Maintaining Security in a Heterogeneous and Changing World

The peril of operating in a permanently degraded state

Jon Geater
Jitsuin
Cambridge, UK
jon.geater@jitsuin.com

Cesare Garlati
prpl Foundation
Santa Clara, California, USA
cesare@prplfoundation.org

***Abstract*—Safety and security concerns are holding back the Industrial Internet of Things (IIoT). Much of this comes down to two very different inconvenient truths: first, that Smart Cities and Connected Infrastructure are by nature composed of highly diverse sets of devices, yet device security standards are highly variable; and second, that those devices are operating in a permanently degraded state.**

Firmware and device data need ongoing maintenance to overcome vulnerabilities and defend against newly-discovered threats, and yet this lack of interoperability makes such patching very difficult to realize. This paper argues for standards and interoperability at a critical layer of the stack – secure boot, firmware, trusted execution environment and identity protection – in order to enable proper security management of the IIoT ecosystem.

Keywords—IoT; blockchain; cryptography; firmware; m2m; digital identity, RISC-V, Trusted Execution Environment, TEE

I. Introduction

The Internet of Things, the technology that promised us utopian smart cities and connected lives, is failing to deliver. Instead of a coherent Internet of Things we have in its place an Internet of Silos, where narrow use cases may work very well, but devices, systems, and economies cannot interoperate. Differences in device standards, a lack of consistency in device security, and a 'land-grab', 'winner-takes-all' mentality on cloud management services means that while vertical walled-garden digital consumer services are making strides into the connected future, the physical world is left frustratingly behind.

Secure By Design initiatives such as NCSC's Secure by Default [1] are starting to address this problem but at the current time they overwhelmingly focus on production and boot-time security only. This is a welcome and necessary step but it is not sufficient for safe and effective operation of devices over their planned 5, 10, or even 30-year lifetimes. The moment a device is shipped its security quality will begin to drop as hacks and vulnerabilities are found. Digital Identities lose trust as organizations come and go. And today's Roots of Trust will fall into unanswerable question as the cryptography and enclave technologies upon which they rely are overtaken by the inexorable march of security research [2] [3].

So now arises a quandary: devices must be equipped with remote update capability in order to patch vulnerabilities, but that requires a connection to the internet which in turn introduces new risks. The recent fashion has been to add secure channel technology (typically TLS) during device manufacture for delivery of firmware in the field, but quite quickly these mechanisms themselves were compromised by hackers [4] [5] and demonstrate that a whole-system, whole-life approach is desperately needed. Although examples date back some dozen years now, it remains clear in 2019 that "IoT Security is (Still) a Gigantic Mess" [6].

In other words, starting trusted is not enough. Devices need constant maintenance to recover from daily security degradation. And simply cobbling together today's fragmented silo approaches to embedded security technologies is impractical: we need open standards and norms for managing large systems of devices if we are to connect, automate, and smarten up our cities and infrastructure.

II. Not all security is created equal

To ask the question "is it secure" always invites another: "secure from what?". Some security is better than others, this much is undeniable. But what does 'better' really mean?

In a fundamentally online world (which IIoT clearly is), interactions happen over remote APIs, generally using cryptography to secure the links and data packets. And in this case, there is no obvious difference at the receiving end between data signed by a secure device, and that signed by an insecure one. The math works in all cases, and it's the combination of hardware and software resilience, enclave technologies, engineering processes and certification that

actually tell you whether that device and its digital identity are actually trustworthy.

Unfortunately, even when 'good' device security is established, 'trustworthiness' of a device is still not a black-and-white issue: exactly the same device in the same state may be viewed as either good or bad depending on the needs and situation of the observer. So, what is important for system security management is not to provide a simple 'yes or no' attestation of goodness, but rather to provide salient information about the state and status of devices from which a relying party can make their own judgements based on a view of the whole system of devices and their own unique risk landscape.

In "Key safety challenges for the IIoT" [7], the example of a remote-enabled door is used to illustrate this principle: for the *security* of the assets behind the door it is best to design a system that fails *closed*, such that it remains secure even in the case of unforeseen and unhandled extreme circumstances; but for the *safety* of people working inside it is preferable to fail *open*, lest they be trapped inside during an emergency. No matter how good the security primitives in the embedded component that drives the door control, it is *impossible* to know at component design time which of the safety and security needs will be prioritized by the system owners once installed.

And once the system architecture is settled there are still many operational and practical considerations that can hinder adoption or compromise IoT projects.

Various open source and open standards initiatives are trying to address this problem (for example, part of the Open Trust Protocol [8] tries to establish a common language for provisioning characteristics, and FIDO's Metadata Service [9] provides a means of checking the hardware characteristics of FIDO authenticators). But still these operate at a very low component level which cannot take account of whole system effects: each is a valuable but small piece of what has to be a larger collaborative solution.

III. The Scale of the Problem

Analyst predictions for the scale of IoT deployments have been booming, up from 26 billion by 2020 (Gartner, 2013) to 40 billion in the same timeframe (ABI, 2018), with chip company intel forecasting a total of 50 billion. Financial figures are equally impressive, with Gartner and IDC both predicting spend and benefits in the high hundreds of billions of dollars.

With 'Connected-X' and 'Smart-Y' being presently fashionable, then, we have a peculiarly special and acute problem with current Industrial IoT programs: capital money is available for showcasing and installing these new technologies – variously for grants, wow-factor, or indeed genuine progress reasons – but security and maintenance spend is not keeping pace, meaning that in a few years' time we might expect to find tomorrow's connected infrastructure crumbling just like yesterday's bridges. We have already seen the first waves of attack and how hard it is to patch and recover: Stuxnet [10], Mirai [11], NotPetya [12], Charlie Miller's Jeep exploit [13] and countless credential weaknesses in devices found trivially through Shodan.

And absent strong intervention things are only set to get worse. Artificial Intelligence and Machine Learning technologies are seeing a huge influx of investment as people see their potential for automating mundane and costly tasks. But AI needs to be fed with data – a lot of data – and so naturally IoT device output becomes an attractive option. Unfortunately, hackers have already noticed this trend and realized the tremendous strategic potential of poisoning the training set data: thereby not needing to hack the AIU at all, but simply teaching it bad habits [14][15]. All the more, then, it is essential to ensure that IoT devices remain trustworthy, updated, and in a known state for their entire lifecycle, since every piece of data they emit has the potential to be an attack vector for a dependent AI.

IV. Heterogeneity is Fundamental and Unavoidable

When we consider any modern city one this is clear: "Rome wasn't built in a day" [16]. Unlike consumer digital systems, which can be created from scratch and renewed from green shoots every couple of years, the physical world is constructed piece-by-piece over hundreds of years. Technologies of all ages, from all kinds of manufacturers, are mixed together and forced to work. Chosen primarily for their immediate extant utility, and not their embedded 'smartness', it is inconceivable that any industrial connected system of any size will feature embedded components exclusively from one architecture or product family.

On the architecture side, then, with MIPS now joining RISC-V in offering open source embedded hardware designs [17], the coming years are bound to see strong and vibrant growth in solutions that diverge from the current ARM- and intel-based architectures as open and lower cost innovation takes hold. In the current world, patches to issues such as the various SGX vulnerabilities [3] have included not updates to software (above the ISA) but updates to the microcode of the processor itself.

In the open source processor world, the implementation of the processor may be as open as the software stack itself and the diversity of implementations is set to explode as designers take advantage of the extensibility of the RISC-V ISA to customize processors to their applications. In a softcore FPGA implementation, this creates an additional component of the FPGA bitstream which needs to be managed and patched, in

the hardcore world a robust patch will require the combined efforts of multiple software and core design teams and be much more tailored to the specific implementation. The result is better, but the process is more complex.

Systems today typically only have one software stack per communications port, most often implemented inside the RichOS. What this means is that any field update of a system relies on the most complex and least trusted component of that system – the RichOS such as Linux or FreeRTOS. With new vulnerabilities being discovered in Linux at a rate of almost one per day during 2017-2018[18], a new model needs to be considered such as pulling the communications stack out of the RichOS [19] or implementing a resilient OTA mechanism that does not rely on the RichOS at all.

That all means more Roots of Trust (RoT) to understand and manage, more software stacks to patch, more diversity of implementations and more security models to build into systems.

Imagine separate devices with secure digital identities built variously with ARM TrustZone TEE, RISC-V with hardware-enforced, software defined enclaves such as MultiZone Security [20], a MIPS device with CHERI [21], and an x86 device with separate SE core. Each of these may be individually the best that they can be but once composed into a system how is it possible to compare the security of the TEE TSM administrative infrastructure with the formal verification of the RISC-V system? Or the security of SE key storage vs application partitioning? It's not that it's not possible to build good systems today. It's just too complicated for most industrial players – who do not, by and large, have a large weight of cryptography Ph.Ds.' – to plan and implement them.

Let's even imagine these hurdles are overcome. In this case, with long supply chains, many device technologies and many stakeholders, exactly who is responsible for maintaining the security of IIoT systems? How do we create a system where the incentive to keep devices patched and current is repaid by a commensurate share of the profits? Is it possible to establish a virtuous circle where value derived from IoT data can fund the maintenance of devices, keeping data truthful and high value?

V. A Collaborative Solution

Security in IoT systems is a Team Sport [22] where multiple stakeholders have to work together with multiple types and levels of security in devices to keep a whole system fully safe and maintained.

Historically in the security industry we have always suffered from the classic economic problem of 'externalities', which is to say (in this case) that the costs of implementing security measures are not borne by those who actually feel at risk; and by that same token that the pain of a security failure is not felt by the person taking the risk. When business cases are reviewed it has always been very difficult for device and software makers to put extra time and money into integrating features that only protect users several steps down the value chain: users who in turn are generally unwilling to pay for such fine gestures.

But lately, off the back of innovative technological progress in cloud, networks, and cryptocurrencies, efficient systems for shared value economies have arisen that at last make it feasible to share both the burden and spoils of good security maintenance throughout the value chain. Such sharing is important because technical (rather than legal) responsibility for fixing issues and bringing degraded devices back to health will lie with different actors at different times: if a firmware issue, then the device maker needs to issue a patch; if configuration then the owner of the system needs to remedy – having first validated with the operators that a system change is safe; and for industry-wide changes it may fall on a government entity to instigate and enforce change rather than any one actor. Working together as peer stakeholders is the only way.

To enable these collaborative systems to reach their full potential, though, we first need the common language, embedded security standards, and RoT-agnostic security protocols in place to talk to. Otherwise how can the device data – the 'new oil' of the World economy – be trusted?

VI. Conclusions

Smart Cities and Industrial IoT more generally are a necessarily multi-stakeholder enterprise and keeping them safe and maintained is a Team Sport. Only through collaborative device lifecycle assurance can the IIoT remain sustainable and secure for the long term, and thereby fulfil its promise to deliver Connected Living. This requires open standards for embedded security interfaces, clear layering and interoperability of the embedded security stack, and a harmonized language for describing and comparing security characteristics in remote systems. With these in place, the possibilities for trade of clean data with known provenance will realize the true value of the Internet of Things.

When man or machine can act on data without fear we can finally build the cleaner, greener, more efficient and more productive connected world that the IIoT has always promised.

References

[1] NCSC, 'Secure by default", https://www.ncsc.gov.uk/topics/secure-default , 2017

[2] NIST CSRC, "Post-Quantum cryptography", https://csrc.nist.gov/projects/post-quantum-cryptography , 2018

[3] J. Van Bulck et al. “Foreshadow: Extracting the keys to the Intel SGX Kingdom with transient out-of-order execution”, SEC’18 Proceedings of the 27th USENIX Conference on Security, 2018

[4] J. Ng, M. Lu and P. Zhang, “Superfish”, https://www.cs.bu.edu/~goldbe/teaching/HW55815/presos/superfish.pdf , 2015

[5] Langner “Original Cyber-Forensic Analysis of Stuxnet”, https://langner.com/stuxnet/, 2010

[6] R. Marvin “IoT Security is (Still) a Gigantic Mess”, https://www.pcmag.com/news/365994/iot-security-is-still-a-gigantic-mess, 2019

[7] Industrial Internet Consortium, “Key safety challenges for the IIoT”, https://www.iiconsortium.org/pdf/Key_Safety_Challenges_for_the_IIoT.pdf , 2017

[8] IETF, “The Open Trust Protocol”, https://tools.ietf.org/html/draft-pei-opentrustprotocol-00, 2016

[9] FIDO Alliance, “Metadata service”, https://fidoalliance.org/metadata/, 2017

[10] Nicolas Falliere, Liam O Murchu, and Eric Chien “W32.Stuxnet Dossier” February, 2011. [Online] https://www.symantec.com/content/en/us/enterprise/media/security_response/whitepapers/w32_stuxnet_dossier.pdf

[11] Manos Antonakakis et al “Understanding the Mirai Botnet” 26th USENIX Security Symposium, August, 2017.

[12] US-CERT “Malware Initial Findings Report (MIFR) – 10130295” NCCIC, Whitepaper, 2018-06-30.

[13] Andy Greenberg “HACKERS REMOTELY KILL A JEEP ON THE HIGHWAY—WITH ME IN IT” Wired Magazine, July 21, 2015.

[14] MIT “Norman: World’s first psychopath AI”, http://norman-ai.mit.edu, 2018

[15] M. Jagielski et al “Manipulating machine learning: Poisoning attacks and countermeasures for regression learning”, https://arxiv.org/pdf/1804.00308.pdf , 2018

[16] https://www.phrases.org.uk/meanings/rome-wasnt-built-in-a-day.html

[17] MIPS “The MIPS Open Initiative Will Accelerate Innovation at the Edge”, https://www.mips.com/mipsopen/, 2018

[18] https://www.cvedetails.com/product/47/Linux-Linux-Kernel.html?vendor_id=33

[19] B. Car and C. Garlati “Trusted Execution Environments: A System Perspective”, Proceedings of Embedded World 2019, (unpublished)

[20] https://hex-five.com/press/hex-five-announces-general-availability-of-multizone-security-the-first-trusted-execution-environment-for-risc-v/

[21] R. Watson et al “Capability Hardware Enhanced RISC Instructions: CHERI Instruction-Set Architecture (Version 6)”, 2017

[22] W. Ashford “Cybersecurity is a Team Sport”, https://www.computerweekly.com/news/252436352/Cyber-security-is-a-team-sport-says-NCSC , 2017