Clearing off the Cloud over the Internet of Things

Carsten Bormann, Stefanie Gerdes, Olaf Bergmann
Universität Bremen TZI
Bremen, Germany
Email: {cabo|gerdes|bergmann}@tzi.org

***Abstract*—As was foreshadowed by product introductions in 2013, the Consumer Electronics Show 2014 has seen the introduction of a large number of "Internet of Things" (IoT) innovations. Almost all of these have in common that they are meant to operate via Cloud-based services.**

In the light of the recent attention to threats by state-level tenacious attackers with significant infrastructure (STASI), in particular to their practice of pervasive monitoring, we discuss the implications of a cloud-centric IoT landscape, and attempt to outline a set of principles as a program to improve the long-term outlook.

I. Introduction

Many recent product introductions in the area of "Internet of Things" (IoT) have in common that they are meant to operate via Cloud-based services.

While the product announcements are not always very clear about the technical details, it seems that even a BlueTooth only toothbrush sends information into the cloud via an app installed on the users' smartphones ("Full use of the KOLIBREE Services requires [...] Internet access" [1]).

(In this paper, the on-demand scaling aspect of the Cloud is of little relevance; we therefore lump together services built upon classically deployed fixed servers with truly cloud-based ones.)

Recently, threats by state-level tenacious attackers with significant infrastructure (abbreviated STASI here) have received heightened attention. In particular, revelations about their practice of pervasive monitoring change the perceived threat landscape considerably.

We discuss the implications of this threat landscape on a cloud-centric IoT approach, and attempt to outline a program that could guide us to improve the long-term outlook.

II. The Lure of the Cloud

For IoT innovators, the Cloud is highly attractive as it lends itself to a scalable business model.

For users, Cloud-based offerings also pose multiple attractions:

- limited initial investment
- sustained operation (outsourcing of maintenance)
- pretty web-based user interfaces
- the promise of easy and seamless integration

(In this paper, we will not discuss where these attractions tend to fall down except for the security considerations.)

Where sensor data need to be stored for extended periods to allow some processing, the sensor device itself is often not the appropriate place for this storage. It is attractive to store the data in the cloud instead of requiring the user to invest in a local storage device, which in turn would need management, a backup solution etc. Any processing associated with the data can receive much shorter upgrade cycles in a data center than in a device that may need firmware upgrades that can go wrong in various ways and are then hard to recover.

The web-based user interfaces are also of interest as they provide a comfortable, standards-based way to configure the IoT devices, including their security properties. Service owners also have a much easier development and upgrade path for the web-based parts of their service as compared to software that would need to be upgraded on the IoT devices themselves. Users are already familiar with the security procedures of web-based service offerings, making them more immediately usable.

One of the technical reasons for embracing a Cloud-based approach for consumer offerings is that IPv4 usually does not provide end-to-end connectivity outside the confines of the home. To be practical before IPv6 becomes widespread, IoT offerings for home environments need to integrate with the widely deployed IPv4 WiFi home networks that only support Internet clients, requiring servers in the greater Internet. The same considerations not only apply to the sensors and actors themselves, but also to historical data stored in

a database, all of which also need to be accessible to mobile devices when outside the confines of the home.

A related reason to use Cloud-based communications are notification services such as Apple Push Notification Services (APNS), Android Cloud to Device Messaging (C2DM), Google Cloud Messaging (GCM). Here, messaging to a device is funneled through a service run by the operating system manufacturer in order to reduce the number of connections needed for multiple applications expecting notifications, and to allow automatic launching of an application when a notification for it arrives.

III. Threats

With all data always in the transit to and from the cloud, STASI-type operations achieve full visibility into the data unless these are properly protected during communication.

There is also a trend to outright outsource STASI operations to commercial entities [2], making not just the communication towards the cloud a concern, but also the data at rest at the cloud services as well as the (potentially security relevant) operations on them.

This comes in addition to the trend to consider personal data stored at a service provider to be non-confidential for the purposes even of legal discovery (third-party doctrine).

When private data are exported into the cloud, this means they are now accessible to anyone who can make a claim that they are relevant to any kind of civil dispute. Even if the access is limited by court order to the lawyers of the accessing party, this may provide little protection in practice [3].

Even where the purchaser of a product might assume its data is never leaving their single home, cloud-based services may provide an attacker with an opportunity to redirect communications (as was recently demonstrated for a WiFi-based baby monitor offering [4]).

Even where data can be confined to a single home, e.g. between a light switch and a light, the setup of the authenticated authorization may benefit from using Web technology. It requires considerable attention to detail to enable cloud-based security setup while still protecting operational communications. Any keying material used for the cloud-based setup must not be useful for gaining access to the operational communication in a passive attack (this is usually achieved using forward secrecy mechanisms).

As with web-based services, realistic Internet-of-Things devices must provide a way to *reclaim* their use when operational credentials have been lost. This reclaim mechanism needs to be protected from being exploitable by pervasive monitoring.

Generally, we distinguish passive monitoring from active, detectable monitoring. To control pervasive monitoring, it is worthwhile to counteract passive monitoring even if the attacker has an easy way to switch to detectable monitoring.

IV. Cloudless or unclouded?

The objective of this paper is not to argue against the use of the cloud (reducing the solution space to "cloudless" approaches), but to argue for adding to the agenda the development of mechanisms that allow a user to keep the dangers of the cloud in check ("unclouded" approaches).

As an example, reclaim, which is an operation that in many cases needs services from the cloud, must be reliably detectable for the concerned parties. More generally, security's trinity of prevention, detection, and recovery applies.

Giving the users control in recovery is probably a more general principle of the Internet of Things. Users of the well-known Nest thermostat report that they woke up to a misfiring 4.0 firmware upgrade that left their furnaces disabled during the coldest winter for a couple of decades [5]. Apart from ripping off the thermostats from the wall, there was no recovery action that the users could take.

V. The ten laws of clear sky

In complex subjects such as the present one, it is often useful to summarize the issues into easily communicated rules. In the area of identity management, this has been successfully done in the seven "laws of identity" [6]. Identity itself is likely to take more of a back seat in the Internet of Things. Still, maybe not too surprisingly, some of the headlines from the "laws of identity" can be immediately reused, after some generalizations of their definition. Others may need to be reformulated and new laws may need to be added.

The initial set of "laws" we propose for discussion follows. The first three are generalized from Cameron's principles, reusing his headlines:

1) **User Control and Consent**
   Technical systems must only reveal information with the user's consent.
   Clearly, Cameron's argument that a system "must earn the user's trust above all" does apply here as well.
   Note that "consent" here must be "informed consent", a non-trivial addition.

Assuming that this can be achieved, the decisions of the user then need to be mapped into authorization information that the devices act upon. Keeping this information radically simple is an important prerequisite to minimizing surprises with respect to the consequences a user decision might have.

2) **Minimal Disclosure for a Constrained Use**
The solution which discloses the least amount of information and best limits its use is the most stable long term solution.
Again, Cameron's argument that "we should build systems that employ [...] information on the basis that a breach is always possible" is more pertinent than ever.
With respect to privacy considerations, information that has been disclosed to the cloud is no longer under control of the information owner. This argues for applying a strict form of the "data minimization" principle to data made available to the cloud. (Or, preferably, not storing or communicating data at all.) Where the storage (and disaster recovery) function is the important contribution of the cloud, the data could be stored in encrypted form based on keying material that cannot be unilaterally reconstructed by the cloud provider.
This may be inconvenient from a "big data" point of view, and may require improved ways of handling data that minimize disclosure.

3) **Justifiable Parties**
Systems must be designed so the disclosure of information is limited to parties having a necessary and justifiable place in a given relationship.
Cameron argues that the "system must make its user aware of the party or parties with whom she is interacting while sharing information." This is a bit harder to generalize, as it is less clear what the boundaries of the "relationship" are: Is my electricity provider part of the "relationship" with respect to specific details of my energy usage such as per-minute consumption (even if it needs access to summary data to perform billing)? Probably, some application of privacy principles such as *contextual integrity* [7] is required to decide this in each case.
Given the third-party doctrine, a cloud provider almost never is a justifiable party just from the fact that it provides a cloud service.

It is somewhat harder to translate Cameron's principle 4 ("directed identity") into the domain of the Internet of Things. We try to capture the essence by focusing on the subjects of discovery and authorization.

4) **Directed Discovery and Authorization**
A system must support both "omni-directional" information sets for general use and "unidirectional" information sets for use within specific private authorization relationships, thus facilitating discovery while preventing unnecessary release of correlation handles.
Note that the need for having omni-directional information sets may be limited to a specific phase in the lifetime of a device.
An important concept here is that of ownership. An owner of a device must be able to fully control authorization of operations on that device, including restricting its discovery and its use to specific other devices. An owner also needs to be able to transfer ownership. (Finally, there may be a need to support involuntary release of ownership, e.g. after a foreclosure.)

There are three more principles (5, 6, 7) in Cameron's seven laws that almost sound like a matter of course today (or at least should be). They are listed (and generalized) here:

5) **Pluralism of Operators and Technologies**
A universal system must channel and enable the inter-working of multiple technologies run by multiple providers.
One important enabler of pluralism that is worth mentioning here is the provision of a data export function, both as an API for permissionless innovation and to foster competition between different operators and technologies. Also, relationships between devices and between them and cloud services need to use indirection mechanisms that can be modified under control of the user (such as the CoRE resource directory [8]) instead of more hardwired mechanisms out of their control (such as the DNS).

6) **Human Integration**
The system must define the human user to be a component of the distributed system integrated through unambiguous human-machine communication mechanisms offering protection against attacks.
Constrained devices often have no or a very limited user interface. It is therefore important to enable other devices under control of the users to stand in for them for the purpose of human-machine com-

munication. This communication must be made resilient to attacks.

7) **Consistent Experience Across Contexts**
The system must guarantee its users a simple, consistent experience while enabling separation of contexts through multiple operators and technologies.
This can be aided by standardizing enough aspects of authorization setup that it then be delegated to user interfaces of different sources.

Finally, we derive three principles of unclouded operation:

8) **Direct Communication**
Communication must be direct between the entities that actually need to communicate, with no diversion to additional parties simply for implementation convenience.
This is a direct corollary of principle 3 "Justifiable Parties", but also of principle 5 "Pluralism of Operators".
It is important to note that this objective is nearly unattainable with NAT-ridden IPv4. With IPv6, there can still be "security" mechanisms in the way, such as Simple Security [9]. The current specification for Simple Security, RFC 6092, only recognizes IPsec (AH and ESP headers) and HIP as protocols that need to be enabled for inbound communication.
As long as some actual competition between different approaches can be maintained on the marketplace, this principle might be relaxed to one where the user is given the ability to opt in for direct communication and possibly forego some conveniences in exchange.

9) **Non-Transitivity**
Communication in the cloud, e.g. for the setup of authenticated authorization, must not be exploitable for deriving keying material for operational communication in a passive attack.
The security mechanisms must be carefully designed such that an active attack on a system in the cloud cannot be used to enable a passive attack of data in use between private devices. Trivially, keying material needs to be kept separate between those applications; mechanisms used for forward secrecy can be employed for this. Also, cloud-based setup may be supplemented by NFC or BlueTooth Low Energy to make attacking it more difficult. More generally, the role of a cloud service in a setup process must not give that cloud service unrestricted access to the operational system.
Given the long expected service lifetimes of some Internet-of-Things installations, it is already prudent to consider hardening operational communications for a post-quantum era.

10) **Think globally, act locally**
Devices must be able to obtain essential processing services locally (under control of the user), without always requiring cloud services for processing the private data of the user.
Where some processing of privacy-relevant data is required, e.g. for the application of rule sets or prediction of user behavior for optimization decisions, there should be a system configuration that does not require disclosure of those data into the cloud. Note that this configuration may be more expensive (require additional purchases) or less convenient (require additional operation and maintenance) according to some measures. This principle is not meant to rule out the use of data from the cloud or even the careful disclosure of processed data to the cloud to enable globally optimized behavior.
As long as some actual competition between different approaches can be maintained on the marketplace, this principle might be satisfied by giving the user the choice between different providers of processing, not just ones under control of the user, but also from a selection of competing providers in different jurisdictions.

References

[1] Terms of use | Kolibree. [Online]. Available: http://www.kolibree.com/terms-of-use/

[2] R. A. Clarke, M. J. Morell, G. R. Stone, C. R. Sunstein, and P. Swire. Liberty and security in a changing world — report and recommendations of the president's review group on intelligence and communications technologies. [Online]. Available: http://www.whitehouse.gov/sites/default/files/docs/2013-12-12_rg_final_report.pdf

[3] J. Ribeiro. (2013, November) Judge considers sanctions against Samsung, lawyers in dispute with Apple. [Online]. Available: http://www.pcworld.com/article/2062320/judge-considers-sanctions-against-samsung-lawyers-in-dispute-with-apple.html

[4] N. Dhanjani. (2013, October) The Belkin WeMo baby monitor, the WeMo switch, and the Wi-Fi NetCam – reconsidering the perimeter security argument. [Online]. Available: http://www.dhanjani.com/docs/Reconsidering%20the%20Perimeter%20Security%20Argument.pdf

[5] TechCrunch. (2014, January) Update: Nest says shut-off heat is sometimes its fault, also pushes thermostat 4.0.1 firmware to fix

4.0 problems. [Online]. Available: http://techcrunch.com/2014/01/06/nest-4-0-firmware-battery-problems/

[6] K. Cameron. (2005, November) The laws of identity. [Online]. Available: http://www.identityblog.com/stories/2005/05/13/TheLawsOfIdentity.pdf

[7] H. Nissenbaum, "Privacy as contextual integrity," *Washington Law Review*, vol. 79, no. 1, pp. 119–158, 2004.

[8] Z. Shelby, C. Bormann, and S. Krco, "CoRE Resource Directory," Internet Engineering Task Force, Internet-Draft draft-ietf-core-resource-directory-01, 12 2011, work in progress. [Online]. Available: http://www.ietf.org/internet-drafts/draft-ietf-core-resource-directory-01.txt

[9] J. Woodyatt (Ed.), "Recommended Simple Security Capabilities in Customer Premises Equipment (CPE) for Providing Residential IPv6 Internet Service," Internet Requests for Comments, RFC Editor, RFC 6092, January 2011. [Online]. Available: http://www.rfc-editor.org/rfc/rfc6092.txt