Published in *2017 IEEE Symposium on VLSI Technology*, pp. T8-T13, IEEE, 2017.

Privacy and Security: Key Requirements for Sustainable IoT Growth

Fari Assaderaghi, Gowri Chindalore, Brima Ibrahim, Hans de Jong, Marc Joye, Sami Nassar, Wolfgang Steinbauer, Mathias Wagner, Thomas Wille

NXP Semiconductors, 411 E. Plumeria, San Jose, CA 95134, USA

fari.assaderaghi@nxp.com

Abstract

As IoT moves beyond a catchphrase and starts to provide meaningful solutions in multiple fields, three of its critical pillars are now well understood:

- Transducers are needed as means of interacting with the environment and machines, and in converting stimuli to data and vice versa. These sensors and actuators form the basis of contextual awareness.
- Given that many end-node IoT devices are size and power constrained, local low-power computing is essential. The need for power-efficient end-node and edge computing becomes more apparent when latency, network bandwidth, and real time analytics are considered.
- Low power communication links to transmit the data between IoT devices and local aggregators or cloud resources form the third pillar.

Missing in this picture, and not fully appreciated yet, is the fourth pillar of IoT: privacy and security (P&S). If IoT is all about data, how P&S is treated will determine IoT's fate: a second phase of rapid proliferation or ultimate demise and collapse. Recent breaches in P&S are starting to change the industry's view on this issue. Even IoT end nodes that are low cost and have limited functionality pose significant risk to the entire system when their security is breached. This is due to the networking nature of the IoT that exposes a massive attack surface, making these devices ideal attack points for causing disruptions and stealing sensitive data. PC-era Internet security has been an expensive afterthought that has cost industry and consumers billions of dollars. Therefore, we should approach IoT differently, making P&S a key requirement at the design phase itself, and address all life-cycle aspects from initial deployment to in-field updates, to end-of-life decommissioning. This is a system level challenge that requires complete end-end HW/SW solutions, developed in partnership with the entire ecosystem.

Keywords: Internet-of-Things (IoT), low power devices, end nodes, transducers, embedded processors, connectivity, privacy and security.

Low Power Sensors and Actuators Spur the Growth of IoT

A dizzying array of sensors and actuators are either now available or being rapidly developed, unleashing the IoT revolution. Several themes emerge from this burgeoning field:

1. Nearly any physical-world phenomena that can be measured or manipulated, is now readily sensed or "actuated", i.e., converted to electrical signals and eventually digital data, and vice versa. Light, sound, linear and rotational motion, temperature, humidity, pressure, electric and magnetic fields, different gases and chemical compositions, and human vital signs all fall within this realm.
2. These transductions create many new capabilities: contextual awareness and environmental sensing, human-machine interactions and intuitive interfaces (touch, gesture, and voice-based), augmented and virtual reality, biometric and health sensing (fingerprint, voice, iris, PPG, blood pressure, oxygen and sugar level), and autonomous machines (cars and drones).
3. The transductions are increasingly carried out by extremely miniaturized systems, often MEMS. Accelerometers, gyroscopes, barometric pressure sensors, microphones, force sensors, magnetometers, optical transducers, chemiresistors, thermal sensors, and haptics now have millimeter-level and smaller dimensions. This miniaturization has relied on many tools and processes of conventional semiconductor manufacturing, and has benefited from its economy of scale. Further, many of these transducers are in very high volume consumer electronics such as mobile phones, leading to their rapid cost reduction.
4. At the same time, these devices are consuming much lower power, becoming smarter, and more accurate, while generating immense amount of data. In fact, it is estimated that by 2020, 40% of all data will be generated by these transducers.

The combination of very small size, low cost, low power, and intelligence enables tremendous amount of use cases of transducers in IoT, while creating a significant dilemma in analyzing, communicating and protecting their data.

Computational Challenges for the Next Generation of IoT

With projected 30+ billion connected devices by 2020, the amount of generated data will stress the bandwidth of wireless networks, and consume an enormous amount of electricity if transported and managed in the cloud. In 2016 annual global IP traffic surpassed one Zetta Byte (10^{21}) threshold, and in the US alone data centers consumed 70 billion kilowatt hours of energy [1]. An effective approach to alleviate this growing challenge is to create a distributed computing network, pushing processing to the IoT end nodes and gateways. This distributed network built with computationally advanced devices handles most of the data processing locally and only sends the relevant or meta information to the cloud servers for more complex decision making and storage purposes. This approach substantially reduces the bandwidth and power burden for wireless networks, while improving response time and security of the critical data.

Adding to this requirement is the increasing demand for end nodes to have smartphone-like user-friendly interfaces with rich 2D/3D graphics and integrated convenience features like touch sensing, voice-based assistance, biometrics for password-less secure access, and facial recognition. This growing trend requires that embedded processors have substantial performance and memory bandwidth to provide seamless, low-latency, and secure implementation of these capabilities. Many end nodes are battery operated and cannot require constant human-user intervention for seamless operation. In fact, many of these nodes are in remote inaccessible areas and need to run without human intervention for months or even years.

We are also seeing an increasing implementation of gateways that act as command and control centers for local networks of end nodes. The use of these local aggregators alleviates computational burden and connectivity requirements of the end nodes, since they need to communicate with the gateway that is located within a short range. The local gateways can be built with more powerful embedded processors, including local-area and wide-area connectivity. Considering the end-to-end IoT network, the challenge therefore is to develop *a scalable set of power-efficient, high performance, secure embedded processors.*

Earlier, the needs of end nodes were sufficiently addressed by MCUs with 8/16-bit cores and limited narrowband connectivity. However, over the past few years the use of embedded processors with 32-bit cores has dramatically increased, outpacing the combined sales of 8-bit and 16-bit processors. Moving to 32-bit processors (in 90 or 40 nm) provides significant improvement in performance to 8/16-bit architectures in use today. The performance improvement opens new possibilities with more deterministic real-time operating systems (RTOS), ability to run complex connectivity software stacks, enhanced security protocols, advanced user-interfaces, and still have processing power left over for the defining features of customer applications.

While there has been significant progress in the process technologies and core architectures to improve the dynamic power efficiency, care must be taken in system-level design to reduce the static power consumption. This is especially critical since many of the IoT end nodes spend much of their time in monitoring mode or even in deep-sleep mode where uncurtailed static leakage can significantly reduce battery lifetime. Through circuit and architecture innovations, great strides have been made in the last few years in improving both static and dynamic power efficiency without compromising performance. For example, MCUs today can deliver dynamic performance of 100-200MHz @ < 50 µA/MHz and consume < 5 µA in deep-sleep mode with fast wake-up time of < 5 µs.

Process technology innovations are also contributing to a much more power efficient local and embedded computing. For example, 28/22 nm FD-SOI technology lends itself well to a balanced power and performance combination. Through management of voltages around the IC, the application processors (APs) built on FD-SOI can be made to deliver peak performance only when needed and can be put into very low leakage mode when not in use. Further, by appropriate use of either forward biasing or reverse biasing, the FD-SOI based APs can deliver GHz-level of performance, but at a power envelope that is comparable to power efficient MCUs. Relying on FD-SOI and other power-efficient process technologies, more features such as dedicated cores and/or hardware blocks for audio processing (DSP cores), advanced 3D graphics (GPU), video processing (VPU), image processing (ISP), and security enablement can be included in the APs. For security, dedicated hardware IP blocks can be integrated for enhancing safe management of the keys, fast encryption/decryption capability, tamper detection & prevention, and many other features that ensure confidentiality, integrity and privacy of the user data.

As Moore's law has guided advances in processor/memory industry for several decades, the demands of IoT will drive the computational capability and power efficiency of embedded processors for many years to come. For example, embedded machine learning

algorithms are now being implemented into the edge nodes. The trend of moving more of computational tasks from the cloud into the embedded and local processors will continue unabated.

Wireless Connectivity for IoT

Although wireline communication is also used in IoT, wireless connectivity is its true growth accelerator. The use of wireless as the medium of communication has allowed significant flexibility and ease of use. It however is faced with its own set of challenges, including interoperability, power constraints, connection robustness, coverage range, and security of communication over an open and shared medium. One convenient way to analyze the IoT wireless connections, is to categorize them based on the distances devices within the network are required to communicate. Therefore, we can classify the IoT networks into ultra-short range, short range, and low power wide area range.

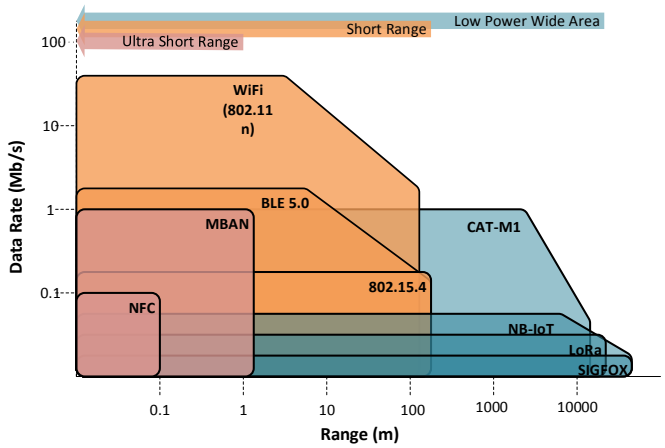

Fig. 1 IoT wireless options: Rate vs range

A. *Ultra-Short Range*

Ultra-short range wireless protocols provide wireless connections from a few centimeters to ~1m. Examples are Near Field Communication (NFC) and Mobile Body Area Networking (MBAN). NFC that was initially designed for secure access and financial transactions can now play multiple roles in IoT networks. One common use case is secure onboarding, where a new device is authenticated and provisioned to join the IoT network. The device credentials are stored locally or in the cloud and are authenticated via a secure NFC transaction. NFC also allows devices that are not battery powered to participate in the IoT network by use of passive tags. These tags can be interrogated to provide information (e.g., smart posters and information kiosks).

With the progress of semiconductor technology, medical sensors continue to shrink and require less power. Adding wireless communication to these devices has enabled more convenient monitoring of patients in hospitals or at homes. In 2012 FCC enabled the use of MBAN in the 2360 MHz – 2400 MHz band. The MBAN network consists of low power sensors attached on or underneath the skin, communicating to a coordinator device that may also be worn by the patient. The coordinator also includes a more conventional link to the internet, such as WiFi or cellular.

B. *Short Range*

Local area wireless protocols are designed for communication links up to a maximum of ~100m. These short-range networks are mainly deployed for in-building applications with a significant number of battery powered sensor or actuator nodes. Short range networks also employ the unlicensed frequency bands, with 2.4GHz UNII being the most popular. Some of the most common short range protocols used in IoT networks are 802.15.4 based (ZigBee(Pro), Thread), WiFi and Bluetooth Low Energy (BLE). Key requirements of short range IoT networks are support of long battery life, range extension, network reliability, IP-based routing, and a standardized application layer.

The batteries of these sensors and actuators need to last from 2 to 10 years, based on their activity level and ease of access. The IoT network protocol must be designed to allow for such long battery life requirements. ZigBee, Thread and BLE are all developed with battery constrained devices in mind where applications are developed with coin-cell or AAA battery as power source. WiFi requires higher power consumption and is not well suited for small battery applications. The OFDMA feature proposed for 802.11ax allowing narrower bandwidth transmission will also help WiFi achieve longer battery life.

To increase the coverage area of link-budget-constrained networks, mesh networking can be implemented. The difference between a point-to-point or star network and a mesh network is that messages are not communicated over a single link, but can be routed over more than one link before arriving at the destination. Therefore, devices can communicate with each other much farther than their link budget allows. Both ZigBee and Thread have mesh functionality built into their specifications. The present Bluetooth 5.0 specification does not include mesh functionality, but the Bluetooth SIG Technology Roadmap for 2016 puts forward the adoption of mesh functionality to make BLE a strong mesh platform.

Both ZigBee and BLE specifications include an application layer that provides defined attributes for sensors and actuators to interact with each other over an IoT network. ZigBee enables this through the ZigBee Cluster Library (ZCL), and BLE through profile specifications based on the Generic Attribute Profile (GATT). Since

WiFi and Thread only provide IoT functionality to the network layer, third party application layers are required. Having common application layer over a common network layer such as IP-based networks allows devices to communicate with each other regardless of the specific transport layer in use.

C. Low Power Wide Area (LPWA)

As the IoT market continues to grow, the use cases for connectivity across urban, sub-urban and rural area have continued to increase. The development of smart cities will also drive use cases including remote metering, environmental monitoring, and real-time traffic monitoring. LPWAs needed for such deployments have several common characteristics:

- Link budget challenged – km range and/or deep building penetration
- Low throughput and infrequent communication
- Required ease of deployment – Battery powered nodes with 10+ year life expectancy
- Low node cost (< $5)
- Support large number of nodes per base station
- Secure communication (authentication and data encryption)

Even though short range specifications include mesh networking for extending range, they will not be practical for range extensions to cover wide areas and meet the above requirements. Dedicated LPWA networks such as SigFox and LoRa are designed from the ground up to meet the requirements of the LPWA IoT use cases. They provide typical network implementation with base stations forming a star network and are deployed in unlicensed bands. They are designed with the requirement to achieve large link budgets for wide area coverage and extremely low power operation for long battery life.

With the increase in demand for LPWA IoT networks, mobile network operators have been using variants of LTE and GSM to address these use cases. The 3GPP Release 13 standardized LTE CAT M1 and NB-IoT specifically to meet the IoT use cases and provide a fully open standard that can be adopted by network equipment providers and endpoint device manufacturers in the cellular licensed bands. The CAT-M1 and NB-IoT standards are designed to reuse the cellular infrastructure and coexist with legacy deployments of LTE, 3G and 2G. This has the significant advantage of reducing overall cost of deployment.

Privacy and Security (P&S) of IoT: More Complex than any P&S Area Before

Privacy and Security pose significant challenges for many systems. These challenges are made more complex in IoT due to its additional constraints such as very low cost, low power, and limited connection bandwidth.

Let us consider a few challenges of secure systems and their extension to IoT.

A. Data Integrity / Confidentiality

When we interconnect devices, the first security challenge is how to prevent data exchanged from being altered or eavesdropped while in transit. Here, data encryption is a key part of the solution. Encryption algorithms can be implemented in SW or accelerated in dedicated HW. The HW accelerator can drastically reduce the power consumption and latency, as well as provide additional defenses.

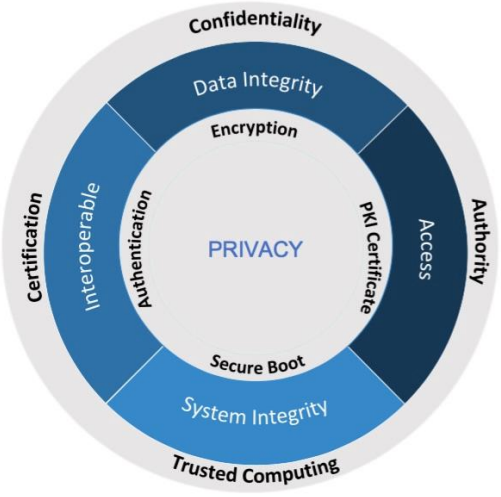

Fig. 2 Key considerations of privacy and security

B. Certification / Interoperability

To ensure interoperability, all devices connected to a certain service should have a minimum level of compliance and in some cases, require certification. Using simple MAC address or an OTP ID can be used to establish the communication, but it is very easy to replicate and clone these features. Cloning will allow counterfeiters to introduce uncertified products that interfere with the quality and reliability of the service. Here, device authentication is the solution, as it creates trusted communication between two devices. It provides the ability to verify that each device is what it claims to be. Many ecosystem providers use authentication to ensure that different vendors are compliant before granting them authentication keys. Authentication keys must be well protected as counterfeiters will spend substantial effort to extract the keys and create clones. A very effective

countermeasure is to hide these keys in a secure-element IC.

C. Integrity / Trust

Most IoT devices allow for remote software update, making them targets of scalable attacks. It is essential to be able to recognize if a device has lost its integrity because an unauthorized change of software has altered its use. Trusted computing environments with a hardware root-of-trust, combined with secure boot and life cycle management are effective countermeasures.

D. Strong Authentication for Access

In many cases the IoT device must access a protected resource, and needs to be assigned credentials for this access. Today, unfortunately most access authentications rely on username and password for cloud services. This method is obviously insufficient when challenged by the latest hacking tools. HW-based Strong Authentication is an effective defense. NXP alongside Google, Microsoft and 250 other companies created the FIDO Alliance protocol that is using HW-based Strong Authentication for logical access to cloud services.

E. Privacy: Dissociating Identity from Data

Personal data has enabled the development of a multitude of new services, applications, and devices: recommender systems, personal assistants, social networking services, monitoring systems, personalized applications, and more. At the same time, data collection poses real privacy risks. This is even more true in the case of IoT as the type of collected data is more personal and the gathered information more detailed. Privacy breaches are numerous. In February 2017, a SmartTV company was convicted for collecting data on viewing habits from 11 million Smart TVs without user's consent. In January 2017 IP-camera security flaws were detected from a company failing to protect its IoT devices from widely known and reasonably foreseeable risks of privacy data lost. In December 2016, the Norwegian Consumer Council carried out an investigation about how the talking doll 'my friend Cayla' operates and interacts with children. Later, in February 2017, Germany banned both the sale and ownership of the interactive doll, alleging that it contains a concealed surveillance device that violates federal privacy regulations [2].

The solution here is privacy by design. A key component enabling privacy by design is *data encryption*. Encryption technologies are now widely available and integrated in a variety of systems. The past decade has seen several advances that go beyond the traditional public-key encryption paradigm where the owner of the private decryption key matching a given public encryption key can decrypt data encrypted under that encryption key. Two recent technologies, fully homomorphic encryption and functional encryption, have made significant progress in this field. They essentially allow the recipient of the data to *operate on encrypted data*, without ever seeing the plain data itself. Other useful technologies include differential privacy, garbled circuits, and zero-knowledge proofs of knowledge [3-4].

F. Design Principles

IoT faces many other security and privacy issues and requirements that make the situation appear intractable. However, adhering to few proven principles can create effective countermeasures:

(a) Consider security at end-to-end system level

Right tradeoffs can be made only when the complete end-to-end system is considered, including services of external parties. One needs to evaluate the impact attacking one node of the system can have on other nodes. For example, using the same key in all nodes (non-diversified keys) creates significant vulnerability – if one node is compromised, the entire systems will be exposed.

(b) Take the long-term view

An IoT system that is put in place may stay deployed for many years. Attacks that we can only imagine today will be tried over the life time of the system, and then more.

(c) Manage end of life

However complex it may be to deploy a system, cleaning up after end of life of its components can be more complex, especially if this was not considered at the introduction – the system may lack necessary resources for creating countermeasures.

(d) Rely on separation

Logical attacks (i.e. that can be carried out remotely, in contrast to physical attacks) are generally the most dangerous, as they scale well. An adversary may attack millions of devices and then use all of them in the attack against other systems. The key defenses here are separation and isolation. Particularly, a very effective approach is for the IoT device to have a *physically* separated part that contains the most sensitive information (e.g., encryption keys). This can be a separate processor with a narrow bandwidth and well controlled connection to rest of the system and to the outside world. It can be implemented as a separate chip or be a specific block on a larger SoC. Such a separation may also make it possible to defend against

physical attacks – attacks for which the adversary in general must have physical access to the device. In such a case, only a small part of SoC needs to be protected against physical attacks where that part may also help protect the rest of the SoC.

(e) *Assume system security failure at some point*

Whatever measures are taken to make the system secure, we need to assume that it will be compromised at some point in time. Therefore, one needs to limit the damage the compromised device can cause on environment. Having a small well-protected enclave that holds the most sensitive data and code will help blocking the compromised system of being used in larger attacks. As important, is implementing an upgrade mechanism that allows recovery from an attack. Each system should be designed such that after an attack, the device is still reachable, software can be reloaded, and the proper owner can regain control.

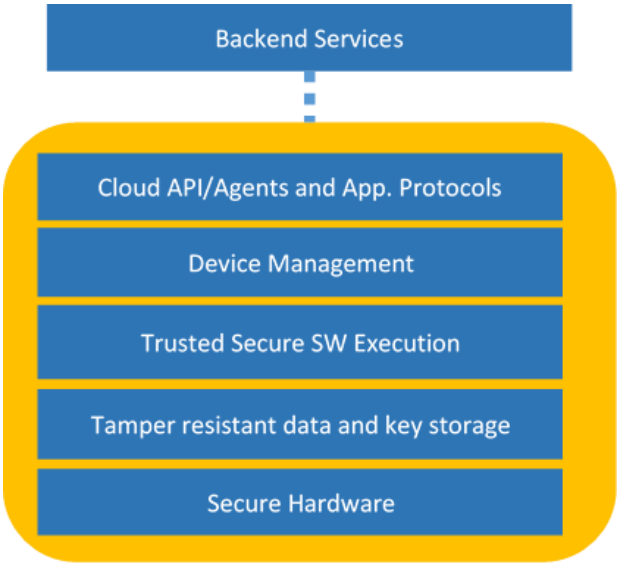

Fig. 3 Key enablers of security in an end-to-end IoT system

To illustrate that it is possible to build high level of security and privacy while creating new business opportunities, we consider the evolution of credit card payment enabled on mobile phones with Secure Element (SE). SE forms the root-of-trust for payment in the mobile device and enables a secure end-to-end operation:

- The SE stores the credit card data and crypto-keys. SE has been personalized in secure environment by entitlement of the issuing bank.
- User/owner authenticates herself versus the SE by biometric fingerprint. The fingerprint circuit is connected securely to the SE.
- SE is entitled to create payment tokens signed by the SE's private key. SE must be activated by the owner. By using payment tokens, the original credit card data never leaves the SE.
- Use of tokens with its cryptographic signature prevents manipulation by merchant's card readers, i.e. the "man-in-the-middle" attack.
- Tokens are fully compatible to the standard payment network and are authorized finally by the issuing bank.

This tokenized mobile payment systems provides strong end-to-end security by:

- Relying on the SE as the root of trust in the mobile device.
- Providing strong owner biometric authentication coupled securely to the SE as root of trust.
- Securing the connection to the acquirer/merchant bank by advanced cryptographic mechanisms.

Conclusions

Tremendous advances in computing, communication and sensing have ushered in the era of IoT. IoT holds the promise of transforming all aspects of human society, from critical infrastructure, to industry and automation, to agriculture, to health, to transportation, and entertainment. This revolutionary development will be derailed if security and privacy are not addressed systematically and as primary factors in the ecosystem design. Our previous experience with PC-era security is a cautionary tale of the pitfalls of treating P&S as an afterthought and an ecosystem patch. Fortunately, this is not an intractable problem, and if certain principles are adhered to, all IoT verticals can be deployed without compromise in P&S, as demonstrated by the mobile phone payment system enabled with Secure Element.

References

[1] Lawrence Berkeley National Laboratory Report
[2] E. Schulz-Kamm, "Protecting the "I" in the IoT: GDPR and future challenges," 2017. [Online]. Available: https://blog.nxp.com/security/protecting-the-i-in-the-iot-gdpr-and-future-challenges.
[3] C. Gentry, "Fully homomorphic encryption using ideal lattices," in 41st Annual ACM Symposium on Theory of Computing (STOC), 2009
[4] D. Boneh, A. Sahai and B. Waters, "Function encryption: A new vision for public-key cryptography," Communications of the ACM, vol. 55, no. 11, pp. 56-64, 2012.