems: security and emergency, smart environment, smart cities, smart metering, smart water, smart animal farming, smart agriculture, industrial control, retail, logistics, domestic and home automation and e-Health.

Quality of Service roadmap for IoT applications

Users expectation from services provided by the Internet of Things are no different from most traditional computer and Internet based services in that they must be delivered with guaranteed levels of quality of service (QoS). For example, in emergency response (ER) IoT application, there is a need to receive and analyse data from deployed sensors immediately and accurately in order to allow for timely response to potential damage that can be caused in natural disaster situations such as earthquakes, floods, and tsunamis [12]. Such IoT applications can be extremely time sensitive and any delay in the collection/transferring/ingestion/analysis of sensor data may have disastrous consequences.

As we note in our previous papers [4][13], engineering IoT applications that can guarantee QoS is a challenging and not feasible with the current state-of-the-art solutions available in context of IoT programming models (e.g. Amazon IoT, Google Cloud Dataflow, IBM Quark) and resource management methods [4][33]. An important difficulty is that IoT application eco-systems are typically consists of several layers involving multiple, heterogeneous hardware and software resources; and data types from digital and human sensors. An example of an IoT eco-system is depicted in Figure 1, which consists of following programming and resource management layers; sensing layer, gateway layer, network layer, and cloud layer. Providing customers with QoS guarantees requires the technical ability to ensure that their QoS requirements will be observed across each of the layers of an IoT application eco-system.

Specification of IoT application specific QoS requirements within Service Level Agreements

In purely business context, QoS requirements are formally specified in a Service Level Agreement (SLA) document [3] which serves as the basis of legal agreement and understanding of service terms, conditions, and commitments between consumers and providers. For example, Amazon Web Services' SLA document stating the terms, conditions, and commitments for its S3 and EC2 services can be found at [34] and [35] respectively.

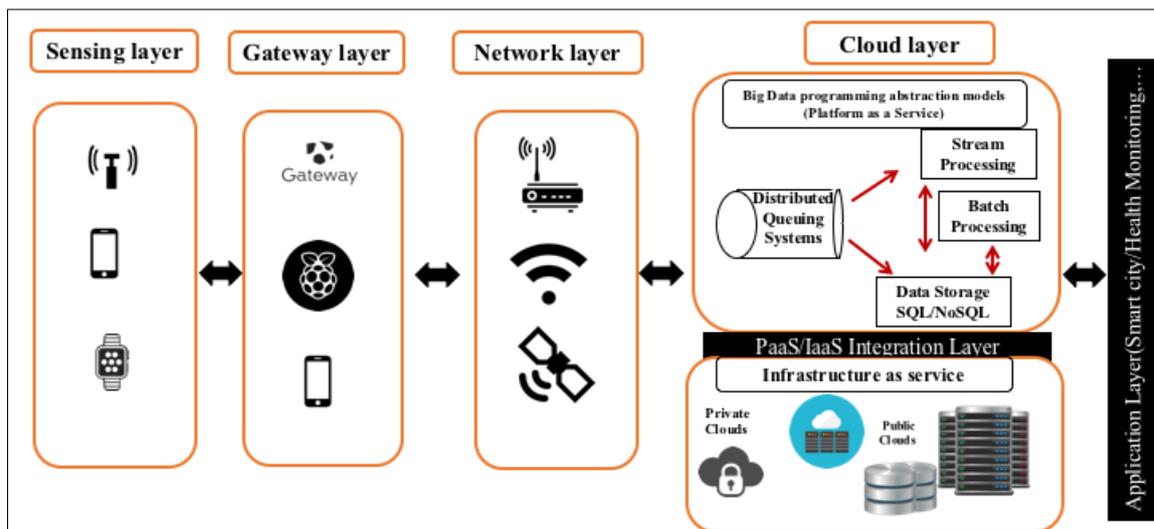


Figure 1: A multi-layered architecture IoT application eco-system involving Sensing, Gateway, Network, and Cloud layers.

As IoT applications have layered architecture and complex Big Data flows across layers, there is a need to first model SLA for individual layers followed by their holistic aggregation. Such aggregated SLA document (template) will form basis for specifying an end-to-end SLA that can be used to specify the service terms, conditions and commitments for an IoT application. Notably, cross-layer SLAs in IoT have a strong dependency relationships with each of its upstream and downstream layers, regardless of whether this component is data, computing hardware, IoT sensor, software, or human. Thus violation of one or more constraints by one or more components (s) affects the adherence to the related SLA's terms.

To illustrate this concept, consider a remote health monitoring IoT application [13] where patients wear sensors and accelerometers to measure their heart rate and sugar levels, reminding them of the time to take medications, and detecting abnormal activities such as falling down. Subscribed patients might ask for a service that can satisfy the following high-level, strict SLA requirement: detecting abnormal activity, such as falling down, within x milliseconds, then alerting/notifying the ambulance, caregivers, and doctors within y minutes. To achieve this high-level SLA requirement, many nested-dependent QoS metrics should be considered, such as high-quality sensors with minimum event detection delay (within x milliseconds), available networks with low latency, and a high-alert detection and notification analytic service to deliver the desirable alerts to relevant healthcare providers and relatives. As patients need to receive the required emergency treatment based on their health status within y minutes, this means that the aggregation of the response time from each layer should be within the time constraints, i.e. less than or equal to y minutes. A delay in the network, for example, would lead to a late response at the alert generation front-end, which could exceed the time the patient and healthcare provider was expecting (y minutes). Specifying SLA requirements with their required level of QoS and monitoring their adherence to these specifications is a non-trivial task and includes many challenges such as:

- A. *Heterogeneity of data sources (IoT sensors) and their distributed locations.*
- B. *Heterogeneity of the key QoS metrics across layers.*
- C. *Heterogeneity of application requirements.*
- D. *Lack of unified/standard methods for collecting the required metrics across-layer and from multiple providers for end-to-end SLA monitoring purposes.*

SLA specification and monitoring current research efforts:

Substantial research on the specification and monitoring of QoS and SLAs has been conducted for computer networks, web services, Grids, and Cloud Computing. But limited literature is available that deal the problem of specifying and monitoring end-to-end QoS and SLAs in an IoT application eco-system. For example, Netlogger provides an API that can be used by applications to check the load on network resources before and after performing operations/sending requests. However, Netlogger only monitors network resources and does not extend to other components of an IoT application [14, 15]. The Web Service Level Agreement (WSLA) standard described in [16] was developed for web service SLA specification. Also, WS-Agreement from the Open Grid Forum (OGF) defines a web service agreement specification as a protocol for launching an agreement between two parties. An illustration of how cloud providers in industry apply SLAs is shown in [21]. Cloud providers such as AmazonEC2, S3 (IaaS provider) and Windows Azure Compute and Storage, serve a pre-defined SLA, and the user can then choose the most appropriate provider that will fit their requirements. After entering into a contract with the selected provider, the SLA can be monitored against violations using third parties such as Cloudwatch, Cloudstatus and Monitis. The LoM2HiS Framework [14] aims to monitor and enforce SLA objectives in the cloud environment, especially; scalability, efficiency and reliability requirements. The framework aims to map low-level resource metrics to high SLAs objectives. However, the LoM2HiS Framework does not extend beyond the Cloud infrastructure layer. A European Commission Report on Cloud Computing Service Level Agreements [24] identifies and describes several interesting research efforts. SLA(T) by the SLA@SOI project [25][27] is a model and language for service description that expresses the dependencies among services within/across layers in the Cloud. Another project (CONTRAIL) provides a quality model [28] for capturing different parameters of interest for customers and providers. The IRMOS project [3] proposes two SLAs at different levels: an application SLA to express high-level application terms between consumers and providers, and technical SLAs to express the low level QoS parameters linked to the infrastructure resources. Cloud4SOA [31], is a project which provides a unified monitoring interface that gives an overview of all of the customer deployments at one time, as well as selecting a set of unified metrics for monitoring both the execution and the usage of an application. IRMOS [32] provides an adaptable monitoring framework that collects data from both the application and technical level to monitor real-time application execution at time intervals based on the collected monitoring information and its associated SLA terms.

Despite a number of impressive research efforts into the specification and monitoring of QoS requirements, none of them are suitable in context of IoT applications. Developing formal approaches for the specification of QoS requirements and monitoring end-to-end IoT ecosystems is what we term as the next “grand challenge” for distributed systems researchers, and current platforms and techniques for monitoring IoT and Cloud computing fall short of this grand challenge.

References

- [1] A. Flammini, and E. Sisinni, “Wireless Sensor Networking in the Internet of Things and Cloud Computing Era.” *Procedia Engineering*, vol. 87, pp. 672-679, Dec. 2014.
- [2] J. Gantz and D. Reinsel, “The digital universe in 2020: Big data, bigger digital shadows, and biggest growth in the Far East,” *IDC iView: IDC Anal. Future*, vol. 2007, pp. 1-16, Dec. 2012.
- [3] A. Galati et al. “A WS-Agreement based SLA implementation for the CMAC platform”. In *Economics of Grids, Clouds, Systems, and Services*, Springer-Verlag Heidelberg: Springer International Publishing, 2014.
- [4] R. Ranjan, “Streaming big data processing in datacenter clouds.” *IEEE Cloud Computing*, vol. 1, no. 1, pp.78-83, May 2014.
- [5] R. Buyya, and A.V. Dastjerdi, eds., *Internet of Things: Principles and Paradigms*. Elsevier, 2016.
- [6] K. Radha et al. “Service Level Agreements in Cloud Computing and Big Data.” *International Journal of Electrical and Computer Engineering*, 5(1), p.158, Feb. 2015.
- [7] IBM What is Big Data? – Bringing Big Data to the Enterprise: [<https://www-01.ibm.com/software/data/bigdata/what-is-big-data.html>]
- [8] Zheng, X., Martin, P., Brohman, K. and Da Xu, L. “Cloud service negotiation in internet of things environment: a mixed approach.” *Industrial Informatics, IEEE Transactions on*, vol. 10, pp. 1506-1515, May 2014.
- [9] M. Diaz, C. Martin, and B. Rubio, “State-of-the-art, challenges, and open issues in the integration of Internet of things and cloud computing.” *Journal of Network and Computer Applications*, 2016.

- [10] M. Chen, S. Mao, and Y. Liu. Big data: a survey. *Mobile Networks and Applications*, vol. 19, no. 2, pp.171-209, Apr. 2014.
- [11] D. Gascon, and A. Asin 50 sensor applications for a smarter world, [online]. Available: http://www.libelium.com/resources/top_50_iot_sensor_applications_ranking; 2015
- [12] N. Li et al. "A new methodology to support group decision-making for IoT-based emergency response systems." *Information Systems Frontiers*, vol. 16, no.5, pp.953-977, 2014.
- [13] P.P. Jayaraman et al. "Orchestrating Quality of Service in the Cloud of Things Ecosystem." In 2015 IEEE International Symposium on Nanoelectronic and Information Systems (pp. 185-190). IEEE. December 2015.
- [14] V.C. Emeakaroha et al. "Low level metrics to high level SLAs-LoM2HiS framework: Bridging the gap between monitored metrics and SLA parameters in cloud environments." In High Performance Computing and Simulation (HPCS), 2010 International Conference on (pp. 48-54). IEEE, June 2010.
- [15] D. Gunter et al. "Netlogger: a toolkit for distributed system performance analysis", 8th International Symposium on Modeling, Analysis and Simulation of Computer and Telecommunication Systems, pp. 267-273, 2000.
- [16] A. Keller, and H. Ludwig. "The WSLA framework: Specifying and monitoring service level agreements for web services." *Journal of Network and Systems Management*, vol. 11, no. 1, pp.57-81, Mar 2003.
- [17] M. Alhamad, T. Dillon, and E. Chang. Service level agreement for distributed services: a review. In Dependable, Autonomic and Secure Computing (DASC), 2011 IEEE Ninth International Conference on(pp. 1051-1054). IEEE.Dec. 2011.
- [18] A. Andrieux et al. Web services agreement specification (WS-Agreement). InOpen Grid Forum, vol. 128, no. 1, p. 216, Mar 14. 2007.
- [19] A. Sahai et al. Specifying and monitoring guarantees in commercial grids through SLA. In Cluster Computing and the Grid, 2003. Proceedings. CCGrid. May 2003.
- [20] M. Alhamad, T. Dillon, and E. Chang. A survey on SLA and performance measurement in cloud computing. In On the Move to Meaningful Internet Systems: OTM 2011 (pp. 469-477). Springer Berlin Heidelberg. 2011.
- [21] L. Wu, and R. Buyya. Service Level Agreement (SLA) in utility computing systems. IGI Global... Apr. 2012.
- [22] K. Alhamazani et al. Cross-Layer Multi-Cloud Real-Time Application QoS Monitoring and Benchmarking As-a-Service Framework. 2015.
- [23] G. Cicotti et al. How to monitor QoS in cloud infrastructures: The QoSMONaaS approach. In Intelligent Distributed Computing VI (pp. 253-262). Springer Berlin Heidelberg. 2013.
- [24] D. Kyriazis. Cloud computing service level agreements: exploitation of research results. European Commission Directorate General Communications Networks Content and Technology Unit, Tech. Rep, 5. 2013.
- [25] SLA@SOI Project, <http://sla-at-soi.eu/>
- [26] RESERVOIR Project, <http://www.reservoir-fp7.eu/>
- [27] 4CaaS Project, <http://4caast.morfeo-project.org/>
- [28] R. Cascella et al. "Contrail: Distributed Application Deployment under SLA in Federated Heterogeneous Clouds", Springer, Lecture Notes in Computer Science, 2013.
- [29]Q-ImPRESS Consortium, Service Architecture Meta-Model (SAMM), Project deliverable d2.1, Sep. 2008 [online]. Available: http://www.q-impress.eu/wordpress/wp-content/uploads/2009/05/d21-service_architecture_meta-model.pdf
- [30] http://irmosproject.eu/Files/IRMOS_NEXOF-RA_SLAs_QoS.pdf
- [31] Cloud4SOA Project, <http://www.cloud4soa.eu/>
- [32] IRMOS Project, <http://www.irmosproject.eu/>
- [33] L. Wang and R. Ranjan, "Processing Distributed Internet of Things Data in Clouds," *IEEE Cloud Computing*, vol. 1, no. 2, 2015, IEEE Computer Society.
- [34] S3 SLA, Amazon Web Services, <https://aws.amazon.com/s3/sla/>, accessed July 2016.
- [35] EC2 SLA, Amazon Web Services, <https://aws.amazon.com/ec2/sla/>, accessed July 2016.