The Identity of Things: Privacy and Security Concerns

By Simon Moffatt – ISSA Senior Member, UK Chapter

The "identity of things" is fast becoming a critical component of the modern web. While this can bring significant benefits in form of personalized content and customized environmental and manufacturing settings, it also brings several concerns regarding data privacy, security, and control.

Abstract

The "identity of things" is fast becoming a critical component of the modern web. Previously "dumb" devices are being upgraded with persistent network connectivity, enabling automatic data generation and facilitating interaction with other devices and people. While this can bring significant benefits in form of personalized content and customized environmental and manufacturing settings, it also brings several concerns regarding data privacy, security, and control.

The future of modern connectivity

Today, we are in the infancy of widespread mobile Internet connectivity, which we typically obtain through Wi-Fi hotspots and 3G/4G network coverage. When not connected, we are invisible to others, unable to get the information we need, and unable to interact with personal and professional networks. However, this concept of ad-hoc connection to the network is slowly altering. The Internet is no longer a separate object that we have to seek and connect with explicitly. Very soon, being "connected" will be so intrinsically tied to us that without it basic human interactions and decision making will become stunted. Switching an object on, purchasing it, enabling it, checking in to it will make that device become "smart" and tied to us. It will have network access and be able to communicate, send messages, register, interact, and contain specific contextual information. A simple example is the many running shoe companies that now provide GPS tracking and training support information for a new shoe. That information is specific to an individual, centrally correlated and controlled, and then shared socially to allow better route planning and training techniques to be created and exchanged. The flow of information requires an "always on" Internet connection, which creates many questions surrounding device management, security, and privacy.

The Internet of things

The Internet of things phenomenon will create a device-, people-, and services-based connected infrastructure of over 50 billions objects by 2020.[1] From a consumer perspective, home automation systems such as context-based lighting and heating or fridge restock systems help reduce energy consumption and billing, while also providing manufacturers and suppliers with powerful usage insights that can help improve products or provide better marketing opportunities. From a manufacturing or logistics standpoint, smart grid energy and electricity systems and improved SCADA (supervisory control and data acquisition) connectivity help automation and improve data flow. Future things-based infrastructures will include the marrying of insured devices such as cars and human bodies to the underwriting of insurance policies. Allowing insurance companies to interact with intelligent devices such as cars and human-wearable monitors provides them with a unique metadata opportunity that could allow insurance companies to create more accurate policies and reduce consumer insurance costs. By allowing cars to capture servicing, distance, and maintenance data, insurance companies can help to identify lower-risk drivers and car owners. In turn, consumers can have much more customized policies at a lower cost. This cost reduction, however, comes at a price: the loss of data privacy.

Why identity is important

The concept of a smart device is not new. Smart phones, for example, shipped over 1 billion units for the first time in 2013.[2] These devices are intrinsically linked to identity. The

1 David Lake, Ammar Rayes, and Monique Morrow, "The Internet of Things," *The Internet Protocol Journal*, Cisco Systems, Vol. 15, No. 3 – http://www.cisco.com/web/about/ac123/ac147/archived_issues/ipj_15-3/153_internet.html.

2 International Data Corporation smart phone shipment data January 2014 – https://www.idc.com/getdoc.jsp?containerId=prUS24645514.

identity of the phone owner—generally tied to a mobile network operator (MNO) via a long-term contract—is the first, but there are other identity components such as a mobile number, the IMEI number, the subscriber identity module (SIM) and associated cryptographic infrastructure, as well as the unique identity attributes associated with the many apps that are available.

While the identity component alone does not make a device smart, without it a device could be considered "dumb." Without a unique identifier or an association with a real physical identity, the object is inanimate, unable to communicate or provide context to the information that it is exposed to or able to generate.

But firstly, what is an identity? The English language definition describes an identity as "the characteristics determining who or what a person or thing is."[3] Those characteristics in a digital sense normally refer to attributes, with the values of those attributes being things such as name, email address, or an alphanumeric unique identifier (UID). The physical unique identifier does not always need to be globally unique. For example, take house addresses. A house number is obviously not unique across the world; it only needs to be unique with a certain context —the street (see figure 1).

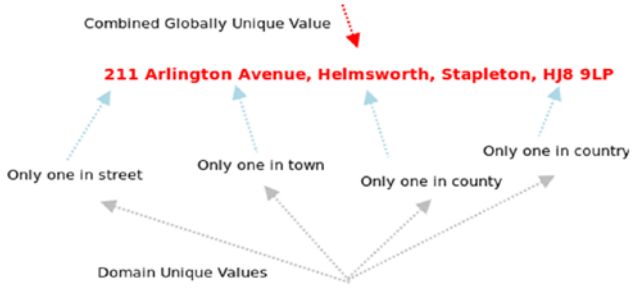

Figure 1 – Physical example of globally unique addressing

The same is true of devices and their association with a context and indeed physical person.

The underlying theme with identity, however, is that it is permanent. The UID should not be reused, even if the object referenced by the UID is not active. The concept of permanent identity is a contentious one. Yahoo announced in 2013 that is was going to reuse previously disabled email addresses.[4] While this gives the service provider the ability to sign up new users, this also poses significant privacy and security issues. A previously used email address, for example, could have been used to register for other services such as social networking sites or personal banking. If the host email address is disabled and reissued, what happens to the emails being sent out by those services? They potentially get into the hands of the wrong individuals.

Figure 2 – Example identity relationships

But what can be used as a unique reference? There are several examples in the different digital layers we use everyday. Figure 2 gives an example of how the joining of locally unique identifiers such as IP and MAC addresses to other identifiers such as email addresses can create chains of device and data identities.

The Extensible Resource Identifier (XRI) is an OASIS-driven initiative for the use of abstract identifiers that are domain, location, application, and transport independent.[5] The XRI format is compatible with the likes of uniform resource identifiers (URIs) that often make up web addresses. This coupled with the likes of more REST-based web technologies[6] paves the way for URIs that can focus on the potential relationships between people and the objects and devices associated with them, replicating the approach used for more common physical concepts such as postal addresses.

Identity of things data landscape

The Internet of things (IoT) has created vast amounts of big data, which is only likely to increase further: usage data, product data, location data, personal preference data, relationship data, health data, all of which could be generated or stored on devices with limited processing and storage capacity. This can bring several interesting questions around data privacy. The example of smart training shoes has several different side discussions. The ability for the training shoe to log GPS location data, time, speed, and other performance-related metrics is certainly a great advantage to the amateur athlete. Firstly, the ability to collect that data, but secondly, to be able to store and share that data socially is certainly a service-improvement use case. The training shoe manufacturer can use that approach as a competitive advantage, offering a gadget-like feature other training shoes do not have. However, what are the consequences of generating and storing such data?

If that data contains GPS information, a individual with malicious intent could use that information to know where the shoe owner is (or will go, based on historical data), and also where the owner is not.

Data on its own may not have significant value until it is combined with other data within a certain context. The Internet of things, while bringing significant service improvements to consumers and manufacturers, also brings concerns regarding the privacy and security of the underlying data. This requires a definition of those involved in the data life cycle,

3 Oxford English Dictionary - http://www.oxforddictionaries.com/definition/english/identity.

4 Yahoo's policy of email account reuse, CNET News article, July 2013 – http://www.cnet.com/uk/news/can-yahoo-recycle-your-username-and-protect-your-data/.

5 Extensible Resource Identifier - http://en.wikipedia.org/wiki/Extensible_resource_identifier.

6 Representational State Transfer protocol - http://en.wikipedia.org/wiki/Representational_state_transfer.

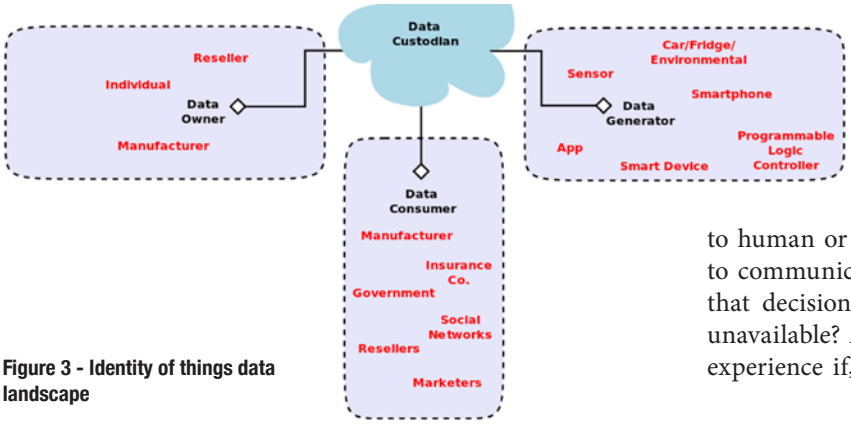

Figure 3 - Identity of things data landscape

alongside the context of their identity. The data owner will generally be an entity or individual. This would be required for accountability and access-control decision making. The data custodian is simply where data is residing—generally at rest, but data transfers need to be considered here too. That could include cloud storage solutions, social networks, or applications.

Data generators could well be the low-level digital devices such as cars, fridges, home automation systems, locks, smart phones, and so on, not to mention manufacturing controllers and sensors. This data landscape is made complete by those who will access the data—the data consumers. This group is potentially the largest, with data consumers from various different organizations and view points, all with different claims and levels of data requirements.

Data generator device security

The data landscape in figure 3 illustrates several basic security concerns. The devices themselves will generally be of low power with limited processing and storage capacity. This could potentially inhibit the ability for the device to perform any sort of encryption, either with regards to data stored locally or as it is transmitted over the either a local mesh-style network or indeed Wi-Fi. The devices themselves also need to perform some sort of registration or mapping, either to a local hub or a physical application or person. The opportunity for "rogue" devices is obviously significant. How can a device be "claimed" or at least prove its own assurance? A paradox also occurs with devices that are capable of performing basic operating system functions—the more complex a device becomes, the attack vectors for things like malware increase.

From a network perspective, technologies such as infrared and Bluetooth, while allowing for simple coupling and data transfer, have historically (pre version 2.1 of Bluetooth, for example) provided opportunities for encryption key compromise.[7]

Another opportunity for security compromise could potentially come from devices when they are not online. One of the foundations of the IoT landscape is that devices will always be on and always be connected. This allows for bi-directional communication with regards to authorization enforcement. Using the simple policy decision point (PDP)/policy enforcement point (PEP) architecture,[8] IoT devices may need to perform enforcement actions, allowing access or responding to human or machine interaction. If the local device needs to communicate to a hub or central policy service to make that decision, what happens when that communication is unavailable? A default fail-safe may impact on the end user experience if, for example, the device is a consumer-facing

7 Bluetooth encryption security concerns - http://en.wikipedia.org/wiki/Bluetooth - Security_concerns.

8 Policy Decision Point authorization architecture - http://en.wikipedia.org/wiki/Policy_Decision_Point.

one. But allowing a caching mechanism or some other local decision-making process could also introduce issues around policy synchronization or decision spoofing.

Devices will also need to capture and send data, generally to a central hub or to a local intermediary. Here device authentication to the hub is critical. Protocols such as Message Queuing Telemetry Transportation allow for the lightweight publishing and subscribing infrastructure that can form the basis for machine-to-machine data transfer. However, authentication of publishing devices is critical to avoid both rogue devices polluting data, as well as rogue clients reading data maliciously.

Data privacy

Assuming device registration, chaining, and capture has taken place, a complete end-to-end data-storage and access-control pattern needs to be implemented. Basic data encryption, from both a transit and storage perspective, is a given. This should, in theory, be no different than any other cloud-based storage solution. A more complex puzzle to solve, is how to implement the *who*, *where*, and *when* questions regarding data access. A fourth question, the *why*, also needs to be considered when focusing on data consumer registration and approval.

Data privacy brings several interesting topics. Firstly, the data itself requires an owner, someone to be accountable when it comes to making consent decisions. The data itself may also need labeling, in a similar way to the classified system used by many governments.[9] Will some data be public? Will some data be accessible by some parties but not others? Will some of the access-control decision making be dynamic and change at run time? How can those access-control decisions be managed and enforced, while simultaneously being understood by a non-technical approver? All of these data-management topics are well understood and observed within the enterprise world, but now they have to be applied to a more scalable, consumer-facing world.

Authentication, authorization, and context

The data landscape shown in figure 2 requires several security components to make it function. Authentication (confirming the truth of an identity) and authorization (confirming what that identity has access to) are the two main components. Both, however, require a context in order for a decision-making process to be fully optimized.

A basic example of authentication for an person-based identity is the username and password. This is the "something you know" concept, as opposed to "something you have" (one-time password generator) and "something you are" (biometric proofing). Authentication plays a significant part in the data landscape. Data owners, consumers, and generators all need to be identified and verified. The authentication approaches here will vary significantly. A physical device performing machine-to-machine (M2M) style communication is unlikely to leverage username and password, for example. Would a physical device need to register with the data custodian, or perhaps be claimed by the data owner?

Either way, the device needs a process that allows proof of its identity in order to validate the data that it can generate. Much M2M-style communication is often encrypted with

9 Data classification standards in the United States government -http://en.wikipedia.org/wiki/Classified_information_in_the_United_States.

other crypto architectures such as Public Key Infrastructure (PKI) or JSON Web Tokens (JWT) being used to perform device authentication.

Data consumers require both an authentication and authorization process. Verifying their identity is a prequel, allowing them access to certain aspects of the data repository. With regards to the training shoe example, figure 4 shows the main actors involved. The shoe manufacturer may just be interested in the shoe age and distance run statistics, while completely ignoring the location and GPS data of the runs completed, for example.

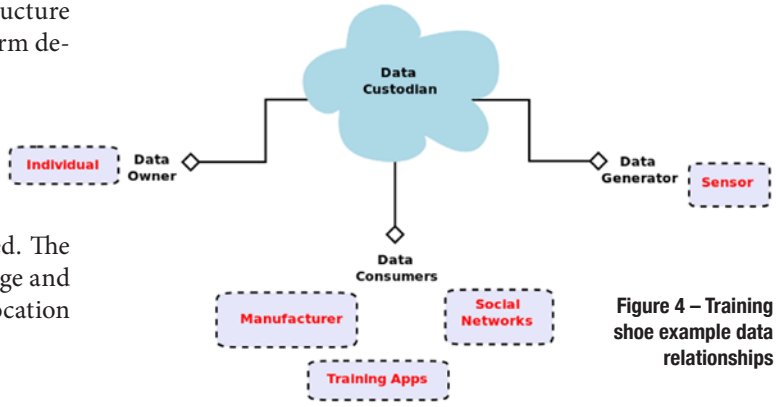

Figure 4 – Training shoe example data relationships

Mechanics and existing standards

The Internet and identity-of-things approach contain many different, complex, and more importantly continually evolving components. These components include mechanics for things like user and data-owner registration and authentication through to the smart–device communication and storage protocols—protocols that potentially need to be optimized to run with a tiny memory and processor footprint.

From a network and data storage and transportation approach, JSON (JavaScript object notation) and REST (representational state transfer) have become a common approach for web developers operating in a platform-agnostic environment, where mobile phones and websites combine.

As a result, JSON- and REST-based authentication and authorization options such as OAuth2 and OpenID Connect are seeing traction by providing a developer-friendly approach. User-centric-based approaches to authorization have also seen the likes of UMA (user-managed access) develop and may have a role to play.

OAuth2 is perhaps the most popular with regards to consumer-based authorization, having been a familiar component of social networking. OAuth2 is often used to allow third-party clients to access social network information such as a user's email address or profile attributes without the need to share password details. OAuth2 also gives the ability for the data owner to remove previously granted access through the revoking of an access token.

Figure 5 gives an example overview of the some of the existing and fast becoming popular protocols for the IoT landscape.

The Curmudgeon

Teaching Moments

ONE OF THE NICE THINGS about being a curmudgeon is that people *expect* you to grumble and complain. That's usually because something didn't work out as expected, such as retaining one's youth.

As information security professionals, we learn to think about things in ways that most people never consider. When we discuss those with other people, we come across as paranoid, distrustful, or malicious. "In order to defend, you must know how to attack."

Decades ago, when I loaded my first "secure" operating system that wanted an identity and password, I realized there was an issue. What happened if I were not immediately available and my spouse needed access to the information? Most particularly, financial information? Technology improved and things got more interesting. I learned not to log in as "admin" or "root," so I set up ordinary user accounts. Then I realized that the operating system did what it was supposed to and prevented me from seeing my spouse's files, and vice versa.

Like many other people, I ended up with logons for many sites and services, which meant a whole lot of passwords. And, like most other people, I wanted to "store" them somewhere securely. But, what happened if I were "unavailable," for an extended period? Or if that file suffered bit-rot?

I learned (the hard way) that "persistent storage" is not, and the greatest threat to it was my <Del> key. I started a system of backups and began using encrypted files.

Then there were the stories from the field: Someone's father died, having kept all the family business records protected on his computer. It was a "going concern," a speech idiom meaning a successful business. By the time they managed to hire someone to crack in, weeks had passed and the business went under. Something that took a lifetime to build, died in weeks, due to lack of "disaster recovery" planning.

Something we can do as professionals is pose situations like these to people outside the field and ask how they would handle the situation. It's a way to discuss the ethics of the profession *and* get them to think things through. Most of the time, they turn around and ask how *we* would handle it.

And that, ladies and gentlemen, is a "teaching moment."

The trick is not to present *all* the ideas and considerations, "what-ifs" and "think abouts" we've gathered from our way of thinking. Just tell them "enough" and let them think some more. That way, they'll listen more, next time.

The next time you join them for a chat, you can extend things by saying: "I was thinking about that conversation about...and I wondered: What if...." and you present the next plot complication.

I've always wondered: Is it paranoia, or is life really out to get me?

Me, I bet on the latter. Harrumph!

Your local, grumpy, tie-wearing, unimpressed, and suspicious Curmudgeon

From a device perspective there are several avenues. Firstly, things like transport protocols, the encryption and privacy of those protocols, as well as approaches for things like data storage, device registration and interfacing—using lightweight concepts such as QR (Quick Response) codes—all need to be considered from an identity and interaction perspective.

The main concern is how so many disparate protocols, devices, and components can be easily integrated to form scalable and lightweight infrastructures to manage the IoT landscape. Modular and loosely coupled services would allow for the rapid provisioning of new consumers and publishers of IoT-related projects. The modular aspect could be based on standard web technologies such as REST, which would allow for powerful, highly customized web- and mobile-based user interfaces to be kept separate from the underlying registration and access-control infrastructure. Standards-based integration into existing identity-provider-based infrastructures such as social networking sites would allow for the rapid sign-up and viral marketing capabilities so often required of new platforms.

The data custodian aspect is probably the most well-serviced today, with numerous cloud-based environments for platform, infrastructure, and software as-a-service style subscriptions that can scale rapidly to allow for the potential of multi-million user deployments.

The area developing the most rapidly would seem to be at the device level. More advanced micro devices requiring crypto

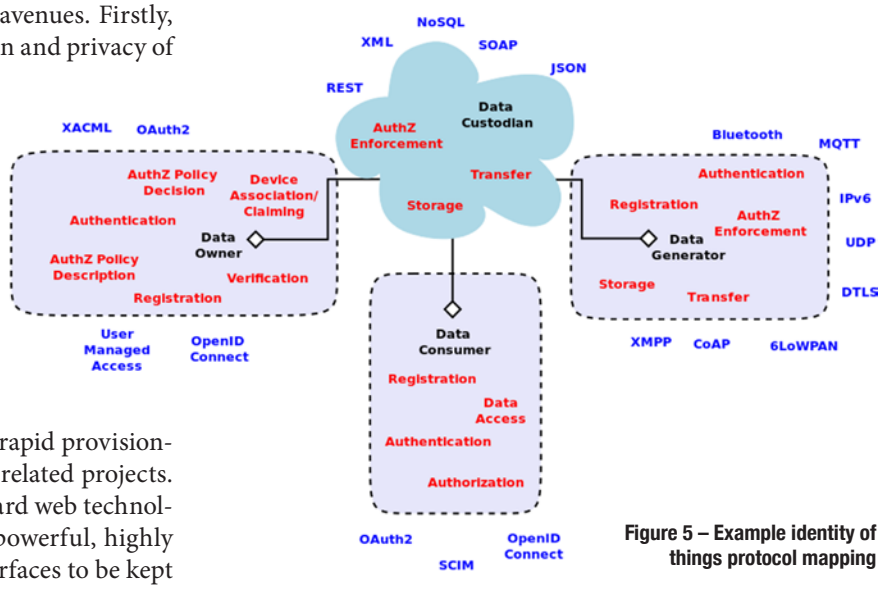

Figure 5 – Example identity of things protocol mapping

processing, JWT integration, and the ability to act both online for data publishing and perhaps offline for authentication and authorization are seemingly being released weekly.

One of the longer-term aspects of the IoT landscape is the analysis of the generated data, user interactions, and access-control decisions. Business intelligence for the IoT world would help drive consumer marketing decisions, manufacturing product research, and many more. The use of an identity context in such analytic processes is key to help define things like behavior profiling, peer comparison, and fraud.

Conclusion

The Internet of things brings together devices, people, and services into loosely coupled but highly optimized chains of data. To protect the privacy of that data, while allowing access to the various data consumers, requires a complex balance of authentication, authorization, and contextual awareness. Underpinning those requirements is a need for unique identities, at both the local and global level, along with strong registration, claims, and implicit approval processes that allow people-to-machine, machine-to-machine, and people-to-service relationships.

While there is a vast opportunity to create new personalized services and content, that opportunity comes at a cost—data privacy. Device registration and ownership need to be carefully mapped to a flexible identity and access-control mesh that is easy to implement, scalable, and easy for a non-technical end user (the ultimate data owner) to understand and embrace.

About the Author

Simon Moffatt has over 13 years information security experience with a specialization in identity and access management. He is currently Principal Engineer at Open Source ISV ForgeRock. He may ber reached at simon@infosecprofessional.com.