IETF IAB Semantic Interoperability Workshop

The Role of Meta-data: Enabling Trust in the IoT

Eve M. Schooler, David Zage, Sung Lee, Moreno Ambrosin, Andrew Brown, Jeff Sedayao
{eve.m.schooler, david.zage, sung.lee, moreno.ambrosin, andrew.brown, jeff.sedayao}@intel.com
Research in Emerging Architectures, Internet of Things Group, Intel Corporation

As part of the IoT business unit in Intel, the Research in Emerging Architectures team is focused on building out a communications, storage and security architecture and infrastructure to support trusted analytics at the network edge. Why is this an important pursuit? How does it relate to semantic interoperability and in particular to the trust needed between "Things" in the IoT? We answer these questions below, and in the process underscore the need for the workshop to include practitioners who are designing, building, deploying and using IoT systems, to better understand the requirements for IoT semantic interoperability.

Trusted analytics at the network edge. The IoT generates a tremendous amount of data at the network's edge. Increasingly, that data is unable to be adequately supported by a data center back-end cloud architecture, particularly when the IoT data generated is delay sensitive, high volume, trust sensitive and/or intermittently disconnected. As a result, we have seen firsthand the migration of cloud functionality to the edges of the network to be more proximate to the data creation [2]. That includes functionality for storing, managing and curating the data at the edges of the network as well as analytics performed on the data locally. Transcoding, aggregating, compressing, sub-sampling the data is done at the edge if there is a need to cache, store or move a less voluminous version of the data up-stream, possibly forming a reverse CDN of sorts as data wends its way towards an intermediate or back-end cloud where it might be archived or processed further. Some are calling this architectural shift *Fog computing*, because of this trend to move cloud capabilities more local or proximate to where data is getting created. One focus of our team has been to evaluate the level of trustworthiness of this new local cloud variant, of its resources and of the data coming from those resources. We posit that the success and acceptance of the IoT will be based on the level to which we can evaluate and guarantee trustworthiness, particularly since that data can affect the physical world. Thus, we must ensure that decision making that leads to actuation is based on trusted analytics, which in turn is predicated on trusted data, which in turn requires trusted devices.

Composability and Semantic interoperability. As a business unit engaged in building IoT products and solutions, we have observed a number of notable and relevant shifts in recent years. The more that IoT moves into every day spaces, the less likely that these edge clouds or local "data centers" look like traditional data center clouds. Resources are increasingly heterogeneous, at all levels (HW, FW, and SW). They are not necessarily manufactured, owned, designed or managed by a single person, vendor or entity. They represent a broad range of capabilities and services, as well as levels of security, privacy and/or trust, and in the extreme, they are less likely to be co-located physically within the confines of a single or secure physical space, as is typical for traditional clouds.

To assemble cohesive IoT solutions from these diverse components requires software-defined techniques that enable composability (1) of systems (with disparate HW, FW, SW components) that combine microservices and services, and (2) of the data associated with those systems. To accomplish this effectively requires a bridging of semantics between Things and creating sufficiently self-descriptive Things. For this task, we believe that Smart Object frameworks (UPnP, DOI/DOA, Haystack, IPSO v1) offer a way to bridge the cyber-physical divide. These frameworks essentially offer up an object as a Thing in the IoT. An object can be a physical tangible element, such as a device (e.g., a phone), a physical space (e.g., a room) or resource (e.g., storage), or it can be something in the cyber realm, like periodic or streaming sensor data, an executable piece of code (a service or microservice), or a cloud (a collection of objects). An object is basically anything to which one can attach an identifier.

In IoTG at Intel, we have embarked on a comparative study of smart object frameworks, establishing a series of metrics by which to evaluate them for the contexts in which we need to use them. Thus far, it is unclear to us to what extent smart object frameworks should handle data objects differently than other kinds of objects, and therefore in what ways semantic interoperability of data is the same or different than it is for other kinds of objects. However, we do know that IoT data creation has a much higher frequency than other types of IoT object creation. Thus, the registries that track data objects have different architectural considerations due to data object dynamics and require greater scalability to account for the ease with which data objects can evolve and spawn new lineages.

Meta-data: Enabling Trust in the IoT. It is abundantly clear that self-description is not only about objects having unique identifiers or having human- or machine-readable names that are handles for manipulation. More importantly, self-description is about being able to attach a rich set of attributes, characteristics, features, descriptors and meta-data (all terms used interchangeably) to IoT objects. Meta-data must capture the interfaces to an objects' exposed capabilities. Meta-data must also be able to describe both functional and non-functional requirements about the objects to which they are attached, e.g., what capabilities might an object require (memory, CPU, TPM presence, a certain SW patch level), as well as by what constraints (only a valid object until the expiration timer), SLAs (never use more than X amount of bandwidth) and/or policies (who under what conditions has access to this object) might an object need to abide.

In fact this is where trusted analytics comes back into play. We believe that meta-data is crucial to the ability to enable Trust in the IoT, and more specifically the ability to enable Trust in the objects that populate the IoT. While it seems straightforward that meta-data should capture an object's owner(s), its lineage (from what other objects it is descended, derived, composed, etc.), and access control policies (terms of engagement), ideally meta-data should enable us to negotiate and broker trust between objects - whether those components are building blocks of a larger IoT system, (microservices being composed into a larger service), or if the negotiation is more along the lines of a device deciding whether it should join a local cloud, or an analytics service deciding if a data stream is trustworthy enough for inclusion in a data computation.

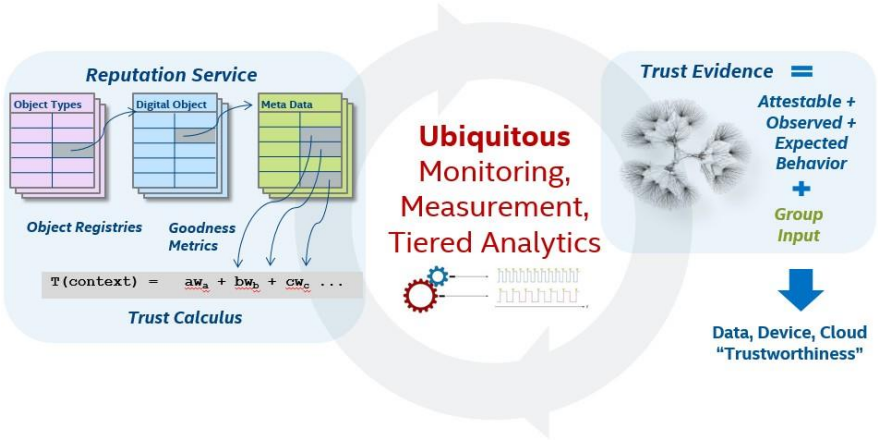
We are working on a set of micro-algorithms and a system design to allow Things to build trust, with or without a hardware anchor. Whereas current security management solutions ask "Should I trust/allow this object?", our work hopes to move solutions to ask, "***How much*** should I trust this object?" and to use meta-data to help answer this questions. In essence, we are enabling IoT systems to utilize the data present in the system (or added with minimal complexity) to autonomously determine which entities should be trusted and to what degree. This becomes essential when Things must operate independently, such as when they are mobile, do not have the luxury of a connection to a back-end data center cloud, are only intermittently connected to the network due to interference, or are in an energy-saving sleep mode. Even if objects are not required to operate independently, objects that wish to interact may come from different trust domains, which once again may require translation or mediation of trust assumptions.

We use the three-step approach pictured in the figure below to evaluate trustworthiness in the IoT: (1) Things monitor and gather local measurements about other Things that have been encountered, (2) Things generate an immediate local trust value by using a trust calculus (a mathematical formula for combining gathered measurements), (3) Things collaborate to construct and share group trustworthiness values (if feasible or allowed) [7][8].

The local trust values represent the direct interactions between Things, analogous to human peers interacting and evaluating each other. These values are based on any set of measurable features, characteristics, attributes, or meta-data deemed *contextually* relevant. We start with computing-oriented features (e.g., CPU class, energy usage, TPM version, percent time in sleep mode, reliability of connectivity, etc.), matching those used in cloud scenarios to assess the suitability of devices for tasks. From there, we extend the set of consulted attributes to take into account policy (e.g., organizational affiliation, patch levels, etc.) and meta-attributes (e.g., proximity to a particular person or Thing, requiring a percentage of heterogeneous sensors to agree as part of a group computation, presence of specific cryptographic keys, etc.). The group trust represents the aggregation of multiple opinions about a peer, similar to asking friends for their opinion of other peers. This larger scope allows entities to evaluate the trustworthiness of Things and to avoid potentially malicious or sub-standard entities without having to interact with them directly.
Once trust is available in an end-to-end IoT solution, it can be used in multiple ways including: the autonomous formation of local/edge clouds from collections of devices working together to provide an IoT service (e.g., a software-defined gateway); flexible access control policies even when failures occur (e.g., network outages) or when physical roots of trust are not available; device ranking for use in trust-based routing, allowing for greater control in choosing beneficial network paths. By bringing trust to the IoT, we are enabling flexible methods for system composition, new security paradigms, and previously untapped analytic capabilities.

Gaps and Challenges. While the standards organizations and consortia have made a series of proposals, some of which have begun to join forces (e.g., OCF), it will take time to sort out how to unify like-minded proposals. Even after some harmonization occurs, there will never be one single solution. Meanwhile, companies building IoT

products need to commit now to existing standards in whatever state they are in or roll their own solutions. Thus in both the short-term and the long-term, it will be necessary to dig into our interoperability toolkit (abstractions, APIs, shims) to mask the differences between models in as agnostic a manner as possible. Currently, there is a need to identify: core aspects that are shared across all schemes, aspects supported by only a few approaches, architectural pieces to a broader eco-system (e.g., naming, registries, tools to manipulate objects/things, bridges that translate between models, etc.). Additional questions and challenges that impact the architecture include: where to find or grow "bridges" (translators, ambassadors, mediators/proxies) between smart object frameworks, what policy/requirements/constraint languages are best suited to interoperability, can we embed security and privacy with the objects themselves (e.g., attribute-based encryption techniques that embed access control policy with data [4][5]), how to build an interoperable solution that works even in the absence of IoT infrastructure, can we leverage trust anchors already available to us while at the same time grow trust, trust anchors/brokers, and trust bridges between security systems, and how does naming at the application layer interact with the routing and caching of data at the networking layer (e.g., information centric networking (ICN))?



Proof of Concept. We have created an end-to-end Proof of Concept (PoC) that endeavors to vet many of these ideas. The PoC is a vehicle to show the utility of trusted local clouds and their interplay with non-local clouds. It demonstrates "Remote Monitoring and Control of IoT Smart Spaces using Video", which generates interactive real-time, high-volume, trust-sensitive data, all of which necessitate moving the computation and storage more proximate to where the data is generated; it showcases a "smart gauges" use case, where legacy non-smart devices are supplemented with IP-enabled cameras to collect status information both for anomaly detection (triggered by analytics algorithms running locally at the edge) and/or for on-demand monitoring of local (not remote) cloud data by remote users; it experiments with naming, caching and security/privacy in ICN, which routes the analytics executables [1] to where the camera data resides at the edges of the network; it aims to expose meta-data of the objects in the system and to perform continuous real-time monitoring to update meta-data status information that is consumed by trust calculus algorithms to derive trustworthiness and to form the basis of an IoT reputation service.

Relevant Publications.

[1] Andrew Brown, Sebastian Schoenberg, Eve Schooler, "NDN and the Internet of Things: Analytics Everywhere", poster, *2nd Annual Named-Data Networking Community Meeting*, NDNComm'15, LA, CA (Sept 2015).
[2] David E. Cohen and Eve M. Schooler, "Data Inversion and SDN Peering: Harbingers of Edge Cloud Migration", *IEEE ComSoc MMTC E-letter, Special issue on Big Data in 5G Networks,* Vol.9, No.6 (Nov 2014).
[3] Moreno Ambrosin, Christoph Busold, Mauro Conti, Ahmad-Reza Sadeghi, Matthias Schunter, "Updaticator: Updating Billions of Devices by an Efficient, Scalable and Secure Software Update Distribution Over Untrusted Cache-enabled Networks", *ESORICS'14* (Sept 2014).
[4] Xinlei Wang, Jianqing Zhang, Eve M. Schooler, "Performance Evaluation of Attribute-based Encryption: Toward Privacy in the IoT", *IEEE International Conference on Communications,* ICC'14, Sydney, Australia (Jun 2014).
[5] Mihaela Ion, Jianqing Zhang, Eve M. Schooler, "Toward Content-Centric Privacy in ICN: Attribute-based Encryption and Routing", *ACM SIGCOMM'13 and SIGCOMM ICN'13 workshop*, extended abstract, Hong Kong (Aug 2013).
[6] Jianqing Zhang, Qinghua Li, Eve M. Schooler, "iHEMS: An Information-Centric Approach to Secure Home Energy Management", *IEEE SmartGridComm'12*, Tainan City, Taiwan (Nov 2012).
[7] David Zage, Carl Livadas, Eve Schooler, "A Network-Aware Distributed Membership Protocol for Collaborative Defense", *IEEE CSE'09, SP4SPNA'09 Workshop,* pp1123-1130 (August 2009).
[8] Denver Dash, Branislav Kveton, John Mark Agosta, Eve Schooler, Jaideep Chandrashekar, Abraham Bachrach, Alex Newman, "When Gossip is Good: Distributed Probabilistic Inference for Detection of Slow Network Intrusions", *Proceedings Twenty-First National Conference on Artificial Intelligence,* AAAI'06, pp. 1115-1122 (July 2006).