# Privacy Aware on-Demand Resource Provisioning for IoT Data Processing

Tom Kirkham[1(✉)], Arnab Sinha[2], Nikos Parlavantzas[2], Bartosz Kryza[3], Paul Fremantle[4], Kyriakos Kritikos[5], and Benjamin Aziz[4]

[1] STFC, Chilton, UK
tom.kirkham@stfc.ac.uk
[2] Inria, Rennes, France
arnab.sinha@inria.fr
[3] AGH, Krakow, Poland
[4] University of Portsmouth, Portsmouth, UK
{paul.fremantle,ben.aziz}@port.ac.uk
[5] FORTH Crete, Heraklion, Greece

**Abstract.** Edge processing in IoT networks offers the ability to enforce privacy at the point of data collection. However, such enforcement requires extra processing in terms of data filtering and the ability to configure the device with knowledge of policy. Supporting this processing with Cloud resources can reduce the burden this extra processing places on edge processing nodes and provide a route to enable user defined policy. Research from the PaaSage project [12] on Cloud modelling language is applied to IoT networks to support IoT and Cloud integration linking the worlds of Cloud and IoT in a privacy protecting way.

**Keywords:** Cloud computing · Scalability · Internet of Things · Models@run.time

## 1 Introduction

The vision of an Internet of Things (IoT) heralds a new dawn in how people and devices relate to each other. Within environments such as the Smart City personalised services can take into account a person's historical behaviour and their current location. Delivery of these services in the environment via personalised messaging or even public displays has the potential to change personal perceptions of space and privacy.

Emerging EU law is set on a course to require personal consent before IoT based services can interact with a person and their data. Without such consent the capture of this data would be illegal. Thus, in order to future proof emerging IoT services privacy assurance is needed and one such way of doing this is by the provision of data filtering at the edge of the IoT network.

Increased processing capability in low power chips used in IoT networks provide the possibility that data can be filtered at source with respect to specific privacy/security rules. This will enable the handling of most sensitive information to be taken out of the

© ICST Institute for Computer Sciences, Social Informatics and Telecommunications Engineering 2016
B. Mandler et al. (Eds.): IoT 360° 2015, Part II, LNICST 170, pp. 87–95, 2016.
DOI: 10.1007/978-3-319-47075-7_11

hands of the service provider and for such networks to comply with the law. However, such privacy filtering adds latency to the core operation of the sensor board and in data intensive applications can cause potential bottlenecks in relation to quality of service.

In order to counter this, hybrid IoT data processing solutions for both privacy and service provision are needed. Such solutions will enable IoT networks to embrace the benefits of both processing at the edge and extra capacity from the Cloud. Existing work in the model-driven Cloud community illustrates how data can be sent to specific cloud infrastructures based on requirements associated with it. Using Smart City requirements from Canary Wharf this paper illustrate how such an approach can be applied to IoT in the Smart City.

## 2 Adapting to Context

Personal interaction with devices and sensors in terms of both passive and interactive engagements are set to change human conceptions on how data is shared. For example, current data shared using traditional social networking technologies such as Facebook is largely reliant on personal input. Within IoT connected environments, data sourced from fixed and mobile sensors is often collected automatically. As privacy awareness in the online domain influences behaviour in terms of choice of websites and data shared, within IoT environments it could change the places people go and choices they make.

### 2.1 Consent

Emerging EU legislation for consent from data subjects prior to data processing in IoT environments is in-line with current approaches to privacy in the online data sharing domain. Within the online community this can be seen manifest in the notification panels asking for consent to track Cookies on most websites. Within the IoT community the approach to achieve this is yet to be defined.

A key challenge in gaining this consent is to determine when and where the consent is required. Personal data in IoT is often produced from multiple sources and varieties of contexts, it differs from web services where data sources are often fixed and application specific. Add to this supported processing on remote infrastructure and the extent to which and prior consent is valid becomes cloudy.

To manage this complexity consent can be better managed in models of deployment and use. In that way the application can investigate such models to ensure consent before the data is processed. Using user defined policy such as in [1] is one way of describing this complex consent as illustrated in Fig. 1.

*Identity + Data / Policy*

**Fig. 1.** Typical model for privacy provision in web service environments

Supporting these policies with deployment models can apply the context and is present in work developed in [2] as illustrated in Fig. 2.

$$Identity + (Data * Context) / Policy$$

**Fig. 2.** Model for privacy provision in IoT environments

Thus applying context to the equation can significantly enhance the sensitivity of the data. This is a particular concern with IoT devices, where the data collected may include significant amounts of meta-data and contextual data which can infringe on privacy. For example, it has been shown that sensors such as accelerometers have unique "fingerprints" that can be used to identify the device [8]. In a typical application data will consist of different privacy levels and how these levels are handled will be described in the model. Taking these concerns into account during processing proposes a problem of adaptation between the device and supporting cloud in both privacy and quality terms.

### 2.2 Adaptation

Edge processing at a significant level in IoT environments is a relatively new phenomenon and related directly to the increasing power in terms of processing and decreasing energy consumption of microchips [3]. From a security perspective, filtering data at the edge enables data marked as private by users to be discarded at source. In addition it can reduce the amount of metadata and contextual data that is published. This reduces both the volume of data to process and the threat of leaked private data. However, for data intensive applications that run complex data analysis, computation at the edge is not always suitable. Edge computation adds delays on data collection and processing and forms a potential bottleneck. A solution to this problem is to support this processing by using either local or remote computing power, one way is to present flexible and on-demand Cloud-based support. The provision of such resource can be realised using the PaaSage platform.

The PaaSage project delivers an open, integrated platform to support model-based lifecycle management of applications executing on multiple cloud infrastructures. Specifically, the PaaSage platform support the generation of application deployment models to best satisfies application owner requirements. When run-time events make the current deployment unsatisfactory (e.g., QoS constraints are violated, or application owner requirements are changed), the platform dynamically adapts this deployment in the most efficient and reliable way. Adaptation in PaaSage relies on the models@run.time approach. Following this approach, the platform maintains models of the running deployment, requirements as well as environment properties. These models are continually updated through monitoring and form the basis of detecting deviations between the current deployment and requirements, of generating a target application deployment, and of transforming the current deployment into the target deployment.

In the context of IoT applications, the PaaSage platform can be used to optimally provide cloud resources when edge resources are insufficient. Specifically, the platform can monitor resource utilisation in the device and automatically trigger the deployment of additional data processing modules on cloud resources. The number and types of virtual machines and the associated cloud provider are selected in order to best satisfy the application's performance, security, energy consumption, and cost requirements. The selected application deployment can then be dynamically adapted when the platform identifies a better target deployment or when environment conditions change (e.g., workload variations, price changes of cloud providers).

PaaSage provides a set of interfaces to configure and monitor the Cloud. It not only enables non cloud specialists from the IoT domain to set specific deployment requirements such as security and quality of service but also to monitor how these requirements are respected during execution. Supporting the deployment and execution are Reasoning components that look to find optimal deployments based on user requirements and monitored metrics from the infrastructure (which can include the IoT network). From an IoT perspective this constant management of the Cloud environment ensures that security and quality can be maintained at the pace of change at the IoT platform.

## 3 Models

The PaaSage platform consists of various components that handle the life-cycle phases of configuration, deployment and execution of multi-cloud applications. Central to the operation of these components is the Cloud Application Modelling and Execution Language (CAMEL). This acts as a thread throughout each phase ensuring application deployment requirements are applied on multiple aspects of multi-cloud applications. These include operations such as provisioning and deployment topology, provisioning and deployment requirements, service-level requirements, metrics, scalability rules, providers, organisations, users, roles, security controls, execution contexts, and execution histories. Applying these models to link user requirements to the operation of IoT networks will enable the edge IoT processor to adopt privacy sensitive flexible Cloud based resource provisioning.

### 3.1 Handling Constraints Towards Privacy

The PaaSage platform can enforce data privacy in various ways through the CAMEL model. Firstly, it uses organisation models in the life-cycle phases of deployment and execution for representing organisations and users associated with a cloud-based application. For this purpose, the organisation package of the CAMEL metamodel is based on the organisation subset of CERIF [10], which is a modelling framework for specifying organisations, users and other entities in the research domain. It is an EU recommendation [11] for information systems related to research databases used for standardising research information and fostering research information exchange.

The CERIF model for an organisation contains blocks of information about the list of data centres offered by the organisation, the organisation itself, its users and user

groups as well as the permissions and role assignments issued by the organisation. CERIF enables varied organisations to express user privileges in relation to data processing and mpa permissions in federated environments. This mapping of identity will provide the edge processor with the ability to handle data from multiple organisations.

Secondly, data privacy could be maintained by specifying location requirements, involving one or more locations. A location can be either a geographical-based location (e.g., region or country) or a cloud location (i.e., a location specific to a cloud provider). This type of requirement is attached in deployment models either at the global level or at the local VM level. In this way, the end-user can specify a set of locations which should hold either for all the specified VMs or for a specific VM.

It is the responsibility of the PaaSage *Upperware* component, and particularly of the *Reasoner* sub-component, to consider such requirements in order to guarantee that all instances of VMs to be generated are situated in the respective locations included in these requirements. This can ensure any constraints in relation to location of data processing can also be applied in the filtering at the IoT edge processing. This is particularly significant for mobile sensors where data collected in some locations could be processed in the Cloud or edge whilst other locations can be marked as private.

CAMEL has the ability to create a digital form of the specification of all possible security controls as they have been identified by Cloud Security Alliance (CSA) and store them. A security control is identified by a name, a particular domain and sub-Domain, a textual description and to a set of security properties and metrics that it links to. In this way, when security requirements will need to be defined, the end-user will have the opportunity to select the security controls that better satisfy his/her needs by either browsing the respective security control list or making focused searches.

Integrating IoT specific controls into this list would enhance the security of distributed IoT networks by ensuring that the Cloud fits to the IoT deployment. A key benefit of edge processing is the simplification of data processing at a local level to the sensor. As this can reduce risk of data propagation as opposed to when it is processed in the Cloud. In cases where data has to be taken from the edge to the Cloud (such as in the need for extra processing power) PaaSage can look to tailor specific Cloud deployments to suit data sensitivity.

This flexibility is of key importance as it is likely that data from the IoT network can be of various levels of sensitivity depending on sensors and context. The ability for a supportive Cloud to adapt to this when providing extra resource to the edge is a key motivation in using PaaSage to support IoT data processing.

### 3.2 Managing Adaptability

CAMEL supports monitoring and scalability information in the deployment model and this is used to trigger dynamic adaptation. Specifically, the platform detects specified events, such as violations of service-level objectives or component failures, and enacts adaptation actions, such as vertical scaling, horizontal scaling, relocating components to different clouds as well as application restructuring.

Within the IoT environment adaptation may also be triggered by monitoring on the device to trigger a Cloud burst or the availability of a deployment model that better satisfies user requirements and goals (e.g., taking into account updated cloud provider offerings). Importantly, the PaaSage platform continually seeks to optimise application operation by finding better deployment models and enacting them in a cost-efficient and safe manner. Deriving deployment models relies on a user-provided utility function that represents the extent to which a given deployment model satisfies user requirements and goals.

## 4 Use Cases and Implementation

The use case in which we have developing an initial deployment of our prototype is focused on the Smart City. Requirements for the platform in terms of business case and function were sourced from Canary Wharf as part of a Smart City Challenge [1].

### 4.1 Smart Cities

Smart Cities can be defined in a variety of ways. A common feature in all definitions is the use of connected devices within the urban environment. This includes connecting existing infrastructure and management systems with sensors in the environment to improve city management, including aspects such as traffic control, parking, air quality and lighting. However, more dynamic uses of technology within the Smart City are embracing increased processing power of devices both personal and at device level.

Such applications include features such as personalisation of retail environments and advanced crowd management. In these scenarios the demand on computing power of the sensors within the environment and data processing modules varies with the numbers of people and the data demands of the application.

Management of the performance of applications in the Smart City typically fall into the hands of various agencies with often different service demands. For example, traffic control systems are usually supplied by local authorities responsible for traffic management across wide areas and demanding high levels of application reliability. Within shopping centres typically the infrastructure is controlled by the owner of the built infrastructure. Here the service is less critical but relies on greater amounts of personal data.

Implementation of IoT within an environment such as Canary Wharf has to balance both the application goals and with support for the reputation of the Smart City brand. Central to reputation management is the control of how data is both used and secured particularly with respect to personal data privacy.

### 4.2 Data Processing

Data processing in our implementation is achieved using the Intel Edison device platform. Collection of data is achieved by the capture of Bluetooth association data from personal devices as they pass into range. In order to better associate identity with

devices the project created a portal for device registration and association with users. During the device registration process personal privacy preferences can be set in relation to data yielded from the device and how it is used. In addition to these user-defined privacy policies, a set of core privacy policies were defined. These core policies implement the requirement to maintain the reputation of Canary Wharf within the Smart City domain.

These requirements captured in CAMEL initially sit at the middleware layer. Pushing them down to the device enables the management of sensed data with respect to privacy preference and identity. Example policies tested on the platform defined what types of data could be collected per user or identity. To implement this a data filtering module was created for the device that configured using policy and identity.

Identity is provided on the portal via user attributes submitted when signing onto the portal. This identity can be expressed using standards such as OAuth or SAML and transferred to the IoT platform. Policies defined by data subjects will enable association of specific context with certain users. DeviceID from sensed data is checked against identity and policy.

Using CAMEL to support the data filtering at the edge the prospect of data processing bottlenecks is reduced. Here, when the performance/processing levels of the core data processing module on the device reaches a pre-set threshold a notification is sent to Cloud burst. In this scenario, the message is sent to the PaaSage platform using the MQTT protocol.

## 5 Related Work

The platform presented in this paper offers a unique combination of data processing depending on the application/user specifications for computation in IoT networks. Significantly established areas for edge processing such as the routing of packets via Switches and Routers are now moving toward supported processing using Cloud based virtual networks and is the focus of newly funded research [2].

In terms of specific IoT and Cloud integration Aneka is an IoT application development Platform-as-a-Service (PaaS) that is capable of utilizing storage and compute resources of both public clouds [4]. It provides various services that allow users to control, auto-scale, reserve, monitor and bill users for the resources consumed by their applications. It also supports resource provisioning on public clouds such as Microsoft Azure, Amazon EC2 and GoGrid as well as on private clouds such as desktops and clusters. The resource provisioning is dynamic for a certain time and cost considering past execution history of applications and budget availability.

In comparison to our work, Aneka follows a similar approach. While on the one hand, the target vision is the same i.e. on-demand resource provisioning for IoT applications, on the other hand the approach for realization the ecosystem is different. PaaSage uses simple CAMEL model to specify the properties of the IoT application i.e. constraints and adaptability for data privacy, application performance and user preferences along with the IoT platform (which also serves for local data processing) and Aneka is itself a dedicated .NET-based application development PaaS.

In [9], the Webinos system pushes XACML policies out to devices to limit the spread of personal and contextual data. While the aims of this are broadly similar, there are two key differences. Firstly, the Webinos system is based around the core concept of devices being in the personal control of users and therefore having a "personal zone" to protect. By contrast, in a Smart City there are many devices that collect data on many different subjects, which is dealt with in our work. Secondly, in contrast with this work, the Webinos system does not implement automatic movement of processing based on load from edge devices into the cloud.

In [5], Aazam and Huh provides a model for Fog computing which provides a layer between IoTs and the cloud. Typically, their model performs resource management for the IoTs taking into account resource prediction, resource allocation, and pricing all in a realistically and dynamically; also considering customers' type, traits, and characteristics. The authors also mention that the Fog could provision for decisions concerning the security and privacy of data collected from the WSNs and IoTs using a Smart Gateway within the layer.

Contrasting with our work, this could be viewed as a different architecture where the Fog layer provides computation, privacy, security etc as services for IoTs. In fact, it overlaps with similar concepts like mobile cloud computing (MCC) and mobile-edge computing (MEC) [6]. Another notable difference as mentioned in Sect. 2.2, these kind of edge processing can add delays therefore leading to bottlenecks. Our PaaSage platform has the flexibility to adapt by using either local or remote processing, through flexible and on-demand Cloud based support. Another drawback as pointed out in [7], Fog devices are prone to greater threats like man-in-the-middle attack as they work at the edge of networks; we use a more tightly coupled architecture with the privacy module embedded within the IOT platform.

## 6 Future Work

This paper documents early stage research and investigations in combining IoT with existing work on the PaaSage project. Future work involves the broadening of the initial investigations to further define links between Cloud models and IoT. Configuration interfaces between the PaaSage platform and IoT devices also require further investigation. More efficient methods for device configuration taking into account combined IoT capability are interesting points of investigation.

## 7 Conclusion

Provision of edge processing in IoT networks can provide enhanced privacy provision and compliance in implementations processing personal data such as the Smart City. In order to support such provision at the edge extra provision for processing of non sensitive data can be provided via the Cloud. Using the PaaSage platform and Cloud modeling language CAMEL, Cloud computing infrastructure can be selected to suit the specific data processing needs and deployment characteristics of the IoT network.

**Acknowledgement.** This project has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under Grant Agreement n° 317715″.

## References

1. Cognicity Smart City Challenge Canary Wharf. http://www.Cognicity.london
2. Becon Horizon 2020 project. www.beacon.eu
3. Spinnewyn, B., Latré, S.: Towards a fluid cloud: an extension of the cloud into the local network. In: Latré, S., Charalambides, M., François, J., Schmitt, C., Stiller, B. (eds.) AIMS 2015. LNCS, vol. 9122, pp. 61–65. Springer, Heidelberg (2015). doi:10.1007/978-3-319-20034-7
4. Gubbi, J., Buyya, R., Marusic, S., Palaniswami, M.: Internet of Things (IoT): a vision, architectural elements, and future directions. Future Gener. Comput. Syst. **29**(7), 1645–1660 (2013)
5. Aazam, M.; Huh, E.-N.: Fog computing micro datacenter based dynamic resource estimation and pricing model for IoT. In: 2015 IEEE 29th International Conference on Advanced Information Networking and Applications (AINA), pp. 687–694, 24–27 March 2015
6. Yi, S., Li, C., Li, Q.: A survey of fog computing: concepts, applications and issues. In: Proceedings of 2015 Workshop on Mobile Big Data (Mobidata 2015). ACM, New York, pp. 37–42 (2015)
7. Stojmenovic, I., Wen, S.: The fog computing paradigm: scenarios and security issues. In: 2014 Federated Conference on Computer Science and Information Systems (FedCSIS), pp. 1–8, 7–10 September 2014
8. Bojinov, H., Michalevsky, Y., Nakibly, G., Boneh, D.: Mobile device identification via sensor fingerprinting (2014). arXiv preprint arXiv:1408.1416
9. Desruelle, H., Lyle, J., Isenberg, S., Gielen, F.: On the challenges of building a web-based ubiquitous application platform. In: Proceedings of 2012 ACM Conference on Ubiquitous Computing, pp. 733–736. ACM (2012)
10. Jeffery, K., Houssos, N., Jörg, B., Asserson, A.: Research information management: the CERIF approach. IJMSO **9**(1), 5–14 (2014). doi:10.1504/IJMSO.2014.059142
11. CERIF Specification. http://cordis.europa.eu/cerif/
12. EU PaaSage project. www.paasage.eu