Securing the Internet of Things: A Proposed Framework

Introduction

By 2020, it is estimated that the number of connected devices is expected to grow exponentially to 50 billion. The main driver for this growth is not human population; rather, the fact that devices we use every day (e.g., refrigerators, cars, fans, lights) and operational technologies such as those found on the factory floor are becoming connected entities across the globe. This world of interconnected things - where the humans are interacting with the machines and machines are talking with other machines (M2M) — is here and it is here to stay.

The Internet of Things (IoT) can be defined as "a pervasive and ubiquitous network which enables monitoring and control of the physical environment by collecting, processing, and analyzing the data generated by sensors or smart objects."

The concepts and technologies that have led to the IoT, or the interconnectivity of real-world objects, have existed for some time. Many people have referred to Machine-to-Machine (M2M) communications and IoT interchangeably and consider them one and the same. In reality, M2M can be viewed as a subset of the IoT. The IoT is a more encompassing phenomenon, which includes Machine-to-Human communication (M2H), Radio Frequency Identification (RFID), Location-Based Services (LBS), Lab-on-a-Chip (LOC) sensors, Augmented Reality (AR), robotics and vehicle telematics. Many of these technologies are the result of developments in military and industrial supply chain applications; their common feature is to combine embedded sensory objects with communication intelligence, running data over a mix of wired and wireless networks. In a broader context, the architecture encompasses the Internet of Things plus business engineering insights captured from the information transmitted by these so-called "smart objects." The focus and scope of this paper is solely on the security aspects of the Internet of Things.

The capability of *embedded and distributed intelligence* in the network is a core architectural component of the IoT for three main reasons:

- **Data Collection:** Centralized data collection and smart object management do not provide the scalability required by the Internet. For example, managing several million sensors and actuators in a Smart Grid network cannot efficiently be done using a centralized approach.
- **Network Resource Preservation:** Because network bandwidth may be scarce and collecting environmental data from a central point in the network unavoidably leads to using a large amount of the network capacity.
- **Closed Loop Functioning:** For some use cases, the IoT requires reduced reaction times. For instance, sending an alarm via multiple hops from a sensor to a centralized system (which runs analytics) before sending an order to an actuator would entail unacceptable delays.

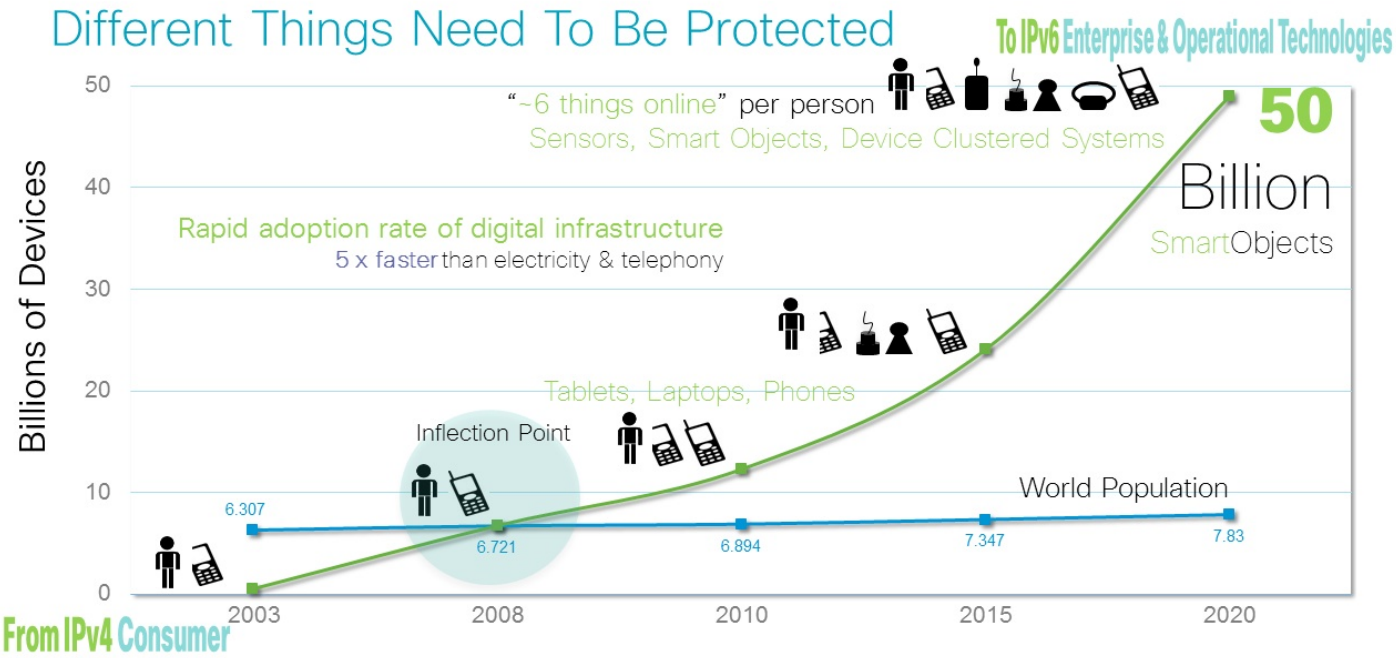
This distributed intelligence capability is known as Fog Computing, an architecture specifically designed to process data and events from IoT devices closer to the source as opposed to a central data center (also known as "Cloud"). In summary, Fog Computing is an expansion of the cloud paradigm. It is similar to cloud computing but closer to the ground. The Fog Computing architecture extends the cloud out into the physical world of things.

Service Management Systems (SMS) (also known as Management Systems, Network Management Systems, or backend systems) are the brains within an IoT architecture. SMS interacts with intelligent databases that contain intellectual capital information, contract information, policy information, manufacturing and historical data. SMS also support image recognition technologies to identify objects, people, buildings, places, logos, and anything else that have value to consumers and enterprises. Smartphones and tablets equipped with cameras have pushed this technology from mainly industrial applications to broad consumer and enterprise applications. The types of data these systems support and maintain continue to evolve.

Growth of the IoT

An important inflection point occurred in 2008, when the number of things connected to the Internet surpassed the human population. The adoption rate of the IoT is trending to be at least five times faster than the adoption of electricity and telephony, shown in Figure 1. This equates to about six things for every person on earth.[8] A interesting trend contributing to the growth of the IoT is the shift from the consumer-based IPv4 Internet of tablets and laptops, that is, Information Technology (IT), to an Operational Technology (OT)-based IPv6 Internet of Machine-to-Machine interactions. This includes sensors, smart objects and clustered systems (for example, Smart Grid).

Figure 1. IoT Growth



Source: Cisco IBSG projections, UN Economic & Social Affairs http://www.un.org/esa/population/publications/longrange2/WorldPop2300final.pdf

From a technology perspective, there are three main drivers that contribute to the growth of the IoT:

- **Ubiquitous Computing:** With intelligence in things at the edge, e.g., lightweight operating systems such as TinyOS running on very small computing platforms
- **Ubiquitous use of IP:** with convergence of protocols to run over IP rather than proprietary transports. Also greater adoption and support for IPv6 in carrier networks
- **Ubiquitous Connectivity:** Including cellular, radio and fixed. This includes low power, personal area wireless mesh networks particularly suited to sensors

Essentially, the enhancements and progress in these technologies have allowed the development of IoT devices such as sensors that have compute, storage and network capabilities built into extremely small form factors with low energy requirements.

Researchers and early adopters have been further encouraged by advancements in wireless technologies, including radio and satellite; miniaturization of devices and industrialization; and increasing bandwidth, computing, and storage power. All of this provides an opportunity to reduce management and operational costs by converting these systems from the legacy platforms, such as Modbus or other serial communication protocols, to an IP-enabled infrastructure.

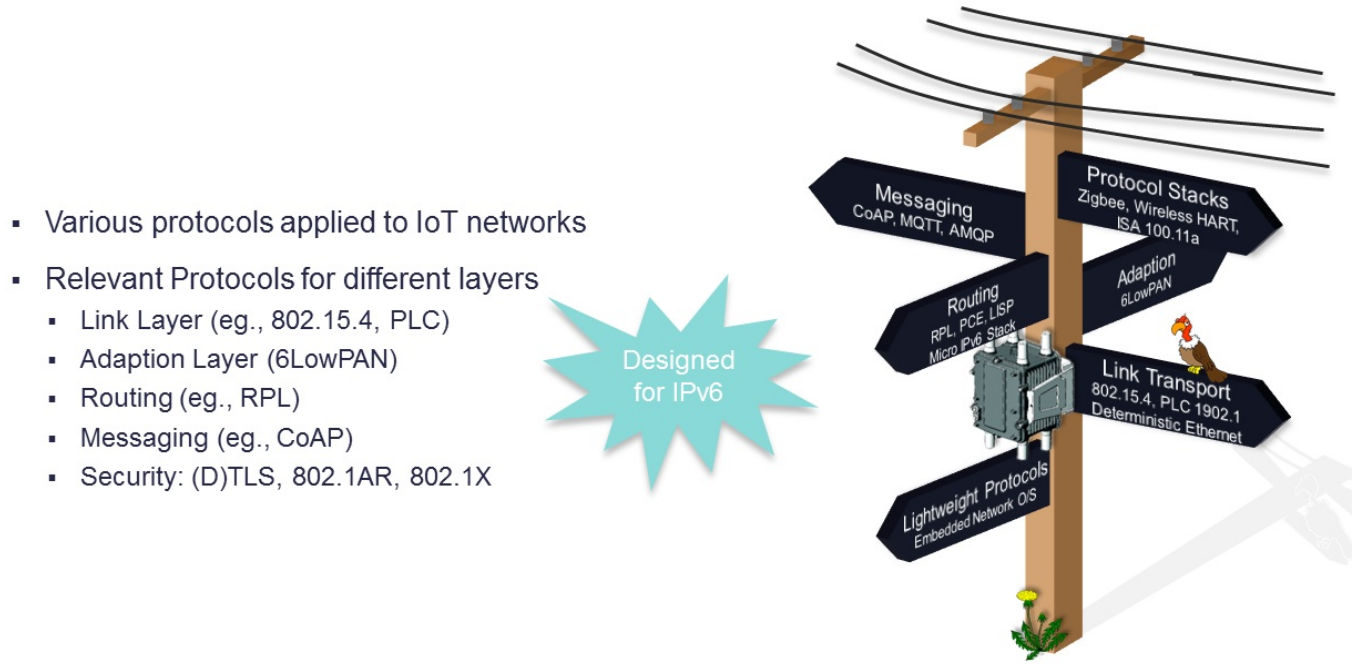
IoT Protocols

As shown in Figure 2, many protocols have been developed at all layers of the International Organization for Standardization (ISO) stack to enable the operation of IoT devices. From messaging protocols such as the Constrained Application Protocol (CoAP), to highly extensible routing protocols such as the Routing

Protocol for Low-Power and Lossy Networks (RPL). The important thing to understand about these protocols is that they have been designed with energy preservation in mind, along with low compute and memory requirements.

The IPv6 Internet is one of the most important enablers of the IoT as it is not possible to add billions of devices to the IPv4 Internet. It follows that the security considerations and implications of IPv6[7] are fundamental to securing the IoT.

Figure 2. Samples of IoT Protocols



Security Challenges within IoT Systems

In many cases, a major disruption of the traditional model brings it's own set of challenges. The following lists some security challenges and considerations in designing and building IoT devices or systems:

- Typically small, inexpensive devices with little to no physical security
- Computing platforms, constrained in memory and compute resources, may not support complex and evolving security algorithms due to the following factors:
  - Limited security compute capabilities
  - Encryption algorithms need higher processing power
  - Low CPU cycles vs. effective encryption

- Designed to operate autonomously in the field with no backup connectivity if primary connection is lost
- Mostly installed prior to network availability which increases the overall on-boarding time

- Requires secure remote management during and after onboarding
- Scalability and management of billions of entities in the IoT ecosystem
- Identification of endpoints in a scalable manner
  - Individual — e.g., Home Smart Meter
  - Group — e.g., All light bulbs in a room/home
  - Scalability challenges of Individual vs. Group
  - Sometimes the location may be more important than the individual identifier (ID)

- Management of Multi-Party Networks
  - For example, Smart Traffic Lights where there are several interested parties such as Emergency Services (User), Municipality (owner), Manufacturer (Vendor)
  - Who has provisioning access?
  - Who accepts Liability?

- Crypto Resilience
  - Embedded devices may outlive algorithm lifetime
  - For example, Smart meters could last beyond 40 years
  - Crypto algorithms have a limited lifetime before they are broken[9]

- Physical Protection
  - Mobile devices can be stolen
  - Fixed devices can be moved

- Tamper Detection techniques and design
  - Always On: High Poll rate, more energy, quick detection
  - Periodic Poll: Less energy, slower detection
  - On-event Push: Minimal energy, no detection

The IoT entities will generally not be a single-use, single-ownership solution. The devices and the control platform on which data may be consumed and shared could have different ownership, policy, managerial and connectivity domains. Consequently, devices will be required to have equal and open access to a number of data consumers and controllers concurrently, while still retaining privacy and exclusivity of data where that is required between those consumers. Information availability while providing data isolation between common customers is critical. We must establish the appropriate identity controls and build trust relationships between entities to share the right information.

There are seemingly competing, complex security requirements to be deployed on a platform with potentially limited resources:

- Authenticate to multiple networks securely
- Ensure that data is available to multiple collectors
- Manage the contention between that data access

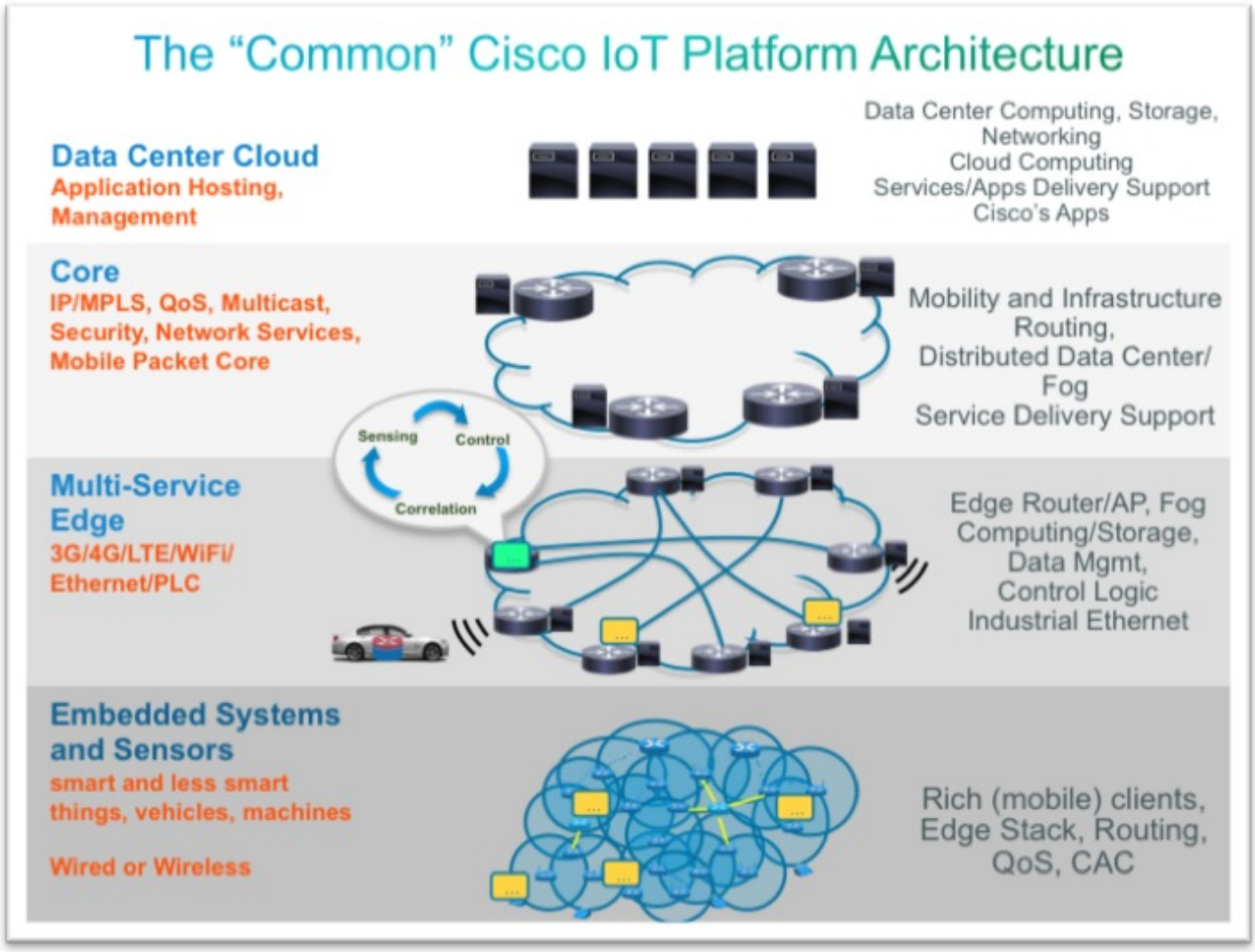
- Manage privacy concerns between multiple consumers
- Provide strong authentication and data protection (integrity and confidentiality) that are not easily compromised
- Maintain availability of the data or the service
- Allow for evolution in the face of unknown risks

These issues have particular relevance in the IoT where secure availability of data is of paramount importance. For example, a critical industrial process may rely on accurate and timely temperature measurement. If that endpoint is undergoing a Denial of Service (DoS) attack, the process collection agent must somehow be made aware. In such an event, the system should be able take appropriate actions in real-time, such as sourcing data from a secondary connection, or delay the information transmission. It must also be able to distinguish between loss-of-data due to an on-going DoS attack and loss of the device due to a catastrophic event in the plant. It might accomplish this by using learning machine techniques (for example, comparing a normal operational state to an attack state previously learned).

Network Layers of IoT Architecture

While many existing security technologies and solutions can be leveraged in a network architecture, especially across the core and data center cloud layers, there are unique challenges in the IoT space. The nature of the endpoints and the sheer scale of aggregation require special attention in the overall architecture to accommodate these challenges. The Cisco IoT/M2M architecture is composed of four layers, some are similar to those described in conventional Cisco network architectures.

Figure 3. IoT/M2M Network Architecture Layers



A. Embedded Systems Layer

As shown in Figure 3, the first layer of the IoT/M2M architecture is comprised of embedded systems, sensors and actuators. As such, these are small devices, with varying operating systems, CPU types, memory, etc. Many of these entities are expected to be inexpensive, single-function devices with rudimentary network connectivity, such as a temperature or pressure sensor. In addition, these devices could be in remote and/or inaccessible locations where human intervention or configuration is almost impossible.

Since the nature of sensors is such that they are embedded in what they are sensingÑone can envisage a new workplace, hospital, or school construction project where these sensors are introduced during the construction phase to collect and monitor data and events. Secondary links will help in cases where the connectivity is lost after the installation teams have left the site.

Additionally, methods must be taken to ensure that the authenticity of the data, the path from the sensor to the collector and the connectivity authentication parameters between the initial installation/configuration of the device, and its eventual presence on the IoT infrastructure cannot be compromised.

B. Multi-Service Edge Layer

The variability in the capabilities of endpoint devices, and their potentially enormous numbers highlight the importance of the multi-service edge in the IoT/M2M architecture. The multi-service edge is multi-modal and supports both wired and wireless connectivity. Even within those two categories, this layer must support many different protocols, such as Zigbee, IEEE 802.11, 3G and 4G, to accommodate a variety of endpoints. In some cases, the protocols used by endpoint devices may not even have any inherent security capabilities at all. It is imperative for security services to protect these inherently insecure endpoints. Additionally, this layer must be modular to scale to meet growth requirements. The components and services offered within one module should be similar so that additional modules can be added in a short span of time.

C. Core Network Layer

The architecture of the core network layer is similar to the architecture deployed in conventional networks. The function of this layer is to provide paths to carry and exchange data and network information between multiple sub-networks. The main differentiator between IoT and conventional core layers is traffic profile. The IoT traffic and data may be different, for example, unique protocols and variable packet size. Security services at the core network ensure that the IoT/M2M system as a whole, and has been hardened to protect against threats such as the following:

- Man-in-the-middle (MITM) is the means by which the attacker can successfully create a connection between two points and eavesdrop into their conversation by relaying the messages it hears from one peer to the other while also capturing the data.
- Impersonation (spoofing) is the means by which an attacker has compromised an identity and thus, through impersonation can send malicious traffic to victim endpoints on the network.
- Confidentiality compromise is the means by which the data that is being relayed can be altered by an attacker.
- Replay attack is the means by which valid data is retransmitted or delayed by an adversary to gain access to an already established session by spoofing their own identity.

D. Data Center Cloud Layer

The architecture of the data center/cloud network layer again is similar to the architectures that are deployed in conventional networks. The function of this layer is to host applications that are critical in providing services and to manage the end-to-end IoT architecture. Again, security services in the data center/cloud network are critical in ensuring that the IoT/M2M system as a whole has been hardened to protect against threats such as the following:

- Denial of Service (DoS) is the attempt by an attacker to make a resource unavailable. A good example

of a resource vulnerable to DoS is the wireless medium. While many technologies exist today to harden the protocols and secure WiFi, Long-Term Evolution (LTE), 3G et al., a simple radio jammer can still be an effective DoS on these wireless media.
- Component and endpoint exploitation is the means by which the attacker can infiltrate a component in the IoT/M2M system (either an endpoint or network element, application or module) and use it to perform further exploits. These attacks have evolved such that a compromise to a single element can lead to further compromise or infiltration within the system. This has been shown by recent attacks such as Stuxnet[1,2] and Duqu[3]. The application servers and devices within this layer may also be exposed to buffer overflow and remote code execution attacks if security hardening and best practices are not followed.

The threats in these layers, whether DoS, transaction replays, or compromised systems typically can be addressed through established cryptographic mechanisms, provisioning of strong identities with credentials to allow them to authenticate into the network, and with strong policies to affect the appropriate access controls.

IoT Threats

IPv6, a foundation of the IoT, is subject to the same attack threats as IPv4, such as smurfing, reconnaissance, spoofing, fragmentation attacks, sniffing, neighbor discovery attacks, rogue devices, man-in-the-middle attacks, and others. Therefore, in the core of the network it requires the same security treatments that exist today for IPv4.

However, the IoT opens a completely new dimension to security. The IoT is where the Internet meets the physical world. This has some serious implications on security as the attack threat moves from manipulating information to controlling actuation (in other words, moving from the digital to the physical world). Consequently, it drastically expands the attack surface from known threats and known devices, to additional security threats of new devices, protocols, and workflows. Many operational systems are moving from closed systems (e.g., SCADA, Modbus, CIP) into IP-based systems which further expands the attack surface

The IoT can be affected by various categories of security threats including the following:

- Common worms jumping from ICT to IoT: Generally limited to things running consumer O/S: Windows, Linux, iOS, Android
- "Script kiddies" or others targeting residential IoT: Unprotected webcams, stealing content, breaking into home control systems
- Organized crime: Access to intellectual property, sabotage, and espionage
- Cyber terrorism: Nuclear plants (For example, Stuxnet virus), traffic monitoring, railways, critical infrastructure

Security in IoT/M2M

As the applications of the IoT/M2M affect our daily lives, whether it is in the industrial control, transportation, SmartGrid or healthcare verticals, it becomes imperative to ensure a secure IoT/M2M system. With continued adoption of IP networks, IoT/M2M applications have already become a target for attacks that will continue to grow in both magnitude and sophistication. The scale and context of the IoT/M2M make it a compelling target for those who would do harm to companies, organizations, nations, and more importantly people. The targets are abundant and cover many different industry segments. The potential impact could span from minor irritant to grave and significant damage to the infrastructure and loss of life.

Although the threats in the IoT environment might be similar to those in the traditional IT environments, the overall impact could be significantly different. That is why there are several efforts in the community to focus on threat analysis[4] and risk assessments to gauge the impact if a security incident or a breach occurs

One of the fundamental elements in securing an IoT infrastructure is around device identity and mechanisms to authenticate it. As mentioned earlier, many IOT devices may not have the required compute power, memory or storage to support the current authentication protocols. Today's strong encryption and authentication schemes are based on cryptographic suites such as Advanced Encryption Suite (AES) for confidential data transport, Rivest-Shamir-Adleman (RSA) for digital signatures and key transport and Diffie-Hellman (DH) for key negotiations and management. While the protocols are robust, they require high compute platformÑ a resource that may not exist in all IoT-attached devices. Consequently, authentication and authorization will require appropriate re-engineering to accommodate our new IoT connected world.

Secondly, these authentication and authorization protocols also require a degree of user-intervention in terms of configuration and provisioning. However, many IoT devices will have limited access, thus requiring initial configuration to be protected from tampering, theft and other forms of compromise throughout its usable life, which in many cases could be years.

In order to overcome these issues, new authentication schemes that can be built using the experience of today's strong encryption/authentication algorithms are required. The good news is that new technologies and algorithms are being worked on. For example, the National Institute of Standards and Technology (NIST) has recently chosen the compact SHA-3 as the new algorithm for the so-called "embedded" or smart devices that connect to electronic networks but are not themselves full-fledged computers[5].

Other elements in security that could be considered include the following:

- Application of geographic location and privacy levels to data
- Strong identities

- Strengthening of other network-centric methods such as the Domain Name System (DNS) with DNSSEC and the DHCP to prevent attacks
- Adoption of other protocols that are more tolerant to delay or transient connectivity (such as Delay Tolerant Networks)[6]

Many of the security considerations for IoT protocols rely on encryption. As new workflows emerge for sensors and elements connected to the Internet, a disparity in time horizons creates an additional gap: devices might outlive the encryption effectiveness. For example, a power meter in a home may last fifty years, where as the encryption protocol might survive half of that time before it is compromised.

Lastly, the communication and the data transport channels should be secured to allow devices to send and collect data to and from the agents and the data collection systems. While not all IoT endpoints may have bi-directional communications, leveraging SMS (automatically or via a network administrator) allows secure communication with the device when an action needs to be taken.

Privacy

Preservation of privacy has been a concern since the dawn of the Internet. IoT will exacerbate the problem because many applications generate traceable signatures of the location and behavior of the individuals. Privacy issues are particularly relevant in healthcare, and there are many interesting healthcare applications that fall within the realm of IoT. We can cite among others the tracking of medical equipment in a hospital, the monitoring of vital statistics for patients at home or in an assisted living facility. In this environment, it is essential to verify device ownership and the owner's identity while decoupling the device from the owner. Shadowing is a mechanism that has been proposed to achieve this. In essence, digital shadows enable the user's objects to act on her behalf, storing just a virtual identity that contains information about her attributes.

Identity management in the IoT may offer new opportunities to increase security by combining diverse authentication methods for humans and machines. For example, bio-identification combined with an object within the personal network could be used to open a door.

Privacy and compliance are intertwined and are under the purview of country regulation. As the technology is evolving so quickly, the consumer must be cognizant of how these issues apply to his or her daily life.

Proposed IoT/M2M Security Framework

To address the highly diverse IoT environment and the related security challenges, a flexible security framework is required. Figure 4 illustrates the security environment from an IoT perspective.

Figure 4. IoT Security Environment

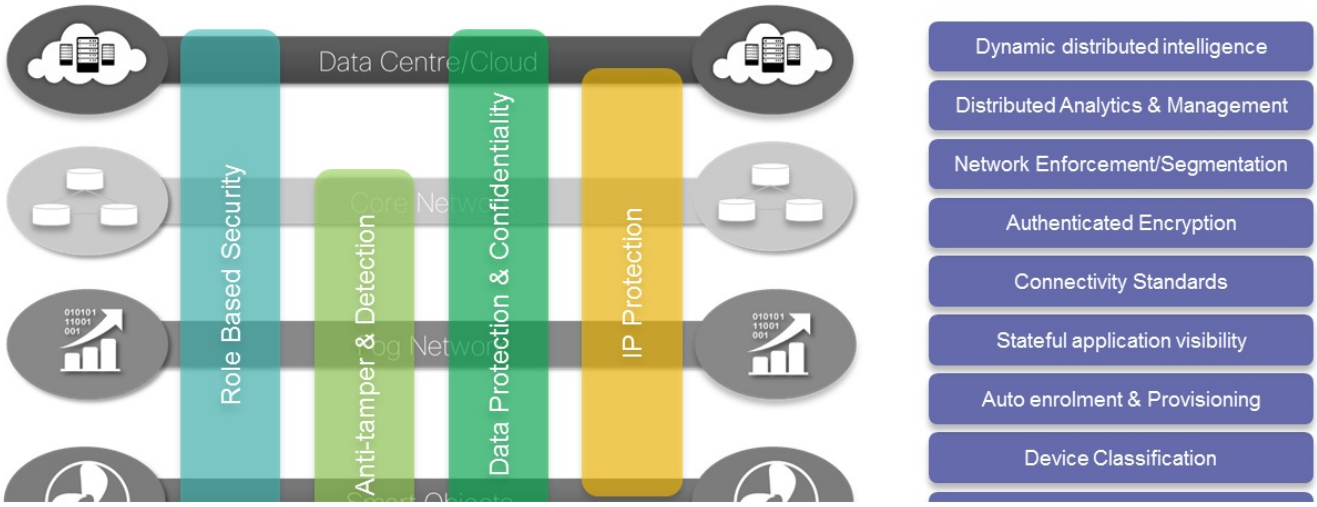
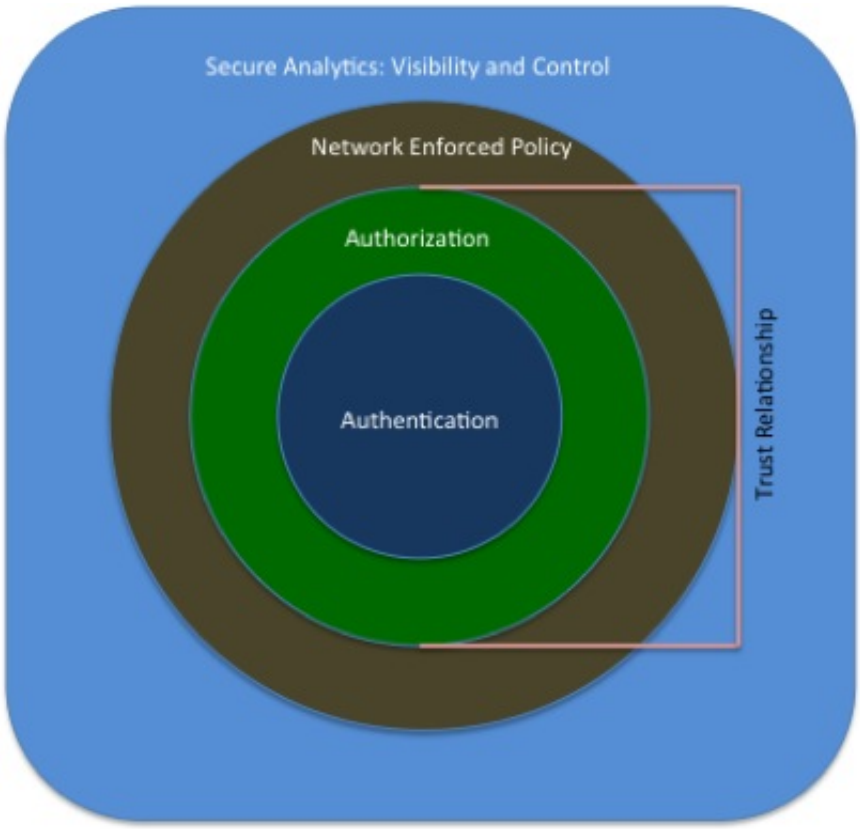


Figure 5 shows a framework to secure the IoT environment and is comprised of four components:

- Authentication
- Authorization
- Network Enforced Policy
- Secure Analytics: Visibility and Control

Figure 5. Secure IoT Framework



Authentication

At the heart of this framework is the authentication layer, used to provide and verify the identify information of an IoT entity. When connected IoT/M2M devices (e.g., embedded sensors and actuators or endpoints) need access to the IoT infrastructure, the trust relationship is initiated based on the identity of the device. The way to store and present identity information may be substantially different for the IoT devices. Note that in typical enterprise networks, the endpoints may be identified by a human credential (e.g., username and password, token or biometrics). The IoT/M2M endpoints must be fingerprinted by means that do not require human interaction. Such identifiers include radio-frequency identification (RFID), shared secret, X.509 certificates, the MAC address of the endpoint, or some type of immutable hardware based root of trust.

Establishing identity through X.509 certificates provides a strong authentication system. However, in the IoT domain, many devices may not have enough memory to store a certificate or may not even have the required CPU power to execute the cryptographic operations of validating the X.509 certificates (or any type of public key operation).

Existing identity footprints such as 802.1AR and authentication protocols as defined by IEEE 802.1X can be leveraged for those devices that can manage both the CPU load and memory to store strong credentials. However, the challenges of the new form factors, as well as new modalities, create the opportunity for further research in defining smaller footprint credential types and less compute-intensive cryptographic constructs and authentication protocols.

Authorization

The second layer of this framework is authorization that controls a device's access throughout the network fabric. This layer builds upon the core authentication layer by leveraging the identity information of an entity. With authentication and authorization components, a trust relationship is established between IoT devices to exchange appropriate information. For example, a car may establish a trust alliance with another car from the same vendor. That trust relationship, however, may only allow cars to exchange their safety capabilities. When a trusted alliance is established between the same car and its dealer's network, the car may be allowed to share additional information such as its odometer reading, last maintenance record, etc.

Fortunately, current policy mechanisms to both manage and control access to consumer and enterprise networks map extremely well to the IoT/M2M needs. The big challenge will be to build an architecture that can scale to handle billions of IoT/M2M devices with varying trust relationships in the fabric. Traffic policies and appropriate controls will be applied throughout the network to segment data traffic and establish end-to-end communication.

Network Enforced Policy

This layer encompasses all elements that route and transport endpoint traffic securely over the infrastructure, whether control, management or actual data traffic. Like the Authorization layer, there are already established protocols and mechanisms to secure the network infrastructure and affect policy that are well suited to the IoT/M2M use cases.

Secure Analytics: Visibility and Control

This secure analytics layer defines the services by which all elements (endpoints and network infrastructure, inclusive of data centers) may participate to provide telemetry for the purpose of gaining visibility and eventually controlling the IoT/M2M ecosystem. With the maturity of big data systems, we can deploy a massive parallel database (MPP) platform that can process large volumes of data in near real time. When we combine this technology with analytics, we can do some real statistical analysis on the security data to pick out anomalies. Further, it includes all elements that aggregate and correlate the information, including telemetry, to provide reconnaissance and threat detection. Threat mitigation could vary from automatically shutting down the attacker from accessing further resources to running specialized scripts to initiate proper remediation. The data, generated by the IoT devices, is only valuable if the right analytics algorithms or other security intelligence processes are defined to identify the threat. We can get better analytical outcome by collecting data from multiple sources and applying security profiles and statistical models that are built upon various layers of security algorithms.

We all know that network infrastructures are becoming more complex. Imagine topologies with both public and private clouds; the threat intelligence and defense capabilities must also be cloud-based. Orchestration

of the visibility, context and control is required to drive accurate intelligence. The components within this layer include the following:

- The actual IoT/M2M infrastructure from which telemetry and reconnaissance data is acquired and gathered
- The core set of functions to coalesce, analyze the data for the purposes of providing visibility, and provide contextual-awareness and control
- The delivery platform for the actual analytics, built from the first two components, discussed above

While the actual IoT/M2M implementations may be different, the framework can be applied to any architecture. The framework is simple and flexible enough to service manned devices as well (e.g., laptops, handheld scanners, etc.) if they reside in the IoT infrastructure.

Can you have an architecture that provides 100 percent protection from threats by leveraging this framework? Unfortunately, that silver bullet does not exist, at least not yet. However, we do believe that big data and analytics platforms will play a key role. Security threats are continuously emerging and require us to develop an architecture that can defend itself against those threats. This security framework provides the foundation from which appropriate security services can be selected. As specific contexts and verticals are considered, gaps can also be identified and addressed.

Conclusion

While the security implications for IoT/M2M constructs are vast, deconstructing a viable IoT/M2M security framework can be the foundation to the execution of security in production environments. The authors have proposed such a framework that may be used in protocol and product development, in addition to, policy enforcement in operational environments.

The authors have also shown how the problem of securing the IoT is much more than IPv6 security. The IoT industry is still evolving, and there is large potential for zero-day attacks. This offers an opportunity to drive the security at the appropriate layer. The embedded endpoint layer is comprised of highly constrained devices, and so far, has limited the growth of malware to this layer. The growth of IP-based sensors corresponds to attack surface growth. This highlights the fact that new security protocols and identification techniques are required, and IoT endpoint security needs to correlate to its enhanced capabilities. Clearly, IoT presents new challenges to network and security architects. Smarter security systems that include managed threat detection, anomaly detection, and predictive analysis need to evolve. In addition, we have offered a point of view on privacy and its implications to security and regulatory compliance.

Acknowledgments

Jazib Frahim (jfrahim@cisco.com)
Principle Engineer

Carlos Pignataro (cpignata@cisco.com)
Distinguished Engineer

Jeff Apcar (japcar@cisco.com)
Distinguished Engineer

Monique Morrow (mmorrow@cisco.com)
CTO-Evangelist-New Frontiers

References

[1] Computerworld, "Siemens: Stuxnet worm hit industrial systems", September 16, 2010.

[2] Steven Cherry with Ralph Langner, "How Stuxnet is Rewriting the Cyberterrorism Playbook", October 2010, IEEE Spectrum.

[3] "Duqu: A Stuxnet-like malware found in the wild, technical report", October 14, 2011, Laboratory of Cryptography of Systems Security

[4] ETSI TR103 167 v0.3.1 "Machine to Machine Communications (M2M); Threat Analysis and Counter-Measures to M2M Service Layer, 2011.

[5] "NIST Selects Winner of Secure Hash Algorithm (SHA-3) Competition", October 2, 2012, http://www.nist.gov/itl/csd/sha-100212.cfm.

[6] Delay Tolerant Networking Research Group: http://www.dtnrg.org/wiki.

[7] Gont, F., "Security Assessment of the Internet Protocol version 6 (IPv6)", UK Centre for the Protection of National Infrastructure.

[8] Source: Cisco IBSG projections, UN Economic & Social Affairs http://www.un.org/esa/population/publications/longrange2/WorldPop2300final.pdf.

[9] Valerie Aurora, "Lifetimes of cryptographic hash functions", 2012, http://valerieaurora.org/hash.html.