Chapter 4
IoT Requirements for Networking Protocols

The success of the Internet is attributed, in part, to the Internet Protocol stack that offers two key characteristics:

- A normalization layer (the IP layer), which guarantees system interoperability while accommodating a multitude of link layer technologies, in addition to a plethora of application protocols. IP constitutes the thin waist of the proverbial hourglass that is the Internet's protocol stack.
- Layered abstractions that hide the specifics of a given layer from the one above or below it. Such abstractions define contracts or "slip surfaces" allowing innovations in one layer to proceed independent of the adjacent layers.

As researchers and technologists started delving into the world of IoT, it was relatively straightforward to justify the benefits of employing a similar layered architectural approach for the IoT protocol stack. However, a topic of lively debate emerged in whether the Internet Protocol stack was suited for the IoT or whether a new stack was needed. In the late 1990s and early 2000s, many researchers in the field of wireless sensor networks did not shy away from denouncing IP networking as unsuitable for that application domain.

It was deemed that the requirements of IoT were sufficiently different to warrant a white canvas approach, rather than reusing the Internet technology, which fell short of addressing the requirements in a number of areas. The decade and a half that followed witnessed an evolution of the IP stack to address many of the cited requirements for sensor networks and the shortcomings of IP technologies at the time.

In this chapter, we will discuss the key IoT requirements and their impact on each of the layers of the protocol stack. In the next chapter, we take a layer-by-layer view and discuss the industry's efforts, to date, to address these requirements. We will also discuss the gaps that remain for further study and require future solutions.

A. Rayes, S. Salam, *Internet of Things from Hype to Reality*,
https://doi.org/10.1007/978-3-030-90158-5_4

4.1 Support for Constrained Devices

The devices that are to be connected to the network in the IoT span a wide gamut of capabilities and characteristics along the facets of computational power, mobility, size, complexity, dispersion, power resource, placement, and connectivity patterns. These and other device characteristics impose a set of requirements and restrictions on the network infrastructure used for interconnecting them. In particular, the devices' computational capabilities, as well as their power resources, introduce challenging requirements for IP networking technologies.

Stepping back and examining the devices that have traditionally connected to the Internet, one can easily categorize them as homogeneous in terms of being fully capable computers or peripherals (e.g., servers, desktops, laptops, printers, etc.) that have an endless source of power (e.g., mains powered or equipped with rechargeable batteries). In the IoT, this homogeneity no longer holds: on one end of the spectrum are devices with very limited processing power which scavenge energy from their environment (e.g., pressure sensors), and on the other end are devices with powerful processors, a generous amount of memory, and replenishable power sources (e.g., smartphones).

Small devices with limited processing, memory, and power resources are referred to as constrained devices. Generally speaking, a constrained device is limited in one or more of the following dimensions:

- Maximum code complexity (ROM/Flash).
- Size of run-time state and buffers (RAM).
- Amount of computation feasible in a specific period of time ("processing power").
- Available power resources.
- Management of user interface and accessibility in deployment (ability to set security keys, update software, etc.).

IETF RFC 7228 defines a taxonomy of constrained devices based on the first two dimensions above, which recognizes three classes of devices as depicted in Table 4.1.

Class 0 devices are the most severely constrained in memory and processing power. In general, such devices do not have the resources to connect to an IP network directly and will leverage the services of helper devices such as proxies or gateways for connectivity. For example, sensor motes fall under this class.

Table 4.1 Classes of constrained devices in RFC 7228

Name	Data size	Code size
Class 0	≪10 KB	≪100 KB
Class 1	~10 KB	~100 KB
Class 2	~50 KB	~250 KB

Class 1 devices are highly constrained in terms of code space and processing capacity; however they are capable of connecting to an IP network directly, without the help of gateways, as long as they are "parsimonious with state memory, code space, and often power expenditure for protocol and application usage." As such, these devices face challenges in running certain demanding IPs such as BGP, OSPF, HTTP, or Transport Layer Security (TLS) and in exchanging data using verbose data serialization formats such as XML.

Class 2 devices are less constrained when compared to the first two classes and are capable of running the same IP stack that runs on general compute nodes today. Nevertheless, these devices can still benefit from lightweight and efficient communication stacks since the resources may then be directed toward applications in lieu of networking.

Another dimension that characterizes constrained devices is power and/or energy resource constraints. These could be attributed to a number of factors such as the device size, primary mode of use, cost, operational environment, etc. Again, with this dimension, there is a spectrum of possibilities ranging from devices that harvest energy from the environment to battery-powered devices where the batteries are replaceable or rechargeable, to non-field replaceable battery-powered devices (which are discarded past the battery's lifetime), and to mains-powered devices. Energy consumption is a major issue for IoT devices. Research studies suggest that communication is over three orders of magnitude more expensive in terms of energy consumption than performing local processing functions. This is especially the case when wireless communication is used, where the radio takes the lion's share of the energy consumed by the device. To this reason, a common strategy employed by power-constrained devices is to remain in sleep mode with no network connectivity for extended periods of time and to connect only long enough to send the local data either based on periodic timers or asynchronous triggers (e.g., when new data is present or an event is detected).

To address the requirements of constrained devices, lightweight, energy-efficient, and bandwidth-conscious communication protocols are required across all the layers of the protocol stack.

4.2 Massive Scalability

Based on an estimate conducted by Cisco, about 99.4% of the physical objects in the world, which could potentially be connected to the Internet, are still unconnected. Conversely, this means that only about 10 billion out of approximately 1.5 trillion global objects are connected. The number of devices connected to the Internet surpassed 26 billion devices in 2020 (Fig. 4.1). The majority of this growth continues to be due to smart objects and "things" connecting to the Internet. This massive scalability imposes requirements on various aspects of the IoT protocol stack, in the areas of device identification and addressing, namely resolution, security, control plane (e.g., routing protocols), data plane forwarding, as well as manageability.

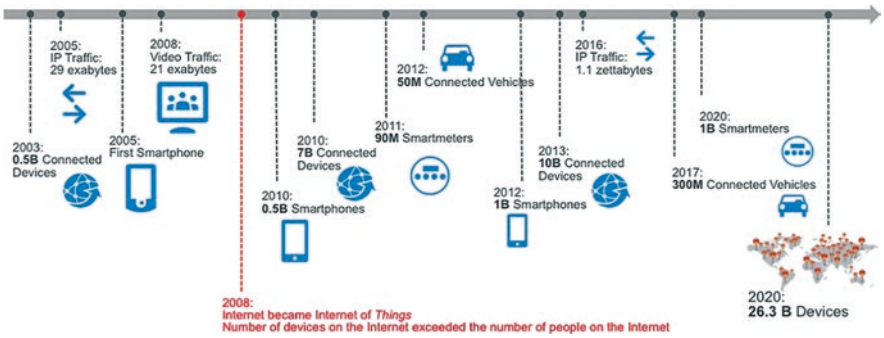

Fig. 4.1 Growth of connected devices. (Source: Cisco)

4.2.1 Device Addressing

The goal of the IoT is to build a uniform network that integrates and unifies all the communication systems between smart objects in the world. To realize the full potential of this vision, the interconnected things need to be individually addressable for ubiquitous communication between systems. In many current deployments of smart objects, the interconnection of things to the Internet, when available, is through gateways or proxies. In this sense, the connected things are proverbial second-class citizens of the Internet. Realizing the IoT vision requires that a global IP address be assigned to each one of the billions of devices that will be connected. Taking into account the fact that the IPv4 address space was completely depleted by February 1, 2011, it becomes clear that the massive scalability of the IoT will accelerate the transition of the Internet to IPv6.

4.2.2 Credentials Management

Security credentials management (e.g., shared key distribution, certificate management, etc.) poses a significant challenge in today's Internet. The addition of billions of devices to the network with IoT will only compound the problem further. Manual mechanisms currently employed for credentials management (e.g., through preconfiguration) are not going to be viable in IoT due to two reasons: the sheer number of devices and the limitations in (or complete lack of) user interfaces on constrained devices. The number of devices renders the use of pre-shared keys impractical for production deployments, especially when the devices have rudimentary user interfaces or no user interface at all.

The massive scalability of the IoT calls for lightweight, low-touch, and highly automated credentials management mechanisms.

4.2.3 Control Plane

The Internet encompasses diverse networks running different control plane protocols for the purpose of discovering topology information, communicating connectivity status or link health, signaling session or connection state, guaranteeing quality of service, and, among other things, quickly reacting to faults. These protocols maintain distributed state that is synchronized using message exchanges between peering nodes. In some cases, these peering relationships are hierarchical in nature (e.g., a client-server model) or flat (e.g., overlay peers). The behavior of the control plane functions together with the syntax and semantics of the messages exchanged defines the specifics of the control plane protocol. As the number of nodes participating in a given protocol increases, both the amount of state to be maintained by each node increases and the volume of messages required for keeping the distributed state tables in synchronization grows. Beyond a specific limit, attempts to scale a specific control plane protocol typically lead to adverse side effects on the protocol's convergence time, the node resources, and the overall network response. The scalability of the IoT calls for elastic control plane mechanisms that can accommodate the massive number of connected devices.

4.2.4 Wireless Spectrum

As the Internet of Things continues to evolve, one fact remains constant: these things require connectivity. This global network of objects, sensors, actuators, etc. must be connected to the Internet in some way, and in many cases wirelessly. The wireless spectrum is a finite resource, and the licensed portion of this spectrum is both expensive and scarce. With billions of devices coming online over the coming decade or so, many of these devices will be contending for the airwaves.

As of now, many IoT systems operate in unlicensed radio frequencies, namely, the industrial, scientific, and medical (ISM) bands, for example, the 900 MHz band for Electronic Product Code (EPC), one of the standards for radio-frequency identification (RFID); the 13.56 MHz band for near-field communications (NFC) supporting mobile payments; and the sub-125 kHz band for physical security systems (video surveillance and access control). These technologies achieve connectivity using a range of different, and in some ways competing, wireless protocol standards, such as Zigbee, Z-Wave, Bluetooth LE, and Wi-Fi, all of which were designed to work in the unlicensed spectrum. There are no spectrum bottlenecks for these bands yet, even though Wi-Fi services are approaching the point where they are maximizing the number of channels that can be fit into the allotted spectrum. However, when it comes to the licensed bands used for cellular communication (e.g., the GSM bands defined in 3GPP TS 45.005), the bottlenecks become more pronounced, especially with the accelerating growth in data traffic over cellular networks. The term "spectrum crunch" has been used in recent years to refer to this

issue. There are two variables at play here: growth in the number of endpoints as well as growth in the volume of traffic per endpoint, both of which contribute to the spectrum crunch phenomenon. Research by Cisco shows that globally, mobile M2M connections grew from 495 million in 2014 to more than 3 billion in 2019, a sevenfold growth. Global mobile data traffic grew 69% in 2014 reaching 2.5 exabytes per month at the end of 2014, up from 1.5 exabytes per month at the end of 2013. Further, global mobile data traffic increased nearly tenfold between 2014 and 2019 (Fig. 4.2).

4.3 Determinism

One of the value propositions of IoT is that the technology will allow for better observation and monitoring of the physical world and will also enable the automated change of that world through closed-loop actuation. IoT opens up the door for supporting use cases that demand mission-critical networking with high requirements for real-time response as well as overall network, protocol, and device robustness. Some of these use cases emerge from industrial automation, such as monitoring systems, movement detection systems for use in process control (i.e., process manufacturing), and factory automation (i.e., discrete manufacturing). Other use cases have a much broader scope that spans mission-critical automation (e.g., rail control systems), motion control (e.g., wind turbines), vehicular networks (e.g., infotainment, power train, driver assistance), etc. With the increasing demand for connectivity and multimedia in transportation in general, use cases and applications are emerging in all elements of the vehicle from head units to rear seat entertainment modules, and to amplifiers and camera modules. While these use cases are aimed at less critical applications than industrial automation, they do share common requirements.

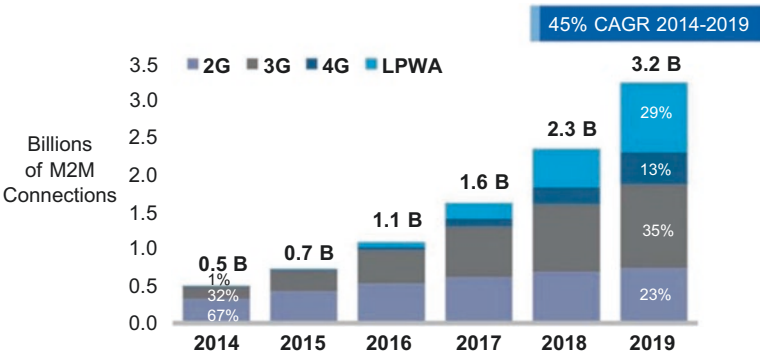

Fig. 4.2 Global machine-to-machine growth and migration from 2G to 3G and 4G. (Source: Cisco VNI Mobile, 2015)

These use cases all share the common requirement to support real-time information transfer: the time it takes for each packet to traverse a path from its source to its destination should be determined; that is, the process must be deterministic. Systems with control loops involving endpoints communicating over a network can function properly only if the networks connecting those endpoints guarantee determinism (imagine what would happen if a network delays a packet carrying a control variable for a high-speed CNC mill).

In this context, a network is said to support determinism and is thereby deemed to be a "deterministic network," if the worst-case communication latency and jitter of messages of interest are decidable based on a reasonable model of the network. A model is considered reasonable when it sufficiently represents reality for the target use cases of the networking system. Determinism does not imply speed. In control functions, both speed and determinism are required. Speed is required to attain the highest possible throughput. Determinism, on the other hand, is required to specify a level of quality for the throughput, i.e., the highest-speed throughput that is in fact usable by the application.

Deterministic Networking enables the migration of applications that have so far relied on special-purpose non-packet-based (fieldbus) technologies (e.g., HDMI, CAN bus, Profibus, etc.) to Internet Protocol technologies to support both these new applications, in addition to existing IP network applications, over the same physical network (Fig. 4.3). When applied in the context of industrial applications, this leads to what is dubbed as the "OT/IT" convergence. Operational technology (OT) refers to industrial networks, which, due to their different goals, have evolved in silo but in a manner that is substantially different from information technology (IT) networks. With OT, the focus has been on transporting fully characterized traffic flows, over a small area (e.g., plant floor), in a well-controlled environment with a bounded latency, extraordinarily low frame loss, and very narrow jitter.

Experience with custom control and automation networks, as well as proprietary audio/video networks, has shown that these applications require one or more of the following characteristics: time synchronization of all hosts and network elements (routers, bridges, etc.) and accurate in the range of 10 ns to 10 μs, depending on the application. The applications also require support for critical packet flows that need guarantees of the minimum and maximum latency end-to-end across the network. Such flows can be either unicast or multicast and can in total consume more than half of the available bandwidth of the network, thereby eliminating the possibility of relying on over-provisioning. The applications mandate packet loss ratios that are at least in the range of 1.0e−9 to 1.0e−12. Furthermore, the traffic for these applications cannot be subjected to throttling, congestion feedback, or stochastic network-imposed transmission delay.

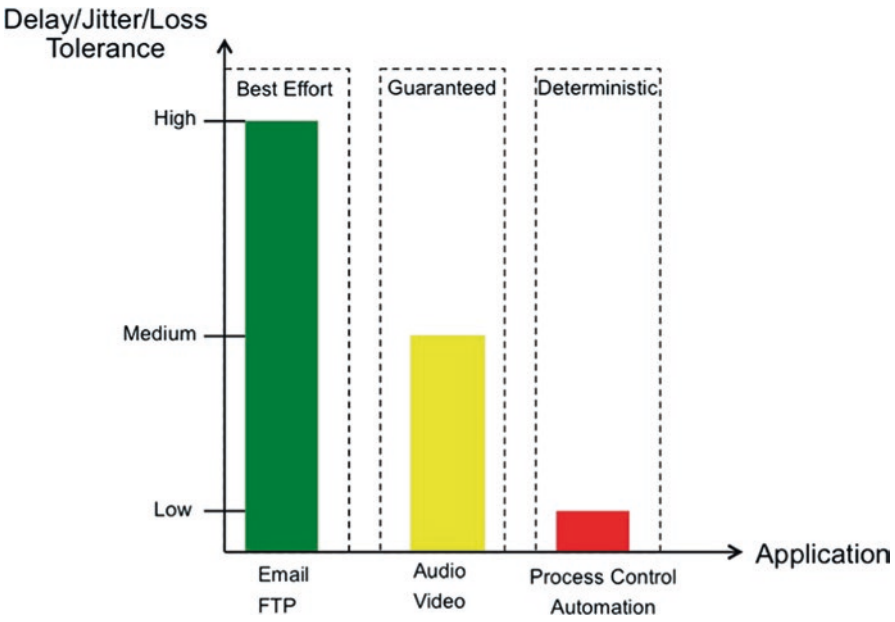

Fig. 4.3 Deterministic vs. guaranteed vs. best effort traffic

4.4 Security and Privacy

The ubiquity of IoT and its potential to extend into all aspects of human life, whether in transportation, healthcare, home automation, industrial control, etc., makes guaranteeing security and privacy paramount. With traditionally offline systems and applications being connected to the Internet, they quickly become targets for attacks that will only continue to grow in magnitude and sophistication. Such targets cover a multitude of industry segments, and the potential impact of security attacks could lead to significant damage and even loss of life.

While the threats in IoT may, at the outset, seem largely similar to those in more traditional IT environments, the potential impact of those threats is more profound. This is why threat analysis and risk assessment efforts are key in IoT to measure the impact of a security incident or breach.

A fundamental pillar in securing the IoT is around mechanisms to authenticate device identity. As was discussed in Sect. 4.1, many IoT devices are constrained devices, which lack the required processing, memory, storage, and power requirements to support state-of-the-art authentication protocols. The state-of-the-art encryption and authentication protocols are based on cryptographic suites such as Advanced Encryption Standard (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 these protocols are battle-proven in deployments, they suffer from two shortcomings when it comes to

applying them to IoT. The first shortcoming is that these protocols are resource hungry and generally demand high-capability compute platforms. Appropriate reengineering is required to accommodate constrained devices. The second shortcoming is that the authentication and authorization protocols are high-touch, requiring user input for provisioning and configuration. In many IoT deployments, access to the devices will be limited or impractical, thereby requiring that the initial configuration be tamper-proof throughout the usable lifespan of the devices, and such lifespan could extend to many years.

In order to address these shortcomings, new lightweight authentication and authorization protocols are required which leverage the experience of today's strong encryption/authentication algorithms but are capable of running on constrained devices.

Encryption is the cornerstone of network security protocols. The effectiveness of encryption algorithms generally decreases with time due to a number of factors including Moore's Law (availability of stronger compute to crack the encryption), public disclosure of inherent vulnerabilities with prolonged exposure to attacks, wide adoption (which increases the attack surface), etc. This creates an interesting predicament for the use of encryption in IoT: deployed devices may outlive the effectiveness of the encryption mechanisms embedded within them. For instance, a smart meter in a home can operate for 50 years, whereas the encryption protocol may lose its effectiveness in about half of that time.

Other aspects of security that need to be considered for IoT include:

- Data privacy levels and geo-fencing of data (i.e., limiting access to data to specific locales).
- Strong identities.
- Strengthening of base network infrastructure such as the Domain Name System (DNS) with DNSSEC and DHCP to prevent attacks.
- Adoption of protocols that are more tolerant to delay or transient connectivity (such as delay-tolerant networks).

Privacy is a major issue even in today's Internet. User data is collected for a multitude of purposes such as targeted advertisements, purchase recommendations, and even national security. IoT will exacerbate the importance of preserving privacy because many applications generate traceable signatures of the behavior of individuals and their physical location. Some IoT applications even involve highly sensitive personal information, such as medical records. For these types of applications, it is imperative to decouple the device from the owner's identity while still providing robust mechanisms for device ownership verification and device identity authentication. Shadowing is one mechanism proposed to achieve this. Effectively, digital shadows enable the user's objects to act on his or her behalf, storing just a virtual identity that contains information about his or her attributes. As a matter of fact, identity management in the IoT paves the way to increase security by applying a combination of diverse authentication methods for humans and machines. For instance, biometric data combined with a physical object could be used as grant access by unlocking a door.

The importance of security in IoT cannot be overstated. More details on this topic are covered in Chap. 8.

4.5 Application Interoperability

M2M deployments, in one form or another, have existed for over two decades now. However, the vision of the Internet of Things is far from being a reality, and the technology is yet to realize its full market potential. The complexity of developing, deploying, and managing IoT applications remains a key challenge for the industry. It constitutes a challenge for network operators who are trying to offer profitable services tailored to the IoT market, for application developers building vertical-specific applications, as well as for service providers who are trying to speed time to market, reduce costs, and simplify robust application deployment. This complexity drives up the cost of building IoT solutions.

The problem of complexity, and associated high cost, can be attributed in part to the closed nature of the solutions, which are developed in vertical-specific silos, thereby leading to each solution provider having to implement all the building blocks required for a minimum viable product, as opposed to reusing standard and open components. The resulting solutions are almost ubiquitously characterized by having strong coupling between application entities. Here, we use the term application entity to refer to an instance of application logic that may be implemented in hardware (analogue or digital), software, firmware, etc. Thus, an application entity denotes any IoT endpoint responsible for producing or consuming data and spans the entire gamut from a sensor/actuator to a cloud application.

The closed nature of existing IoT solutions renders them not only expensive to implement initially but also expensive and difficult to maintain and evolve over time. This is primarily because application code often needs to be updated or changed in the scenario where a device is swapped with another that is functionally equivalent albeit manufactured by a different vendor, let alone the scenario where a new device type needs to be integrated into the solution.

The above challenges lead to the requirement for application-level interoperability for the IoT. This requirement can be further broken down into requirements for abstractions and standard application programmatic interfaces (APIs) as well as requirement for semantic interoperability.

4.5.1 Abstractions and Standard APIs

Realizing the full vision of the IoT will be difficult unless the application programmatic interfaces (APIs) that control the functionalities of the devices and smart objects adhere to common standards that guarantee interoperability. To reach full API interoperability, the industry must converge on mechanisms for identifying the

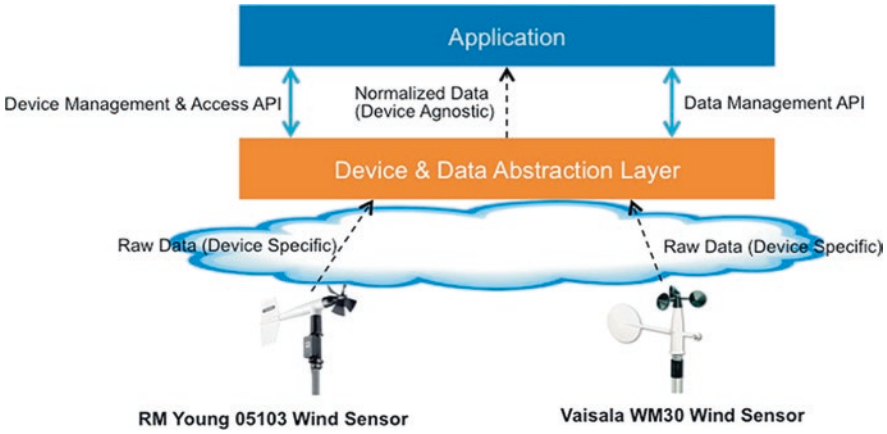

Fig. 4.4 Abstractions and APIs

data that application entities will share and methods for sharing it. APIs expose the data that enables disparate devices to be composed in innovative ways to create new and interesting workflows. With the availability of standard APIs, it is possible to introduce abstractions for common IoT functions, including:

- Device management (activation, triggering, authentication, authorization, software/firmware update, etc.)
- Data management (read, write, subscribe, notify, delete, etc.)
- Application management (start, stop, debug, upgrade, etc.)

The abstractions provide logical representations of the functions while hiding all implementation nuances and variations. They define service contracts that are governed by the syntax and semantics of the APIs and which formally specify the methods for interaction with modules supplying those functions. In other words, the use of standard APIs introduces "slip surfaces" that eliminate coupling between functionally discrete modules of a given IoT solution. This allows modules supplied by different IoT vendors to seamlessly interwork and integrate into a cohesive system. A given module can be replaced by another supplied by a different vendor as long as it subscribes to the standard API governing the associated slip surfaces between the system's building blocks (Fig. 4.4).

4.5.2 Semantic Interoperability

Semantic interoperability guarantees that application entities in the IoT can access and interpret data unambiguously. Providing unambiguous data descriptions that can be machine processed and interpreted by application entities is one of the key enablers of automated information communications and interactions in IoT.

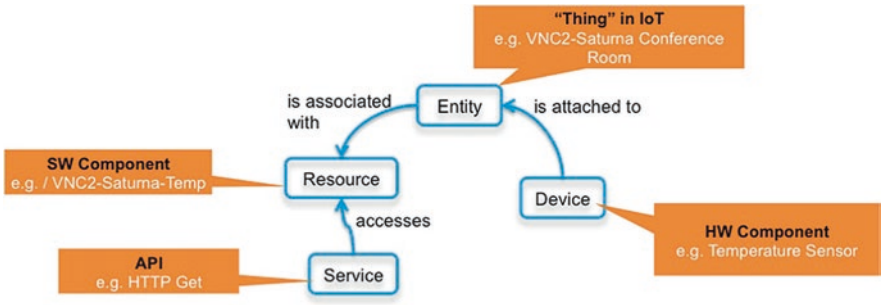

Fig. 4.5 Simple IoT ontology

Without semantic interoperability among communicating systems, sharing IoT data in a useful way is impossible. Semantic interoperability guarantees a common vocabulary that paves the way for accurate and reliable communication between applications and systems. This fluent machine-to-machine communication depends on the ability of different systems to map data to shared semantics, or meaning. If we were to use the analogy of a pyramid to visualize the different tiers of application interoperability, the base of that pyramid would be syntactic or structural interoperability: it defines the structure or format of data exchange between applications. Structural interoperability is a prerequisite; it is necessary but not sufficient for two applications to successfully work together. The top part of the pyramid is reserved to semantic interoperability. It deals with the content of the messages exchanged and their associated meaning, not just the message formats.

Semantic interoperability can be achieved in a number of ways. One is through the development of pervasive and common information models, or ontologies (Fig. 4.5), that capture the knowledge associated with a specific vertical domain. Another is through providing semantic mediators, or translators, that perform conversion of the information to a format that the application entity understands.

4.6 Summary

The Internet Protocol (IP) stack was among the factors that contributed to the success of the Internet. While this IP stack provides a strong foundation for building the IoT, a number of shortcomings need to be addressed to meet the peculiar requirements of IoT. These requirements include support for resource-constrained devices that have very limited compute capabilities and limited power; support for the massive scalability of IoT, with billions of connected devices; the need for deterministic networks to support real-time mission-critical applications; the requirement for lightweight security protocols and ensuring data privacy; and finally the requirement for application interoperability through the use of APIs and unified data semantics.

Problems and Exercises

1. What are "constrained" devices? Name their classes and characteristics.
2. What makes a network "deterministic"?
3. In what three areas does the massive scalability of IoT impact networking protocols?
4. What is the importance of standard APIs in the success of IoT?
5. Why is scalability a major requirement for IoT protocols?
6. What is an ontology? Why are ontologies applicable in the IoT?
7. Name three key IoT requirements that have impact on networking protocols.
8. What characteristics of the IP stack contributed to the success of the Internet?
9. Was the choice of the Internet as the underlying network for IoT always a given or agreed upon fact?
10. Name the various options by which IoT devices can be supplied with power.
11. Describe the characteristics of Class 0-constrained devices.
12. What is "semantic interoperability"? Why is it important in IoT?
13. How does scalability impact the network control plane? Explain the various dimensions impacted.
14. How much of the IPv4 address space is still available for allocation?
15. What common IoT functions can be abstracted through APIs in order to simplify application development and improve the time to market new IoT applications and services?
16. What types of applications can be migrated to IP technologies with the advent of Deterministic Networking?
17. Which is more expensive in terms of power consumption: Communication or local processing? What does this imply to IoT devices?
18. How does the addition of billions of devices to the Internet affect the wireless spectrum?
19. How does the complexity of developing, deploying, and managing IoT applications today affect the state of the industry?
20. What makes existing credentials management techniques inadequate for IoT?
21. What are two shortcomings of the state-of-the-art security protocols (for authentication/authorization/encryption) when applied to the IoT?

References

1. D. Estrin, R. Govindan, J. Heidemann, S. Kumar, Next century challenges: Scalable coordination in sensor networks, in *MobiCom '99: Proceedings of the 5th Annual ACM/IEEE International Conference on Mobile Computing and Networking*, (ACM, New York, 1999), pp. 263–270
2. Bormann, et al., "Terminology for Constrained-Node Networks". Internet Engineering Task Force RFC 7228. May 2014

3. V. Cantoni, L. Lombardi, P. Lombardi, Challenges for data Mining in Distributed Sensor Networks, in *18th International Conference on Pattern Recognition (ICPR'06)*, (2006), pp. 1000–1007
4. J. Bradley, J. Barbier, D. Handler, Embracing the Internet of Everything To Capture Your Share of $14.4 Trillion, Cisco Whitepaper, (2013)
5. The Zettabyte Era: Trends and Analysis, Cisco Whitepaper, (June 2016)
6. D. Evans, The Internet of Things – How the Next Evolution of the Internet is Changing Everything, Cisco Whitepaper, (April 2011)
7. "Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2014–2019", Cisco Whitepaper, February 2015
8. http://www.ieee802.org/802_tutorials/2012-11/8021-tutorial-final-v4.pdf, IEEE 802.1 Tutorial on Deterministic Ethernet, November 2012
9. N. Finn, P. Thubert, "Deterministic Networking Problem Statement", draft-finn-detnet-problem-statement-01, work in progress, (October 2014)
10. W. Steiner, N. Finn, Deterministic Ethernet: Standardization in Progress and Beyond, RATE Workshop, (December 2013)
11. P. Barnaghi et al., Semantics for the internet of things: Early progress and back to the future. Int. J. Semant. Web. Inf. Syst **8**(1) (2012)
12. Securing the Internet of Things: A Proposed Framework, Cisco Whitepaper