A Fog-enabled Smart Home Analytics Platform

Theo Zschörnig[1], Robert Wehlitz[1] and Bogdan Franczyk[2,3]
[1]*Institute for Applied Informatics (InfAI), Leipzig University, Goerdelerring 9, 04109 Leipzig, Germany*
[2]*Information Systems Institute, Leipzig University, Grimmaische Str. 12, 04109 Leipzig, Germany*
[3]*Business Informatics Institute, Wrocław University of Economics, ul. Komandorska 118-120, 53-345 Wrocław, Poland*

Keywords: Smart Home, Fog Computing, Internet of Things, Analytics Architecture.

Abstract: Although the usage of smart home devices such as smart speakers, light bulbs and thermostats has increased rapidly in the past years, their added value, compared to conventional devices, is mostly limited to simple control and automation logic. In order to provide adaptive smart home environments, it is necessary to gain deeper insights into the data generated by these devices and use it in sophisticated data processing pipelines. Providing such analytics to a multitude of consumers requires specialised architectures, which are able to overcome various challenges identified by scientific literature. Currently available smart home analytics architectures are not designed to tackle all of these issues, specifically fault-tolerance, network-usage, latency and external regulations. In this paper, we propose an architectural solution to address these challenges based on the concept of Fog computing. Furthermore, we provide insight into the motivation for this research as well as an overview of the current state of the art in this field.

1 INTRODUCTION

In the past years, Internet of Things (IoT) technologies and solutions have been adopted in a variety of domains for personal and business use. In the year 2018, the number of connected Internet of Things devices worldwide has already risen to 16.8 billion in communications, 5.4 billion in commercial and industrial electronics as well as 5.9 billion in terms of consumer devices (IHS, 2018). All these devices are sources of data, which provide their users, but also businesses, industry and researchers, with the opportunity to gain valuable insights into everyday life and industrial value creation. Furthermore, state of the art analytical algorithms and methods play a key role in achieving an even deeper understanding of the use of IoT devices and their surroundings, hence further increasing their added value.

In this regard, a major challenge for businesses is to provide their customers with technical solutions to utilize the full potential of their IoT devices in terms of data insight and value-added information. In this context, the solutions, offered by businesses and researchers alike, mostly involve Big Data technologies, embedded into cloud platforms, which offer an abundance of processing and storage resources. While this seems appropriate for a variety of scenarios, especially IoT environments characterized by sensitive data, such as smart home ecosystems, expose additional requirements regarding data security, fault tolerance and latency, but at same time ease of use, therefore creating the need for new solutions regarding analytical architectures. In order to tackle these challenges, we present an IoT analytics platform designed for smart home environments based around the Fog computing paradigm.

In this paper, we motivate our research as well as the technical challenges to be addressed by our solution proposal and the opportunities it provides (Section 2). We give an overview of the state of the art in IoT analytics regarding technologies and architectures. In addition, we show that these are not fully suitable in terms of smart home analytics (Section 3). The main contribution of this paper, an approach to build an analytics platform architecture to be used in smart home environments, is described in Section 4. In conclusion, we provide ideas to further the research in this field as well as our own (Section 5).

2 MOTIVATION

It is estimated that the number of annually sold smart home devices will grow to 939 million in the year

Zschörnig, T., Wehlitz, R. and Franczyk, B.
A Fog-enabled Smart Home Analytics Platform.
DOI: 10.5220/0007750006160622
In *Proceedings of the 21st International Conference on Enterprise Information Systems (ICEIS 2019)*, pages 616-622
ISBN: 978-989-758-372-8

2022 (IDC, 2018). Consumers anticipate benefits from the use of these smart home technologies mainly in the areas of energy management, home automation and control as well as home security (Wilson et al., 2017). In this context, a major part in the creation of added value from smart devices is the ability to generate insights from the data they are collecting. While research focuses on developing new and applying existing methods of data processing on smart home environments (Brush et al., 2018), current data processing and analytics architectures evolve around cloud-based Big Data solutions.

These architectures aim to resolve a variety of different challenges for IoT analytics in general, already identified in scientific literature:

- *Limited resources at IoT devices* (Stolpe et al., 2016)
- *High data volume, heterogeneity and velocity* (Chen et al., 2015; Stolpe et al., 2016; Marjani et al., 2017)
- *Occasional connection loss of IoT devices* (Chen et al., 2015; Rozik et al., 2016)
- *Real-time data processing* (Chen et al., 2015; Zaslavsky et al., 2015; Stolpe et al., 2016)
- *Personalized analytics* (Biswas et al., 2014; Auger et al., 2017)
- *Data security and privacy* (Zaslavsky et al., 2015; Stolpe et al., 2016)

Analytical architectures in common smart home scenarios need to be able to address these requirements. In contrast to other fields of application of the IoT, they also need to offer capabilities to handle long-running, rather static, Big Data problems as well as smaller, more intimate analytical problems, which are often changed in terms of data sources and requirements. For example, the training of outlier detection or non-intrusive load monitoring algorithms based on large datasets for energy management, are as important as small processing tasks of a few data sources, e.g. temperature tracking of a single thermostat sensor. At a technical level, this causes the necessity for an analytics architecture, which is scalable and elastic while at the same time providing fast and flexible ways to change and extend analytical processing. In addition, because of possible connection loses, data processing needs to be fault-tolerant at both the cloud and the local level to ensure the ongoing operation of the smart home. With time critical use cases such as home security and disaster detection and prevention, being an essential component of smart home expectations by end-consumers (Brush et al., 2018), low latency is key in terms of data processing. Another important aspect of analytics architectures is data security and privacy. Especially with the introduction of the General Data Protection Regulation (GDPR) in the European Union, processing of sensitive data, as it is the case in most smart home environments, needs to be in accordance to legislative regulation.

Looking at current smart home analytics architectures, these requirements seem to have only been insufficiently met. Especially in terms of fault-tolerance, latency and data privacy cloud-only data processing solutions seem to be ill equipped for the aforementioned requirements. In contrast, fog computing provides a promising approach to address these issues. The term “fog computing” was first used by CISCO in 2012 and describes a concept where data processing capabilities are moved from the center of the cloud to the edge of the network. It is therefore an extension of the cloud computing paradigm and fog components cannot stand alone. (Mouradian et al., 2018)

In general, the fog computing paradigm offers several advantages over cloud computing in terms of reduced latency and network usage, higher fault-tolerance, resource availability at the source of data and data processing in compliance with specific legislation (Byers, 2017; Klonoff, 2017; Ravindra et al., 2017; Velasquez et al., 2017). In terms of smart home analytics, enabling local networks of smart devices to process their data cloud-independent ensures the continuous operation of all processes and tasks in case of connectivity issues to cloud services. Moreover, offloading processing and analytics tasks to local processing nodes further reduces processing latency. Furthermore, reducing the volume of data send in-between and from IoT devices to cloud services will decrease latency issues even more (Stojkoska & Trivodaliev, 2017). In addition, local data processing addresses concerns regarding data privacy and security. This involves anonymization at the location of data generation as well as reduced transmissions of sensitive user data to cloud services.

3 STATE OF THE ART

Smart home networks typically include various IoT devices such as smart thermostats, light bulbs, speakers, locks, but also voice control devices and cameras. These devices are connected either directly to their respective cloud backend or via a central entity, a so-called “gateway”, utilizing IoT-based communication protocols, such as Z-Wave, ZigBee, etc. The cloud backend services are used to control and access IoT devices remotely and, furthermore,

most of the data processing of IoT devices is done via the underlying cloud infrastructure after transferring the required data. Especially large internet companies like Amazon rely solely on cloud-based services for data processing (Amazon, 2019).

Looking at the scientific literature regarding IoT analytics architectures, dedicated smart home solutions are rare. They focus on specific topics such as device security (Haddadi et al., 2018) or energy management (Al-Ali et al., 2017) and offer cloud-based solutions for data processing with data collection nodes at a local level. The area of application of these works is limited to specific use cases, hence missing the needed flexibility to react to changing requirements of smart home environments as described in Section 2.

Other domains of application have yielded additional approaches to IoT analytics architectures. Several publications in the field of industrial IoT (IIoT) propose fog computing-based architectures (Rehman et al., 2018; Alexopoulos et al., 2018) for data processing and analytics. Although, these approaches are able to address most of the requirements of IoT analytics in general, they are specifically designed around the processes and roles of industrial manufacturing. It is therefore questionable if their deployment in smart home environments is possible without extensive adjustments.

There has also been research into analytics architectures in Smart City scenarios. These works usually offer cloud-based Big Data solutions (Cheng et al., 2015; Ta-Shma, 2018) and are therefore not well suited for smart home scenarios because of the requirements described in Section 2.

General fog computing architectures for IoT use cases are described in Alturki et al. (2017) and Ravindra et al. (2017). They offer valuable insights into fog architectures regarding important components and data flow modelling. Nevertheless, their experimental setups are rather static and are hence missing flexibility in data processing locations as well as use case adoption.

None of the analysed related works is able to provide full coverage of all challenges for analytics architectures in smart home environments. Hence, this paper aims to provide an architectural approach to fill this gap.

4 SOLUTION PROPOSAL

In order to provide sophisticated data manipulation, analytics and persistence in a flexible manner for smart home environments, we propose the following architectural solution. This proposal is based on the previous works of the authors which has been adapted to reflect the changed requirements in smart home environments as mentioned before. In this regard, the formerly presented architectural proposal was extended to include a fog layer, which aims to address the challenges described in section 2. The main purpose of this layer is to allow for offloading data processing tasks from the cloud layer. The complete architecture is shown in Figure 1.

4.1 Cloud Layer

The cloud layer is based on the concept of the Kappa architecture, which treats all data as a stream and drops the batch layer of the more traditional Lambda architecture in favour of only a speed and serving layer (Kreps, 2014). This allows for flexible data processing while still providing capabilities to handle large, high velocity amounts of data from various sources. The data processing and analytics are executed by the *stream processing system*. This system includes a *log data store* for handling data streams and additionally container-based microservices, which represent individual processing tasks of analytics pipelines.

The core component to orchestrate the processing jobs is the *flow engine*, which executes analytics flows. These flows are typically designed by power users and may be accessed and deployed by regular users as well. The execution of analytics flows is triggered directly at the *flow engine* by either API access, frontend or mobile applications.

Surrounding the *flow engine* are auxiliary services providing data and metadata for analytics flow execution. In this regard, the *flow repository* is the central point for storing analytics flows. These flows employ a simple flow-chart methodology with nodes and edges, in which nodes represent tasks of data manipulation or analytics, so called operators, and the edges represent the data flow between nodes. The capabilities of operators as well as additional metadata are stored in the *operator repository*. This information is used to design analytics flows, but also by the *flow parser* to check analytics flow validity before execution. A running instance of an analytics flow is an analytics pipeline and registered in the *pipeline registry* by the *flow engine*.

In addition to these, already by Zschörnig et al. (2017) established, components of the cloud layer, we introduced a *reasoner*. Its main purpose is to decide whether analytics operators may be offloaded to the fog layer of the overall architecture. Therefore, a set

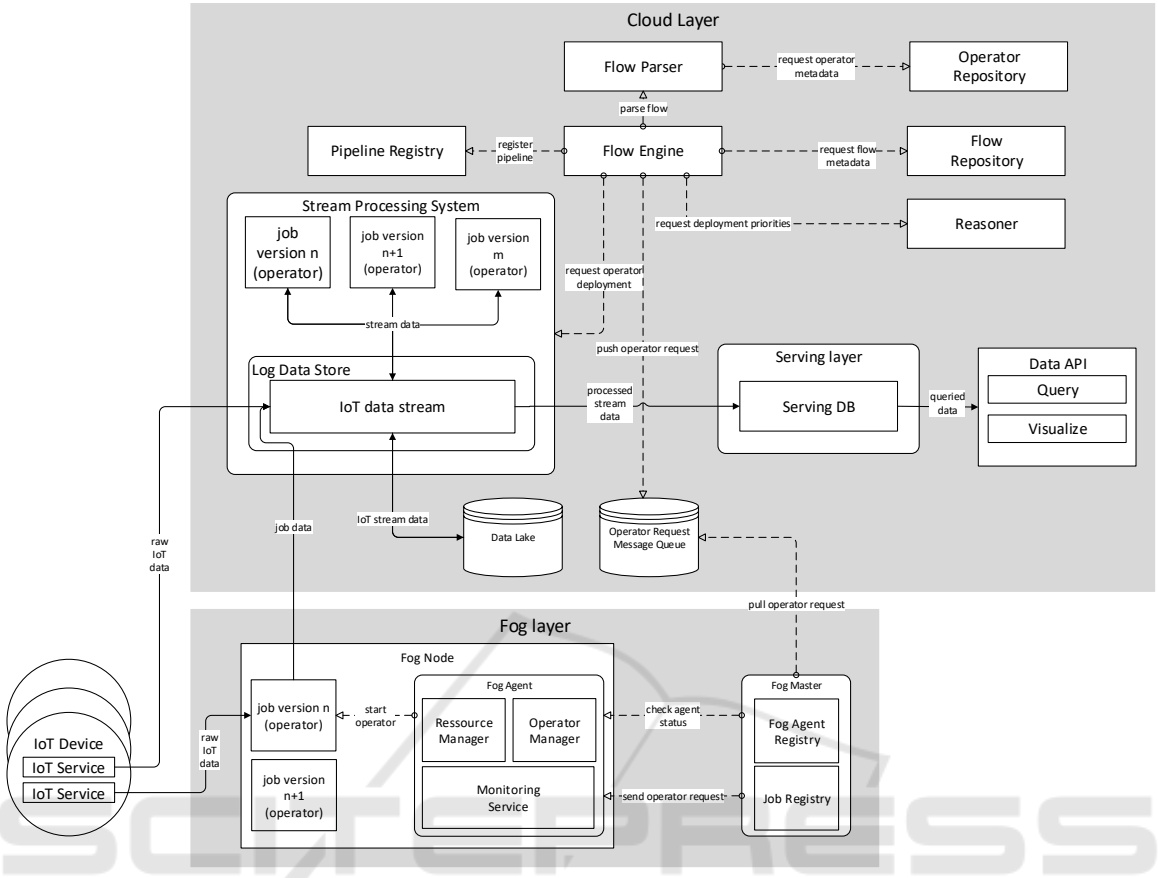

Figure 1: Proposed architecture with data and control flows, adapted from Zschörnig et al. (2017).

of rules and conditions needs to be in place to support these decisions. The conditions have to be designed in a way that allows to conclude hard (MUST and MUST NOT) and soft (SHOULD and SHOULD NOT) scheduling decisions for the *flow engine*. An example for a condition could be: "Sensitive data used → MUST be deployed locally". Following the scheduling rules, the *flow engine* either requests operator deployment in the cloud *stream processing system* or issues an operator request to the fog layer. To ensure fault-tolerance requests to the fog layer are decoupled using a message queue.

The cloud layer by itself already provides a wide range of capabilities to address the challenges of IoT analytics architectures. In terms of smart home analytics, fault-tolerance, low latency and data privacy issues remain. Therefore, the fog layer specifically addresses these issues.

4.2 Fog Layer

The design of the fog layer components and its integration in the existing architectural proposal is based on the challenges to be addressed by fog computing architectures as described in scientific literature.

Fog resources are mobile as well as added and removed in an unpredictable manner (Dastjerdi & Buyya, 2016; Byers, 2017; Velasquez et al., 2017), therefore, a fog architecture has to provide mechanisms to cope with the resulting insecurities concerning connection stability and data flow volatility. In addition, the processing of data is supposed to be done in the proximity of its generation, requiring optimal orchestration of processing tasks (Byers, 2017; Ravindra et al., 2017) while taking into account the limited resources of fog devices (Dastjerdi & Buyya, 2016). Beyond that, fog architectures need to handle a large variety of different use cases (Byers, 2017) and provide the means for secure data transmission (Dastjerdi & Buyya, 2016; Velasquez et al., 2017). Literature regarding this topic also mentions a high number of fog nodes, which need to be managed by the overall architecture (Mouradian et al., 2018), which requires scalable components for orchestration and

communication. Probably the most important challenge is the support of real-time data processing and analytics under the assumption of external constraints. These include limited and heterogeneous resources of fog devices along with organisational or judicial regulations (Dastjerdi & Buyya, 2016; Ravindra et al., 2017; Velasquez et al., 2017; Mouradian et al., 2018).

The fog layer comprises all components of the architecture, which are not deployed in the cloud but rather at a local hardware level such as IoT gateways. These fog nodes are not as mobile as edge devices and therefore more reliable in terms of connectivity and availability. Still, architectural components need to be able to run independently from the cloud layer to ensure the ongoing usefulness of smart devices in smart home scenarios. In order to achieve this, the proposed solution is based on similar approaches like Brito et al. (2017) and employs two main components, the *fog master* (FM) and the *fog agent* (FA). Every fog node, which is used to execute analytics operators, needs to have a FA deployed. This component is able to manage and monitor hardware resources as well as analytics operators. In addition, at least one fog node has to deploy the FM component. Its main tasks are to register all available FAs and to orchestrate operator deployments at the fog layer. The FM component may be deployed at multiple fog nodes thus providing additional fault tolerance. In this scenario, all decisions are made via a quorum of all FMs. During the registration process of a FA, their available resources and location are registered as well. This is used by the FM to determine if an operator request may be executed, taking into account the external constraints of an analytics flow.

The FM constantly pulls operator deployment requests from the corresponding *message queue* of the *flow engine*. When an operator request is received, the FM checks its own FA registry, if the operator request can be satisfied using the available FAs. The status of all known FAs is continuously checked by the FM. This includes their overall health as well as available resources.

If a FA is available for operator deployment, the FM sends it a request containing the necessary metadata to start the operator. The FM prioritizes FAs, which are near the source of the data to be processed. The information regarding the data source location is relayed along with the operator request from the *flow engine*. In addition, a new entry in the *job registry* of the FM is created. This ensures that in the case of an offline FA, all its jobs may be reassigned to other FAs to ensure fault-tolerance. In the case of no available FA for operator deployment, the FM informs the *flow engine*, which either cancels the flow execution or tries to deploy the operator at the cloud layer.

A FA comprises three components to handle operator deployment. The *resource manager* checks for available hardware resources and the current load of the device. This information is constantly sent to the FM. The *operator manager* checks, if an operator to be executed is available or even possible to deploy with regards to available resources. FA operators are container-based comparable to their cloud counterparts. This allows for processing isolation. In addition, missing operators are pulled conveniently using the mechanisms current container software solutions offer. The *monitoring service* checks if all deployed operators are running and sends FA health data to the FM.

5 CONCLUSIONS AND OUTLOOK

In this paper, we presented and architectural approach for IoT analytics in smart home environments. It is based on the concept of fog computing to enable low latency, fault-tolerant and privacy observing data processing, all of which are requirements for smart home analytics architectures identified by scientific literature.

We provided insights into the current state of the art in IoT analytics architectures research and showed that already existing solutions are not sufficient to address all architectural challenges identified for smart home analytics. We found that fault-tolerance, low-latency data processing as well as external regulations are key aspects when designing an analytics architecture. The presented approach utilizes several concepts from system and software engineering, such as fog computing, Kappa architecture, microservices and container virtualisation to solve the surrounding problems. Moreover, the solution architecture comprises components to allow for optimal orchestration of data processing along individual analytics pipelines.

Future research in this field needs to focus on identifying conditions and requirements for deployment rules of analytics pipeline tasks. It seems plausible to derive the resulting rules from “static” sources such as resource availability, legislation, etc., but also from human behaviour using already gathered IoT data. Understanding the usage of smart devices by consumers may lead to the use of machine

learning algorithms to establish optimal distribution of analytics pipeline jobs between cloud and fog nodes. Finally, the research concerning choreography-based scheduling approaches of analytics tasks needs to be furthered in order to provide increased fault-tolerance of the overall architecture.

A prototypical implementation of the solution proposal has already been developed. The resulting software prototype needs to be evaluated in future research with regards to the challenges for IoT analytics as well as fog computing architectures, but also regarding performance compared to different architectural concepts.

ACKNOWLEDGEMENTS

The work presented in this paper is partly funded by the European Regional Development Fund (ERDF) and the Free State of Saxony (Sächsische Aufbaubank - SAB)

REFERENCES

Alturki, B., Reiff-Marganiec, S., and Perera, C. (2017). A hybrid approach for data analytics for internet of things. In S. Mayer (Ed.), *ICPS: ACM international conference proceeding series, Proceedings of the Seventh International Conference on the Internet of Things* (pp. 1–8). New York, NY, USA: ACM.

Amazon. (2019). Alexa Voice Service. Retrieved from https://developer.amazon.com/de/alexa-voice-service

Auger, A., Exposito, E., and Lochin, E. (2017). Sensor observation streams within cloud-based IoT platforms: Challenges and directions. In *2017 20th Conference on Innovations in Clouds, Internet and Networks (ICIN)* (pp. 177–184). IEEE.

Biswas, A. R., and Giaffreda, R. (2014). IoT and cloud convergence: Opportunities and challenges. In *2014 IEEE World Forum on Internet of Things (WF-IoT)* (pp. 375–376). IEEE.

Brito, M. S. de, Hoque, S., Magedanz, T., Steinke, R., Willner, A., Nehls, D., . . . Schreiner, F. (2017). A service orchestration architecture for Fog-enabled infrastructures. In *2017 Second International Conference on Fog and Mobile Edge Computing (FMEC)* (pp. 127–132). IEEE.

Brush, A. J., Hazas, M., and Albrecht, J. (2018). Smart Homes: Undeniable Reality or Always Just around the Corner? *IEEE Pervasive Computing*, *17*(1), 82–86.

Byers, C. C. (2017). Architectural Imperatives for Fog Computing: Use Cases, Requirements, and Architectural Techniques for Fog-Enabled IoT Networks. *IEEE Communications Magazine*, *55*, 14–20.

Cheng, B., Longo, S., Cirillo, F., Bauer, M., and Kovacs, E. (2015). Building a Big Data Platform for Smart Cities: Experience and Lessons from Santander. In B. Carminati (Ed.), *2015 IEEE International Congress on Big Data (BigData Congress): June 27, 2015 - July 2, 2015, New York, New York, USA* (pp. 592–599). Piscataway, NJ: IEEE.

Dastjerdi, A. V., and Buyya, R. (2016). Fog Computing: Helping the Internet of Things Realize Its Potential. *Computer*, *49*, 112–116.

Haddadi, H., Christophides, V., Teixeira, R., Cho, K., Suzuki, S., and Perrig, A. (2018). SIOTOME: An Edge-ISP Collaborative Architecture for IoT Security.

IDC (2018). New IDC Smart Home Device Tracker Forecasts Solid Growth for Connected Devices in Key Smart Home Categories. Retrieved from https://www.idc.com/getdoc.jsp?containerId=prUS43701518

IHS. (2018). IoT Trend Watch 2018. Retrieved from https://cdn.ihs.com/www/pdf/IoT-Trend-Watch-eBook.pdf?utm_campaign=PC10273-2_MD_eT1_MT_TMT_GLOBAL_IoT-Theme_3rd-IoT-eBook_2018_customers&utm_medium=email&utm_source=Eloqua

Klonoff, D. C. (2017). Fog Computing and Edge Computing Architectures for Processing Data From Diabetes Devices Connected to the Medical Internet of Things. *Journal of Diabetes Science and Technology*, *11*, 647–652.

Kreps, J. (2014). Questioning the Lambda Architecture: The Lambda Architecture has its merits, but alternatives are worth exploring. Retrieved from https://www.oreilly.com/ideas/questioning-the-lambda-architecture

Marjani, M., Nasaruddin, F., and Gani, A. (2017). Big IoT Data Analytics: Architecture, Opportunities, and Open Research Challenges. *IEEE Access*, *5*, 5247–5261.

Mouradian, C., Naboulsi, D., Yangui, S., Glitho, R. H., Morrow, M. J., and Polakos, P. A. (2018). A Comprehensive Survey on Fog Computing: State-of-the-Art and Research Challenges. *IEEE Communications Surveys & Tutorials*, *20*(1), 416–464.

Ravindra, P., Khochare, A., Reddy, S. P., Sharma, S., Varshney, P., and Simmhan, Y. (2017). ECHO: An Adaptive Orchestration Platform for Hybrid Dataflows across Cloud and Edge.

Rehman, M. H. u., Ahmed, E., Yaqoob, I., Hashem, I. A. T., Imran, M., and Ahmad, S. (2018). Big Data Analytics in Industrial IoT Using a Concentric Computing Model. *IEEE Communications Magazine*, *56*, 37–43.

Rozik, A. S., Tolba, A. S., and El-Dosuky, M. A. (2016). Design and Implementation of the Sense Egypt Platform for Real-Time Analysis of IoT Data Streams. *Advances in Internet of Things*, *06*, 65–91.

Stojkoska, B. L. R., and Trivodaliev, K. V. (2017). A review of Internet of Things for smart home:

Challenges and solutions. *Journal of Cleaner Production*, *140, Part 3*, 1454–1464.

Stolpe, M. (2016). The Internet of Things: Opportunities and Challenges for Distributed Data Analysis. *ACM SIGKDD Explorations Newsletter*, *18*, 15–34.

Ta-Shma, P., Akbar, A., Gerson-Golan, G., Hadash, G., Carrez, F., and Moessner, K. (2018). An Ingestion and Analytics Architecture for IoT Applied to Smart City Use Cases.*IEEE Internet of Things Journal*,*5*, 765–774.

Velasquez, K., Abreu, D. P., Goncalves, D., Bittencourt, L., Curado, M., Monteiro, E., and Madeira, E. (2017). Service Orchestration in Fog Environments. In M. Younas, M. Aleksy, and J. Bentahar (Eds.), *2017 IEEE 5th International Conference on Future Internet of Things and Cloud: FiCloud 2017: Prague, Czech Republic, 21-23 August 2017 : proceedings* (pp. 329–336). Piscataway, NJ: IEEE.

Wilson, C., Hargreaves, T., and Hauxwell-Baldwin, R. (2017). Benefits and risks of smart home technologies. *Energy Policy*, *103*, 72–83.

Zaslavsky, A., and Georgakopoulos, D. (2015). Internet of Things: Challenges and State-of-the-Art Solutions in Internet-Scale Sensor Information Management and Mobile Analytics. In *2015 16th IEEE International Conference on Mobile Data Management (MDM)* (pp. 3–6).

Zschörnig, T., Wehlitz, R., and Franczyk, B. (2017). A Personal Analytics Platform for the Internet of Things: Implementing Kappa Architecture with Microservice-based Stream Processing. In *Proceedings of the 19th International Conference on Enterprise Information Systems* (pp. 733–738). SCITEPRESS - Science and Technology Publications.