An experiment on deploying a privacy-aware sensing as a service in the Sensor-Cloud

Thiago Moreira da Costa, Hervé Martin, Elie Rachkidi, Nazim Agoulmine

An experiment on deploying a privacy-aware sensing as a service in the Sensor-Cloud

Thiago Moreira da Costa[1], Hervé Martin[1], Elie Rachkidi[2], and Nazim Agoulmine[2]

[1] LIG, Université Grenoble Alpes, Grenoble, France
{thiago.moreira-da-costa, herve.martin}@imag.fr
[2] IBISC, Université d'Evry Val-d'Ossonne, Evry, France
{erachkidy, nazim.agoulmine}@univ-evry.fr

Abstract

As pervasive computing spreads, the physical reality and its phenomena are mapped into the digital world, generating large amount of data that can be interpreted and correlated to a variety of personal information. Its development has incorporated the advances in sensor, networking, web service, and data processing technologies, creating a new era of connected things, so called the Internet of Things (IoT). The IoT sensing service, scattered by the network – Virtual Sensor Network (VSN) –, has been leveraged by the Cloud Computing, delivering scalable, virtualized and geographical proximity resources. In this paper, we present an experiment on deploying Sensor-Clouds that provides a privacy-aware Sensing as a Service (pSaaS) using a new privacy model. This model provides two-fold Privacy Enforcement Point (PEP) that intermediates connected data providers and data consumer, implementing an *in-network* verification process that reasons about inference intention and personal information in order to deny access or degrade data utility to specific parts of the IoT data stream. We investigate how this novel pSaaS paradigm can be implemented to provide *on-demand* sensing of meaningful personal information through *in-network* data processing while still enforcing privacy. In addition, we investigate how to support its on-demand deployment on the Sensor-Cloud using a real VSN middleware.

1 Introduction

The increasing omnipresence of sensors is enabling the perception of physical phenomena into digital information, sensing humans in almost every context and aspect of their lives. The advances in sensory, networking, web service, and data processing technologies pushed the pervasive computing to the new era of the IoT [1]. Its sensing service were conventionally formed by connected entities, uniquely identified, that sensed the environment and published data to data consumers through a VSN [2]. With the convergence of the IoT sensing service and the Cloud Computing, so called Sensor-Cloud, some limitations of these VSN were addressed. In particular, Sensor-Cloud extends the conventional VSN by enabling the management of sensors connected through the Cloud infrastructure, which offers better storage and processing capabilities, supporting a number of innovative services both for the Cloud Computing and the VSNs [2]. As a result, a number of Sensing as a Services (SSaaSs) were developed, providing a versatile remote sensing service through the Sensor-Cloud infrastructure [3].

Privacy and ethical issues related to the manipulation and discovery of personal information become more critical in the IoT given its pervasiveness. While individuals (*data providers*) are highly concerned about their privacy [4], their privacy risk perception and their privacy behavior are paradoxical, since they continue consuming services that explicitly outbreak their privacy [5]. This is partly explained by the choices in privacy strategies and mechanisms that privilege data

processing in *data consumers*' side, which often misleads users to trust *data consumers* and to reveal more information than necessary by offering access to their raw sensor data, such as geographic location and microphone from smart-phones.

In this paper, we present an experiment on deploying Sensor-Clouds that provides a privacy-aware Sensing as a Service (pSaaS) using our novel privacy model. This model provides PEPs that intermediate connected *thing* (*data providers*) and *data consumer*, implementing an *in-network* verification process that reasons about inference intention, personal information, and privacy policies in order to grant/deny access or degrade data utility selectively in the IoT data stream. In order to interpret and reason about inference intention, our privacy model relies on the modern SSaaS paradigm that provides *in-network* data processing capacity, shifting the Knowledge Discovery and Data Mining process (KDDM) process implementation toward the Cloud infrastructure. We extend this concept of *in-network* data processing to guarantee that KDDM processes are implemented as Semantic Perceptions (SPs), which is an inference process that perceives semantics from observation of detectable qualities [6]. The objective of this study is to investigate how this SSaaS paradigm can be implemented to provide *on-demand* sensing of meaningful personal information, while still enforcing privacy. In addition, we investigate how to support the *on-demand* deployment of this privacy-aware Sensor-Cloud using the Multisite Orchestration SysTem (MOSt) [7], an open-source cloud deployment orchestrator, and eXtended Global Sensor Network (xGSN) [8], an open-source VSN platform, to deliver our pSaaS.

The remaining of the paper is organized as follows. Section 2 defines the objective and problem statement of this work. Section 3 briefly review works related to privacy in the IoT. Section 4 describes our privacy model and the extension for the xGSN architecture and MOSt. Next, in Section 5 we show the proposed experiment. Lastly, Section 6 we present preliminary results and future works.

2 Objective and Problem Statement

Motivated by the need of a privacy model for the IoT that is capable of addressing privacy issues in the SSaaS and its deployment *on-demand*, we strive at defining an experiment to verify the viability of our pSaaS using MOSt and xGSN. Our privacy model is based on *privacy by design* and *privacy by policy* approaches, following design principles of privacy engineering [9]. The PEPs modeled into Virtual Sensors (VSs) in the VSN assure by design that private sensor data stream are processed and verified *on-the-fly*. These PEPs implements a privacy by policy mechanism that evaluate privacy policy conditions that specify which personal information are allowed to be produced and released.

In this paper, we study how this *in-network* data processing can be implemented in SSaaS as SPs, producing streams of meaningful information in the Cloud infrastructure. This new SSaaS paradigm and the PEPs benefit from the Cloud service provisioning capacity to enforce privacy *on-the-fly* by reasoning over semantic annotated information stream. Figure 1 depicts this scenario, highlighting the work-flow between MOSt and xGSN, and the data-flow between *data producers* and *data consumers*.

The *virtual sensor developer* represents the capacity of implementing any KDDM process in the xGSN, registering and providing through MOSt *processing classes*, *wrappers*, and *virtual sensors*. The *data producer* represent the ability of *data producers* to register themselves into MOSt, providing information about their physical sensors, such as IP and observed properties, and privacy policies. In this paper, we intent to present the study and preliminary results of the viability to implement this new pSaaS paradigm using VSNs and its *on-demand* deployment.

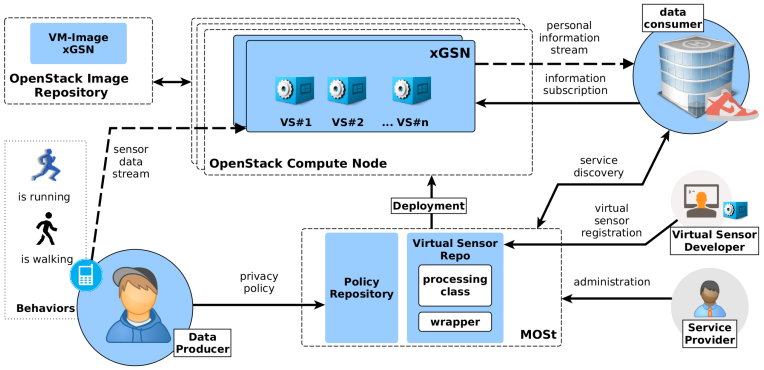

Figure 1: Privacy-aware Sensing as a Service Workflow

3 Related Works

Two approaches are commonly employed to address privacy in the IoT sensor stream: Privacy Preserving Techniques (PPTs) and Access Control Models (ACMs). PPTs, such as anonymization, can minimize the probability of extracting of sensitive personal information from the sensor data stream by degrading data utility. ACMs consist of authorization mechanisms that allow specifying policy conditions to grant or deny authorization to execute an operation over sensor data stream. However, while ACM approaches falls short to provide intermediary options between access grant and deny, PPT approaches are inefficient in the SSaaS and do not consider *contexts*. In general, PPTs are unable to decrease data utility only of parts of data stream that can be used to extract sensitive personal information. In both case, the trust model implemented on the *data consumer*'s side is doubtable and should be avoided. None of the related works implements a two-fold PEP process, neglecting the potential in preventing access according to the interpretation of the data stream and KDDM process. Besides that, the released data can be used to infer unintended personal information, even when anonymized.

From the several types of ACM available, the Attribute-based Access Controls (ABACs) [10] paradigm offers the finest-grained policy condition definition that allows defining policy conditions based on attributes (data stream samples), requester, and operations in a logic expression [11], which facilitate to express policy rules and control access in the IoT. Chronologically, first contributions for privacy in ubiquitous systems were proposed in [12] based on anonymous and secure connections to data source devices. In [13], the Butterfly framework is proposed aiming to apply PPTs over *generalizations*, such as patterns and association rules. Previously classified private *generalizations* are suppressed or degraded to enforce privacy. In [14], a privacy-aware framework provides an expressive policy and rule definition, encompassing not just the access request evaluation, but the subsequent usage of this information based on policy and rule conditions. The concepts of *data handling* and data handling policies allow specifying the strategy in terms of processing sphere (server-side, user-side, or customized), access level, meta-data type, restrictions based on provisioning (pre-conditions) and obligations (during and post-conditions). In this case, certified platforms guarantees that pos-conditions obligations are fulfilled. In [15], trusted KDDM applications are allowed to extract information using private datasets, which are encrypted in trusted Cloud platforms. The strategy consists in releasing only a subset of private datasets, dencrypting according to privacy categories defined in the Service Level Agreement (SLA), and allowing only authorized *data consumers* to have access to

the output. In [16], ontology and Semantic Web technology are incorporated to decide which PPT or ACM should be performed over a content defined according to a domain ontology. The privacy protection ontology contains concepts of threat, information, identifier, mechanism, protection mechanism, component, and protection component. The approach also conceives the concept of KDDM components that access personal information and generates a result item. The capacity of reasoning over the KDDM workflow in order to anticipate inference intention, however, is limited, making impossible to prevent unintended inference. In [17], the approach consists of a user controlled privacy preserved access control protocol, and a context-aware k-anonymity privacy application that are applied according to privacy protection mechanisms, secured by the exchange of public keys that allows specific data processing operations. The privacy policy is specified using external ontologies and relies on third-part definitions to determine PPT executions. A privacy preserving model based on VS is proposed in [18], providing a modular privacy by design solution. A PEP is implemented after the VS data processing execution, controlling release of private personal information that is produced or simply tunneled through the VS.

4 Privacy-aware Sensing as a Service and the Sensor-Cloud

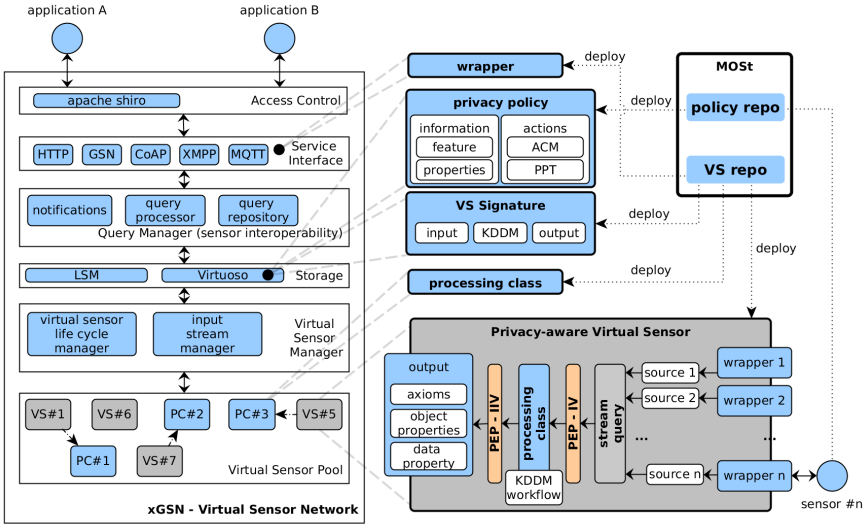

Figure 2: Privacy-aware xGSN

The concept of VS was originally defined in [19] and has evolved to address shortcomings inherently observed in physical sensors and its integration to the Sensor Web, such as limited storage and processing capacity, restricted energy autonomy, and connection protocols. Along with its capability to filter, VSs have become data processing units that ultimately intermediate data sources (connected *things*) and final IoT applications, constituting a new SSaaS paradigm that provides meaningful information, instead of raw data or data mining *generalizations*. Virtualized sensors in IoT platforms, such as the eXtended Global Sensor Network (XGSN) [8], implements the concept of VS. The xGSN relies on semantic representations of sensor and observation meta-data to implement the process of annotating and publishing sensor data on the

Sensor Web, developed in the context of the OpenIoT[1]. The system is constituted by the Global Sensor Network (GSN) middleware and the Linked Stream Middleware (LSM). The former provides Virtual Sensors (VSs) which offer several wrappers that implement IoT networks protocols, such as CoAP, XMPP, MQTT, and data processing capabilities through *processing classes*. The latter corresponds to the interface to create and request semantic annotations about VSs and observations. Along with GSN, LSM provides information to represent VS input, output, and processing class (KDDM process). In the OpenIoT project, the ontology used to represent VS input is the Semantic Sensor Network Ontology (SSN-O) [20].

Our privacy model is designed into the VS structure as a modular PEP formed by a two-fold verification step, performed before and after the KDDM step, as depicted in Figure 2. In our experiment, we restrict the KDDM processes in these VSs to perform only SPs, which guarantee that personal information will be produced along with its semantic annotation. Based on that, we propose to implement our pSaaS paradigm through the extension of the semantic representation of VS output; while using SPARQL Protocol and RDF Query Language (SPARQL) end-point of the xGSN to reason about privacy policy conditions into VS *on-the-fly*. Instead of field structures, described only by datatype, we extend the VS output specification to allow the semantic representation of any personal information using axioms, object properties and data properties of any domain-specific ontology. The **Inference Intention Verification** (IIV) step aims to anticipate the KDDM inference based on the VS output specification in order to apply PPTs over the input data stream or prevent the execution of VSs which produce only unauthorized outputs. The **Inference Verification** (IV) step aims to filter inferred personal information that is classified as private according to privacy policy conditions.

The GSN provides an initial set of *processing classes* that can be extended to implement any KDDM process, loading *jar* libraries available in the VSN. These *processing class jars* must be available during the VSs execution. If a specific *wrapper* is required to connect a physical sensor, except those already available in the platform, the default *wrapper* can also be extended. We also reuse the environment defined in the OpenIoT project where the ontology for sensor annotation is the Semantic Sensor Network (SSN) and the Virtuoso[2] database is the relational database, quad-store back-end, and SPARQL end-point.

The injection of VS semantic representation file, along with its dependencies of *processing class* and *wrapper*, is realized by MOSt during the Sensor-Cloud deployment process. MOSt is a placement and post-configuration system responsible for managing the deployment of compound services in a federated Cloud environment, which relies on the Iterative Graph Mapping (IGM) placement algorithm described in [21] to optimize resources allocation in Cloud infrastructure. It uses Open Cloud Computing Interface (OCCI) infrastructure specifications [3] to describe the application requirements in JSON. Once MOSt receives the manifest, it processes it to find the optimal placement of different services in the infrastructure. In our experiment, MOSt executes the process of deployment of xGSN middleware, after injecting the all dependencies of VS. Moreover, it connects the xGSN instances to appropriate services in the Cloud that request the Sensor-Cloud services.

5 Experiment

In order to test our solution, an experimental *testbed* is being built. The experimentation is split in two sub-experiments: i) evaluation of the pSaaS implemented in the xGSN platform; and ii)

[1] http://www.openiot.eu/
[2] http://virtuoso.openlinksw.com/
[3] http://occi-wg.org/about/specification/

evaluation of the deployment of this new platform using MOSt. In the first sub-experiment, we intent to compare the execution time of traditional VSs against our pSaaS paradigm. In this case, SPARQL is used to infer about privacy policy conditions defined using semantic annotation. For each PEP, one SPARQL query is executed. The ontological framework for reasoning about the Privacy Policy Condition (PPC), stored in the Virtuoso quad-store, are: (1) a domain specific personal information ontology (PEO); privacy enhancing technology (PET) ontology; and (2) the Semantic Sensor Network Ontology (SSN-O). In addition, the RDF graphs created during the deployment process will constitute the facts on which the PPC are evaluated. An example of SPARQL query is illustrated in Listing 1, using SSN-O, PEO, and PET axioms.

Listing 1: SPARQL-DL query to retrieve privacy policy conditions

```
SELECT * WHERE {{ ?iPPT a pet:PrivacyPreservingTechnique.
 ?iPPT ssn:ofFeature ?iAxiom.
 ?Axiom rdfs:subClassOf* peo:personalInformation; rdfs:subClassOf+ ssn:FeatureOfInterest.
 ?iObservationValue ?dataProperty ?Axiom. ?dataProperty a owl:DataProperty.
 ?iSensorOutput ^ssn:isRegionFor ?iObservationValue; ssn:isProducedBy ?iVSensor.
 } FILTER (?iVSensor = <:ID-Virtual-Sensor>) }
```

For the matter of brevity, the structure of these ontology are not described in this paper. An on-going ontology project developed in the context of this research can be accessed in https://github.com/thiagomoreirac/opis, containing ontologies developed for this purpose. We refrain to explain that, in this example, an individual of a specific VS class is extended from the ssn:sensor class, which produces a sensor output. This output has a region defined according to data properties that point to IRIs. Each of these IRIs represent axioms to personal information, which, in turn, are extended from the SSN-O feature of interest. In the Listing 1, the PPT is defined in the PET ontology, and it is associated to a personal information through a ssn:ofFeature, that defines the relationship between any entity and the feature of interest. Each result for this query means that a privacy policy condition is valid, and therefore, the associated privacy technique should be executed, either preemptively – in the Inference Intention Verification step – or posteriorly – in the Inference Verification step. We expect to compensate the gain of executing selectively and intermittently PPTs on the sensor data stream with extra execution time caused by the SPARQL queries and execution prevention of VSs whose output are sensitive, and therefore, denied to be release. In the second part, the objective is to implement an automated process to deploy a compound service in Sensor-Cloud that requires a specific sensor data (feature, property, datatype, conditions) using MOSt, as depicted in Figure 3. Hence, the configuration of one or more VSs for each request is retrieved from the VS repository in MOSt based on its input and output defined in each manifest. Then, based on the requested input and conditions for physical sensor selection, MOSt generates a deployment VS file with links to physical sensors and *data consumers*. Lastly, MOSt triggers VS registration using LSM in the xGSN virtual machine, and then initialization of xGSN service.

6 Preliminary results and Future Works

We ran preliminary queries using a virtual machine provided by the OpenIoT project, which has the Virtuoso 6 as back-end for SPARQL queries. We simulated the execution of the SPARQL queries to understand the footprint of the most important part of our pSaaS. The complexity analysis of its algorithms are directly related to query response times and the size of the SP outputted by the VS. Since these outputs are restricted to be atomic, similarly to the SSN-O design, the size of SP should not impact in the overall response time and should be compared to a constant complexity. The preliminary results shows that the response time was directly

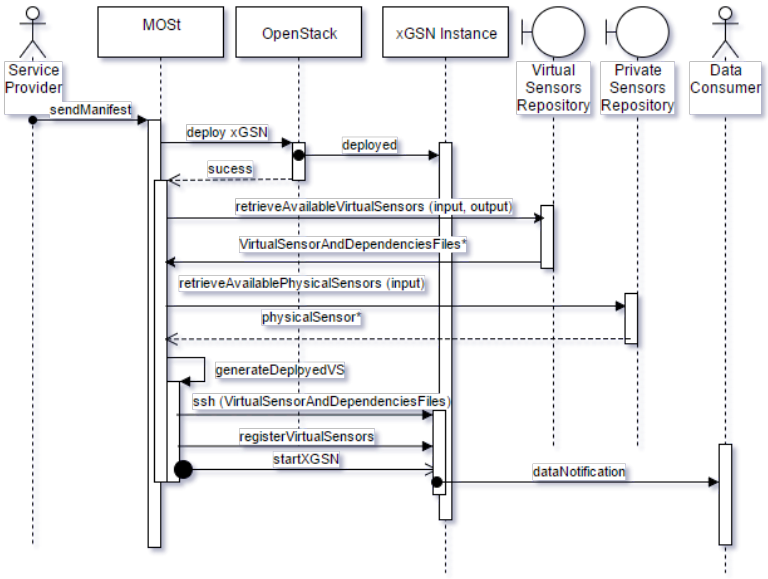

Figure 3: MOSt sequence for deploying, configuring, and launching pSaaS in xGSN

related to the size of the RDF pattern graphs used to request the result set. However, even thought we do not address this issue in our approach, query performance can be addressed by using SPARQL query caching [22]. This meet our privacy model requirements, once the results of those queries achieves a constant if no new virtual sensor or PPC is installed in the system. A further investigation with a set of virtual sensors, using different ontologies for Privacy-Enhancing Technology (PET) and personal information should be carried in the future.

ACKNOWLEDGMENT

This work was supported by *CAPES Foundation, Ministry of Education of Brazil* and VNET AmSud #3647WL Project. Thanks to all the partners who helped with discussions to improve this work.

References

[1] Shancang Li, Li Da Xu, and Shanshan Zhao. The internet of things: a survey. *Information Systems Frontiers*, 17(2):243–259, apr 2015.

[2] Atif Alamri, Wasai Shadab Ansari, Mohammad Mehedi Hassan, M. Shamim Hossain, Abdulhameed Alelaiwi, M. Anwar Hossain, Atif Alamri, Wasai Shadab Ansari, Mohammad Mehedi Hassan, M. Shamim Hossain, Abdulhameed Alelaiwi, and M. Anwar Hossain. A Survey on Sensor-Cloud: Architecture, Applications, and Approaches, A Survey on Sensor-Cloud: Architecture, Applications, and Approaches. *International Journal of Distributed Sensor Networks, International Journal of Distributed Sensor Networks*, 2013, 2013:e917923, 2013.

[3] S Abdelwahab, B Hamdaoui, M Guizani, and T Znati. Cloud of Things for Sensing as a Service: Sensing Resource Discovery and Virtualization. *2015 IEEE Global Communications Conference (GLOBECOM)*, pages 1–7, 2015.

[4] FTC. IoT Privacy & Security in a Connected World. (January):71, 2015.

[5] Alessandro Acquisti, Laura Brandimarte, and George Loewenstein. Privacy and human behavior in the age of information. *Science*, 347(6221):509—-514, 2015.

[6] Cory Andrew Henson. *A Semantics-based Approach to Machine Perception*. Doctoral dissertation, Wright State University, 2013.

[7] Emanuel Coutinho, Marcelo Santos, Stenio Fernandes, José Neuman de Souza, Thiago Moreira da Costa, Elie Rachkidi, Nazim Agoulmine, and Javier Baliosian. Research Opportunities in an Intercloud Environment Using MOSt in SLA4CLOUD Project. *Advance*, 2015.

[8] Jean-paul Calbimonte, Sofiane Sarni, Julien Eberle, and Karl Aberer. XGSN: an open-source semantic sensing middleware for the web of things. In *Joint Proceedings of the 6th International Workshop on the Foundations, Technologies and Applications of the Geospatial Web, TC*, pages 51—-66. 2014.

[9] Seda Gürses and Jose M. Del Alamo. Privacy Engineering: Shaping an Emerging Field of Research and Practice. *IEEE Security and Privacy*, 14(2):40–46, 2016.

[10] D R Kuhn, E J Coyne, and T R Weil. Adding Attributes to Role Based Access Control. *Computer*, 43(6):79–81, 2010.

[11] Vincent C. Hu, D. Richard Kuhn, and David F. Ferraiolo. Attribute-based access control. *Computer*, 48(2):85–88, 2015.

[12] Marc Langheinrich. A Privacy Awareness System for Ubiquitous Computing Environments. In *UbiComp '02: Proceedings of the 4th international conference on Ubiquitous Computing*, pages 237–245. 2002.

[13] Ting Wang and Ling Liu. Butterfly: Protecting Output. *ICDE*, pages 1170–1179, 2008.

[14] C. A. Ardagna, M. Cremonini, S. De Capitani Di Vimercati, and P. Samarati. A privacy-aware access control system. *Journal of Computer Security*, 16(4):369–397, 2008.

[15] Wassim Itani, Ayman Kayssi, and Ali Chehab. Privacy as a service: Privacy-aware data storage and processing in cloud computing architectures. In *8th IEEE International Symposium on Dependable, Autonomic and Secure Computing, DASC 2009*, pages 711–716, 2009.

[16] Martin Kost and JC Freytag. Privacy Analysis using Ontologies. *... on Data and Application Security and Privacy*, pages 205–216, 2012.

[17] Xin Huang, Rong Fu, Bangdao Chen, Tingting Zhang, A.W. W Roscoe, Xin Huang, Rong Fu, Bangdao Chen, Tingting Zhang, and A.W. W Roscoe. User interactive Internet of things privacy preserved access control. In *Proceedings of the 2012 International Conference of Internet Technology and Secured Transactions*, pages 597–602, 2012.

[18] Evangelos Pournaras, Izabela Moise, and Dirk Helbing. Privacy-preserving ubiquitous social mining via modular and compositional virtual sensors. *Proceedings - International Conference on Advanced Information Networking and Applications, AINA*, 2015-April(October 2014):332–338, 2015.

[19] Wb Heinzelman, Al Murphy, H.S. Carvalho, and M.A. Perillo. Middleware to support sensor network applications. *IEEE Network*, 18(1):6–14, jan 2004.

[20] Michael Compton, Payam Barnaghi, Luis Bermudez, Raúl García-Castro, Oscar Corcho, Simon Cox, John Graybeal, Manfred Hauswirth, Cory Henson, Arthur Herzog, Vincent Huang, Krzysztof Janowicz, W. David Kelsey, Danh Le Phuoc, Laurent Lefort, Myriam Leggieri, Holger Neuhaus, Andriy Nikolov, Kevin Page, Alexandre Passant, Amit Sheth, and Kerry Taylor. The SSN ontology of the W3C semantic sensor network incubator group. *Journal of Web Semantics*, 17:25–32, 2012.

[21] Khanh-Toan Tran, N. Agoulmine, and Y. Iraqi. Cost-effective complex service mapping in cloud infrastructures. In *Network Operations and Management Symposium (NOMS), 2012 IEEE*, pages 1–8, April 2012.

[22] Nikolaos Papailiou, Dimitrios Tsoumakos, Panagiotis Karras, and Nectarios Koziris. Graph-Aware , Workload-Adaptive SPARQL Query Caching. *Sigmod*, pages 1777–1792, 2015.