What does this device do?

Edoardo Pignotti
Computing Science & dot.rural
Digital Economy Hub
University of Aberdeen
Aberdeen AB24 5UA, UK
e.pignotti@abdn.ac.uk

Stanislav Beran
Computing Science & dot.rural
Digital Economy Hub
University of Aberdeen
Aberdeen AB24 5UA, UK
s.beran@abdn.ac.uk

Peter Edwards
Computing Science & dot.rural
Digital Economy Hub
University of Aberdeen
Aberdeen AB24 5UA, UK
p.edwards@abdn.ac.uk

ABSTRACT

This paper describes the Trusted Tiny Things project which is investigating some of the challenges inherit in making the Internet of Things (IoT) more transparent to users. We present a semantic framework for reasoning about the capabilities of IoT devices based on provenance information collected from devices and their associated services. As part of this framework we have developed a semantic model, services and a smartphone app to represent, store and query IoT provenance. The semantic model and app was informed via a series of participatory design activities with users. In this paper we discuss the use of the system with two distinct IoT devices: an NFC tag used at bus stops to provide a means to access real-time bus timetables, and a black-box device installed into vehicles by insurance companies to track driving behaviour.

Keywords

IoT, provenance, transparency

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Miscellaneous ; H.1 [**Information Systems**]: Models and Principles

General Terms

Theory, Documentation, Management

1. INTRODUCTION

The vision of the Internet of Things (IoT) is a dynamic global network based on standard and interoperable communication protocols where physical and virtual 'things' have identities, physical attributes, and capabilities and are seamlessly integrated into the existing internet infrastructure [4]. The IoT is thus built upon a range of sensors and other devices that together represent the 'things'; these devices range from passive radio tags to internet connected sensor platforms and embedded computers. Deployments of such devices in urban spaces are increasingly commonplace. For example, passive NFC (Near Field Communication) tags are currently in use by Aberdeenshire Council in Scotland to provide smartphone access to timetable information for a particular bus stop. Active IoT devices include the in-car black boxes [6] being introduced by insurance companies to assess the behaviour of drivers and affect their premiums, and smart meters providing information about electricity consumption to energy suppliers. Such applications raise a number of issues, not least of which is the extent to which users understand these devices and their capabilities. Questions that a user might like to ask include: *What kind of data does the thing collect? Is the data transmitted? If so, how and to whom? For what purposes are the data used? What control do I have over any aspects related to the generation and use of this data?*. These questions are reflected in the "TRUSTe Internet of Things Privacy Index - GB Edition[1]" study where more than 80% of the 2,005 people interviewed were concerned about such issues. Similar issues are also discussed in Vermeulen et al. 2010 [11] where it was argued that allowing users to pose why and why not questions about context-aware systems resulted in better and stronger feelings of trust. We argue that creating an infrastructure that would enable users to interrogate IoT devices to discover this (and other) information would significantly increase trust in such devices and would provide reassurance to users (and wider society) as the roll-out of such technologies accelerates.

In this paper we discuss how the "Trusted Tiny Things" (T^3) project is investigating some of these issues by proposing a solution based on metadata describing the context surrounding devices (e.g. manufacturer, owner, data transfer method). We argue that this can be provided by publishing information about devices according to the linked data principles [3]. As 'things' become more interconnected this context should also include provenance information: a record of the entities (devices or services) and processes (data transmission, data analysis, decision making) involved in the creation and use of data. A formal representation of provenance has been identified as essential to support users (and machines) to better understand and trust data [8]. For example, in the car black box scenario, provenance could be used in order to understand what kind of data the box is collecting, what agents or services are using this data, and for what purpose.

Urb-IoT'14 October 27-28 2014, Rome, Italy

http://dx.doi.org/10.1145/2666681.266669.

[1]https://www.truste.com/gb-internet-of-things-index-2014/

In the remainder of this paper we introduce two transport-related case studies that are being investigated in the context of urban spaces. We continue by discussing a trusted things framework based on Semantic Web technologies and provenance. We describe the components of the system architecture used to reason about the capabilities of IoT devices and to present this information to the user via a smartphone app. We conclude by highlighting the potential impact of the project and our future plans.

2. CASE-STUDIES

The T^3 project is considering two case-studies, the first of which relates to the deployment of passive NFC tags by Aberdeenshire Council to provide smartphone access to timetable information for bus stops. Passengers interested in obtaining real-time bus information can scan the NFC tag with a capable smartphone. The NFC tag embeds a URL containing a unique ID identifying the bus stop. This URL is used to redirect the smartphone web browser to a third party website displaying live timetable information. This scenario raises some questions regarding the privacy of the user. For example: *Where is the web browser re-directed? Who is running the service? What information are they collecting from my smartphone? Are there any charges for the service? Would the service take contact-less payments?* (A feature normally associated with NFC technology). In its current form, the existing service does not provide the information required to answer the questions above. For example, the user is not aware that their web browser is redirected to a website (*www.rslpublic.co.uk*) and that the service is not managed by Aberdeenshire Council. There is also no explicit indication that the service is free of charge. Furthermore, the user is unaware that the service is collecting the IP address of the device used and details about the smartphone's operating system and version.

The second case-study focuses on the use of in-car black boxes. Such boxes record information about driving style and location of the vehicle using a range of sensors including GPS and accelerometers. The information captured by the sensors is then transmitted to an insurance company, typically via a 3G connection. This scenario also raises a number of questions regarding privacy. For example: *What kind of data is being recorded? When and where is the data transmitted? Who is using the data? Is the data being sent to other third-party companies? For what purposes?* Current in-car black boxes solutions do not allow users to find information on how the data is transmitted and used by the insurance company. For example, the user might want to be informed if some of the data collected by the device is shared with third party companies, e.g. a car manufacturer.

3. PARTICIPATORY DESIGN ACTIVITIES

As a result of a collaborative R&D roadmapping activity between the UK's Technology Strategy Board and the UK Research Councils conducted in 2012, a report was produced highlighting the priorities for research and innovation in the IoT. The report [10] identified several priorities in this area including the need to understand how researchers, developers and end users can become involved in co-designing IoT services, especially with respect to information interfaces. In the T^3 project we have embraced this idea by involving users in co-design of certain aspects of our semantic models and software applications.

To date, we have conducted a number of participatory design events involving a total of 77 participants with different technological backgrounds. The events involved the participants themselves determining the direction of the group discussion, through their answers to some initial questions. Each event lasted for 90 minutes and was divided into two stages: We began by exploring with participants the capabilities of IoT devices. They were presented with a number of pictures illustrating IoT devices such as an internet enabled alarm clock, a telemetry blackbox, a number of smart appliances, the NFC tags used at bus stops and an internet enabled toy. Questions were then posed such as: *What do you think are the capabilities of this device?* and *What kind of capabilities would you want to be aware of before interacting with this kind of device?*. Information points and other thoughts were captured on post-it notes by participants themselves and by the event facilitator. In the second stage we asked participants to design a mobile app to visualise the kinds of information identified during the earlier stage. Participants recorded app design ideas on A3 sheets of paper, using words or images as they preferred. An example of the material generated by the participatory design activities is presented in Figure 1. The data extracted from the first stage was categorised in a process where similar statements were clustered together, and subsequently formed into categories describing attitudes towards the capabilities of IoT devices. The categories identified were as follows:

- **Understanding Who** - It is important to understand who controls the devices and who has access to the data generated (people, organisations or software agents). It is also important to make explicit if the owner of the device is different from individuals and organisations that have access to the data.
- **Understanding Why** - It is important to know for what purposes personal data are used (e.g. statistics, quality of service, advertisement, etc.)
- **Accessing Data** - It is important to have access to personal data generated by the devices.
- **Exercising Control** - It is important to know if it is possible to limit the data being sent by devices and if it is possible to turn the device off (and how to do this).
- **Receiving Notifications** - If the capabilities of the device change, users should be notified and additional consent should be required.

In the subsequent analysis, the key categories identify during the study were translated into requirements that informed the the components of a provenance framework for IoT devices including the rules required to infer capabilities of devices (described in Section 4). The app design ideas identified during the second stage were also categorised and used to inform the development of a prototype mobile app (described in Section 5).

4. DESIGNING A PROVENANCE FRAMEWORK FOR IOT DEVICES

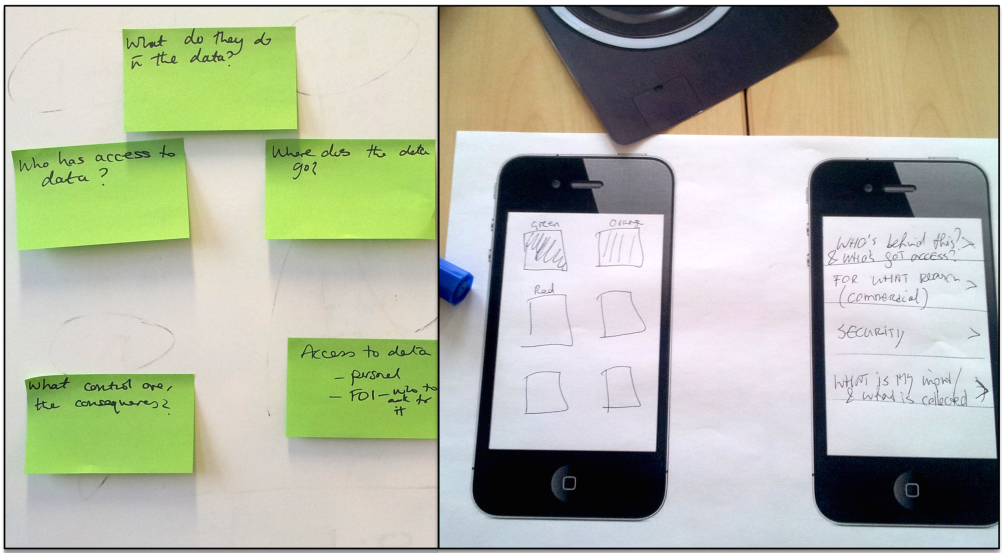

Figure 1: An example of the material generated by the participatory design activities.

The Semantic Web[2] is a vision in which today's Web will be extended with machine readable content, and where every resource will be marked-up using machine readable meta-data; a family of XML-based technologies, most notably the Resource Description Framework[3] (RDF) provide a mechanism for representing that meta-data. The Web Ontology Language[4] (OWL) is used to capture the meaning of meta-data terms and their interrelationships.

In order to describe the provenance of IoT devices we have developed a semantic framework able to characterise information such as: capabilities, security properties, ownership and provenance of devices and their use. This framework is summarised in Figure 2.

4.1 Linking Physical and Digital Entities

In order to retrieve information about IoT devices (characteristics, provenance, capabilities, etc.) it is necessary to be able to identify things (e.g. bus stop, fridge) and their IoT components (tag, device, sensor or service). Kortis et al. [7] describe an ontology that represents knowledge about 'Things' in the IoT domain and the way they should interoperate. The authors have created a model describing IoT concepts by introducing ontological definitions such as Physical Entity, Control Entity, Electronic Device, Smart Entity Cluster and Smart Network. However, this ontology is focused on finding a common framework to allow deployment of IoT devices into the existing Internet infrastructure for service discovery and it is not suitable for our needs as it is too focused on low level service descriptions which do not align with the requirements gathered from our participatory design activities. The Internet of Things Architecture[5](IoTa) is another project working towards building a common architecture for the future Internet of Things. They have developed a conceptual model [2] to describe the IoT domain based on previous work from Serbanati et al. [9] and Haller [5]. The main aim of this model is to characterise the different entities in the IoT domain (e.g. User, Service, Device, Physical Entity, Virtual Entity and Resource). We have created our own OWL ontology of the conceptual model introduced by the IoTa project. This ontology describes the the following concepts:

- *iota:PhysicalEntity* representing physical objects such as a bus stop, a fridge, etc.
- *iota:Device* describing an Internet of Things device such as an embedded computer, a sensor, an actuator, etc. Devices can contain other devices (e.g. a car blackbox contains different sensors) and this is described using the *iota:contains* property. An *iota:Device* can represent an *iota:PhysicalEntity* using the *iota:represents* property.
- *iota:Tag* is a special kind of *iota:Device* that can be attached to an *iota:PhysicalEntity* (denoted using the *iota:attachedTo* property) and identifies an *iota:Device* using the *iota:identifies* property. An *iota:Tag* can be used to represent an NFC tag or an RFID tag.

An extract of this ontology is illustrated in Figure 2 (top left).

4.2 Provenance of IoT Devices

During the participatory design events participants identified the need to make certain information about IoT devices transparent such as who controls the device, who uses the data, and for what purposes. This is consistent with the findings of Weber et al. [12] who argue that transparency allows for a certain level of "democratic" legitimisation and predictability through active involvement of citizens as well as through certain control over the decision-making process. Bandara et al. [1] proposed a first semantic model for describing devices. While this model is capable of describing device characteristics, it is not capable of capturing crucial provenance information such as the processes associated with the devices and what human or computational entities are involved.

[2]http://www.w3.org/2001/sw/
[3]http://www.w3.org/RDF/
[4]http://www.w3.org/TR/owl-features/
[5]http://www.iot-a.eu

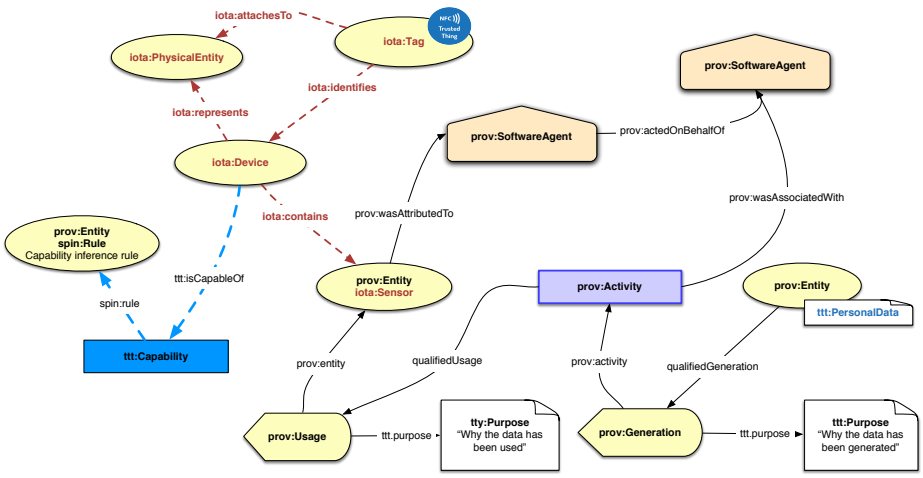

Figure 2: An extract of the ontological framework describing the provenance of IoT devices.

In our model we decided to describe provenance using the emerging W3C PROV-O[6] ontology as it is designed to be applicable to a wide range of applications and domains. PROV-O defines concepts such as: *prov:Entity* (physical, digital, conceptual); *prov:Activity* (something that occurs over a period of time and acts upon or with entities); and *prov:Agent* (something that bears some form of responsibility for an activity). Using this ontology it is possible to describe how, for example, data from a device was produced (e.g. a location *prov:Entity* was generated by a Position Calculation *prov:Activity* using a GPS sensor) and who used the data and for what purpose (e.g. the location observation was used by the insurance company *prov:Agent* to determine if a car is kept on the street at night). The diagram in Figure 2 illustrates how the PROV-O model has been integrated with the IoTa ontology. In this example an *iota:Sensor* is also characterised as a *Entity* in PROV. The is attributed to a *prov:SoftwareAgent* using *prov:wasAttributedTo*. The graph also shows that an *Activity* (associated with a different agent) uses information from the sensor to generate another entity.

When managing provenance of IoT devices it is not always possible to instrument devices and services to generate information about their usage and operation (retrospective provenance). In some cases, manufacturers can only provide information on how devices are intended to operate (prospective provenance). In our provenance framework we therefore make provision for both kinds of provenance.

4.3 Inferring Device Capabilities

Guided by user requirements we have designed an ontology to support inferences about device capabilities using provenance described according to the PROV-O and IoTa ontology. This ontology (referred as the T^9 ontology) provides the metadata and supporting logic required in order to determine the capabilities of a device. Firstly, the ontology provides annotations over the provenance of devices capturing the kind of information identified by our participants. These annotations include:

- The *ttt:PersonalData* class is used to identify if a *prov:Entity* represents information that can be associated with a particular individual.

- The *ttt:Purpose* class is used to provide an explanation of why certain entities (described as personal data) are being generated or used. Using the *ttt:purpose* property it is possible to associate *prov:Usage* and *prov:Generation* qualified relationships with a description of purpose.

- The *ttt:Capability* class defines different kinds of capabilities (e.g. *ttt:DataConsumption*, *ttt:DataGeneration* and *ttt:DataSharing*) that can be associated with *iota:Devices*. These associations (described by the *ttt:isCapableOf* property) are made on the basis of a number of inference rules.

In order to infer the capabilities of IoT devices using our ontological framework we can associate rules to specific classes of *ttt:Capability*. We make use of the SPIN ontology[7] to support the use of SPARQL to specify rules and logical constraints necessary to reason about capabilities. The SPIN ontology allows SPARQL queries to be represented in RDF and associated to classes in an ontology using a pre-defined *spin:rule* property that can be used to specify inference rules using SPARQL CONSTRUCT, DELETE and INSERT statements. Figure 3 (left box) shows an example of such a rule for the *ttt:DataConsumption* class. The rule is designed to traverse a PROV-O provenance graph starting from an instance of an *iota:Device* and looks for activities that have used or generated entities classified as personal data. Once such activities have been identified the rule specifies how an annotation about the data consumption capability is generated, including a link to the agent responsible for the activity and the specific purpose. In this ontology we have also specified two rules (Figure 3 top right and bottom right boxes) that are used to determine what provenance has been used to infer a specific device capability. The ontologies described in this section are both available on GitHub[8].

[6]http://www.w3.org/TR/prov-o/

[7]http://spinrdf.org/spin.html

[8]http://t3.abdn.ac.uk/ontologies/t3.owl
http://t3.abdn.ac.uk/ontologies/iota.owl

```
CONSTRUCT {
   _:b0 a :DataConsumption .
   _:b0 :consumes ?data .
   _:b0 :consumer ?agent .
   _:b0 :purpose ?purposeDescription .
   ?this :isCapableOf _:b0 .
}
WHERE {
   ?virtualentity iota:represents ?this .
   ?activity (prov:qualifiedUsage)+ ?usage .
   ?usage prov:entity ?data .
   ?data a :PersonalData .
   ?usage :purpose ?purposeDescription .
   ?activity (prov:wasAssociatedWith)+/prov:actedOnBehalfOf ?agent .
   ?agent a foaf:Organization .

   NOT EXISTS {
      ?this :isCapableOf ?capability .
      ?capability :purpose ?purposeDescription .
      ?capability :consumer ?agent .
      ?capability :consumes ?data .
   } .
}
```

```
CONSTRUCT {
   ?device :prospectiveCapability ?capability .
}
WHERE {
   ?device prov:wasAttributedTo ?agent .
   ?agent :qualifiedFollow ?follow .
   ?follow :shouldGenerate ?bundle .
   ?device :isCapableOf ?capability .
   ?capability :consumes ?data .
   ?bundle :contains ?data .
}
```

```
CONSTRUCT {
   ?device :retrospectiverCapability ?capability .
}
WHERE {
   ?device prov:wasAttributedTo ?agent .
   ?bundle a prov:Bundle .
   ?bundle prov:wasAttributedTo ?agent .
   ?device :isCapableOf ?capability .
   ?capability :consumes ?data .
   ?bundle :contains ?data .
}
```

Figure 3: Example of device capability inference rule (left box) and two rules used to distinguish between prospective and retrospective provenance (top right and bottom right boxes).

5. THE T^3 SYSTEM

In order to support our semantic framework we have developed a software infrastructure and a mobile app (see Figure 4) that can be used to query, update and register IoT devices and to notify the user of any changes in the capabilities of a registered device. We have created a custom NFC tag (branded as "Trusted Thing") that can be attached to any physical entity so it can be identified in our system. This tag also serves as an indication that a "Thing" (physical entity) is part of the IoT and it can be interrogated using our system. The app[9] continuously monitors events generated by the NFC sensor on the phone and detects if one of our custom tags has been scanned. This initiates a connection to our services (hosted at http://t3.abdn.ac.uk) to retrieve information about the device associated with the tag. If the user has not interacted with the device before, information about the device is presented via a mobile client interface (see Figure 4). The interface consists of a brief description of the device at the top of the screen followed by an infographic representing the user, the personal information he/she will be sharing and with whom. In the example provided in Figure 4, if the user decides to interact with the device he/she will be sharing four items of personal information (document icon) with three organisations (group icon). Clicking on each of those icons the user is able to obtain additional details via a popup box such as the contact details of the organisations or the type of personal information being shared. For example, in the bus stop scenario one of the pieces of information being shared is the IP address of the user's mobile phone. The app also presents a list summarising the device capabilities (e.g. Personal Data Consumption). Clicking on any item in this list provides the user with a different screen showing details about a capability (e.g. what personal data has been collected, who is consuming the data and for what purposes). If the user is comfortable with the capabilities of the device he/she can click the Accept button or otherwise they will be able to click Decline. Either choice is recorded in our system for future interactions including giving the ability to our system to provide notifications if the capabilities of a device have changed.

Information about IoT devices in our system are stored in the form of RDF statements described by the ontological framework presented in Section 4. This metadata is stored in a OpenRDF Sesame[10] triplestore. Additionally, we utilise a MySQL database server to store other information such as the ID associated with a user's smartphone, a list of IDs representing the devices the user has interacted with, etc.

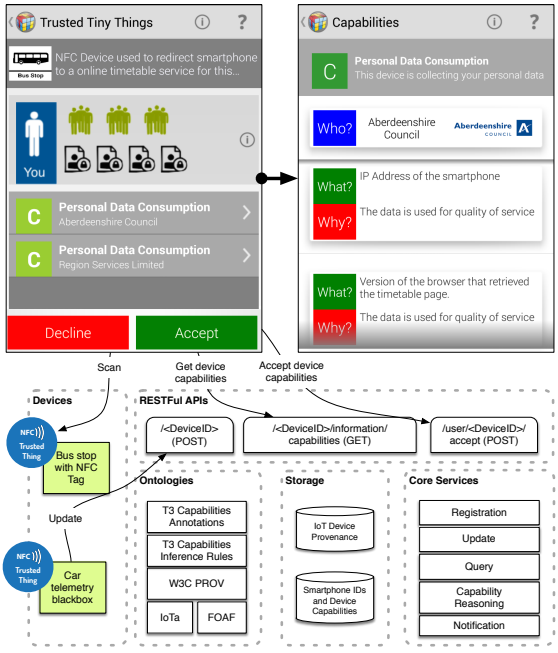

Figure 4: The T^3 system architecture (bottom) and the client smartphone app (top)

The system consists of a number of Java services which are accessible from our mobile app via a RESTful API and JSON[11] is used as the data interchange format. Devices in

[9]The app is called "Trusted Tiny Things" and is available on the Google Play store.

[10]http://www.openrdf.org

[11]http://www.json.org/

our API are recognised by a custom URL *http://<domain>/devices/{DeviceID}* where the DeviceID represents the device identifier encoded in our Trusted Thing NFC tag. Using the devices URL we can support different types of GET and POST actions to retrieve or create information about devices.

6. CONCLUSIONS & FUTURE WORK

The project outlined in this paper is investigating how Semantic Web technologies can be used to manage information about IoT devices so that their capabilities are transparent to users. This should allow users to make informed decisions about the trustworthiness of such devices based on their provenance.

During the participatory design activities we learnt what users want to know before interacting with IoT devices. This is summarised in five high-level requirements: understanding who controls the device and has access to the data; understanding why the data is used; being able to access personal data; being able to exercise control over the device; and receiving notifications if the capabilities of a device change. We argue that in order to create a "transparent" IoT ecosystem it is important to consider how to capture provenance, how to monitor activities within the ecosystem and how to control the behaviour of devices so the five requirements highlighted above can be fulfilled.

In order to demonstrate how such an ecosystem can be created we have designed a semantic software infrastructure and a custom Trusted Thing NFC tag that can be attached to any IoT physical entity so it can be identified in our system. We have built an app and supporting services that can provide users with information about the capabilities of IoT devices (e.g. what personal data is used, by whom and for what purposes). We are in the process of evaluating our solution with real users based on the bus stop scenario as the tag is currently deployed across 2400 bus stops in the Aberdeen and Aberdeenshire region.

While the Trusted Tiny Things infrastructure is designed to provide greater transparency about the capabilities of IoT devices, there are several limitations to our approach. One of these is being able to ensure that the provenance about devices and services used to derive capabilities is truthful. We provide a set of rules in the form of guidelines[12] for registering and managing information about IoT devices in the Trusted Tiny Things system. However, such rules will have to be enforced by a trusted authority in order to guarantee the reliability of the provenance represented in our system. Understanding the role of this authority in policing such a system will require further research.

In the future we are interested in exploring how policy-based reasoning can be used to control the behaviour of active IoT devices. In an urban environment where potentially hundred of devices could have access to information about people, it is important that a user is able to specify high-level privacy preferences, and that devices are able to enact such preferences by providing or denying access and notifying users of violations. Such preferences can be represented in the form of policies. For example, if a car's black-box is capable of sending real-time information to the insurer and other third parties, a user might specify that he/she is only willing to send information to the insurer but not to third parties. We are interested in exploring how policy-based reasoning can be used to control the behaviour of active IoT devices in such situations.

This research should help to stimulate debate about transparency of IoT devices deployed in urban spaces. This work also has the potential to influence Web standards including the W3C Semantic Sensor Network ontology and the W3C Provenance recommendations by providing a number of real-life application scenarios.

[12]http://t3.abdn.ac.uk/guidelines/

Acknowledgements

This research is supported by the UK Research Councils Digital Economy IT as a Utility Network+ (EP/K003569/1) and the dot.rural Digital Economy Hub (EP/G066051/1).

7. REFERENCES

[1] A. Bandara, T. R. Payne, D. de Roure, and G. Clemo. An ontological framework for semantic description of devices. *McIlraith, S.A., Plexousakis, D., van Harmelen, F. (eds.) ISWC 2004. LNCS, vol. 3298, Springer, Heidelberg*, 2004.

[2] A. Bassi, M. Bauer, M. Fiedler, T. Kramp, R. van Kranenburg, S. Lange, and S. Meissner. *Enabling Things to Talk*. Springer, 2013.

[3] C. Bizer, T. Heath, and T. Berners-Lee. Linked data - the story so far. *International Journal on Semantic Web and Information Systems (IJSWIS)*, 5(3):1–22, 2009.

[4] S. Gusmeroli, S. Haller, M. Harrison, K. Kalaboukas, M. Tomasella, O. Vermesan, H. Vogt, and K. Wouters. *Vision and Challenges for Realising the Internet of Things*. European Commission, 2010.

[5] S. Haller. The things in the internet of things. In *Internet of Things Conference, Tokyo, Japan, November 2010*, 2010.

[6] E. Hossain, G. Chow, V. C. M. Leung, R. D. McLeod, J. Mišić, V. W. S. Wong, and O. Yang. Vehicular telematics over heterogeneous wireless networks: A survey. *Comput. Commun.*, 33(7):775–793, May 2010.

[7] K. Kotis and A. Katasonov. An iot-ontology for the representation of interconnected, clustered and aligned smart entities. Technical report, VTT Technical Research Center, Finland VTT Technical Research Center, Finland, 2012.

[8] L. Moreau. The foundations for provenance on the web. *Found. Trends Web Sci.*, 2(2-3):99–241, Feb. 2010.

[9] A. Serbanati, C. M. Medaglia, and U. B. Ceipidor. Building blocks of the internet of things: State of the art and beyond. *Deploying RFID-Challenges, Solutions, and Open Issues, Dr. Cristina Turcu (Ed.), InTech*, 2011.

[10] R. Tafazolli, C. Upstill, H. Aghvarni, R. Cooper, and W. Dutton. A roadmap for interdisciplinary research on the internet of things. Technical report, Swindon, GB, Technology Strategy Board, 2012.

[11] J. Vermeulen, G. Vanderhulst, K. Luyten, and K. Coninx. Pervasivecrystal: Asking and answering why and why not questions about pervasive computing applications. *IE*, pages 271–276, 2010.

[12] R. H. Weber and R. Weber. *Internet of Things*. Springer, 2010.