CHAPTER 4

FUNDAMENTAL IoT MECHANISMS AND KEY TECHNOLOGIES

This chapter looks briefly at some fundamental issues and technologies that have to be considered in the context of Internet of things (IoT) design and deployment. In fact, there are indeed many issues that have to be considered; only a small set of such issues are covered here as a way to highlight some of the underlying logical and technological infrastructure needed. A hybrid view of the IoT both as a service (application) concept and as an infrastructure is utilized in this discussion.

4.1 IDENTIFICATION OF IoT OBJECTS AND SERVICES

There are a number of key underpinning issues that come into play in the (architectural) design and field deployment of IoT applications. The discussion that follows is synthesized from a number of published documents including References 1–3 among others.

An important first issue is the identification of objects and services. There are various types of identifiers with different purposes and practicality. Globally unique identifiers are highly desirable. Identification codes can be classified as (i) object IDs (OIDs) and (ii) communication IDs. Examples of the former include but are not limited to radio frequency identification (RFID)/electronic product code (EPC),

Building the Internet of Things with IPv6 and MIPv6: The Evolving World of M2M Communications, First Edition. Daniel Minoli.

content ID,[1] telephone number, and uniform resource identifier (URI)/uniform resource locator (URL); examples of the latter include media access control (MAC) address, network layer/IP address, and session/protocol ID. A number of researchers advocate defining an identity layer for objects that is logically independent of the networking addresses; according to these proponents, the IoT should be identity oriented. One such practical, but not necessarily elegant, approach might be to use RFIDs physically attached to the objects in question that would act as electronic ID for objects to which they are linked.

Among other identification approaches, one can use the general approach briefly described in Chapter 1. There we noted that it is both desirable and feasible for all objects to have a permanent unique identifier, an OID. It is also desirable as well as feasible for all end-point network locations and/or intermediary-point network locations to have a durable, unique network address (NAdr); the IPv6 address space enables the concrete realization of these location identification goals. When objects that have enough intelligence to run a communications protocol stack (so that they can communicate), are placed on a network, these objects can be tagged with a NAdr.

Every object then has a tuple (OID, NAdr) that is always unique, although the second entry of the tuple may change with time, location, or situation. In a stationary, non-variable, or mostly static environment, one could opt, if one so chose, to assign the OID to be identical to the NAdr where the object is expected to attach to the network; that is, the object inherits the tuple (NAdr, NAdr). In the rare case where the object moved, the OID could then be refreshed to the address of the new location; that is, the object then inherits the tuple (NAdr', NAdr'). However, there is a general trend toward object mobility, giving rise to a dynamic environment; hence, to retain maximal flexibility, it is best to separate, in principle, the OID from the NAdr and thus assign a general (OID, NAdr) tuple where the OID is completely invariant; however, the OID can still be drawn from the NAdr space, that is, from the IPv6 address space.

The basic requirement for an identification scheme is that it affords global uniqueness. Additionally, it is useful to have mechanisms for hierarchical grouping to deal with large populations. The aggregation feature of IPv6 address provides such hierarchical grouping. For a number of applications, there is a need to map/bind IP addresses (communications IDs) with other relevant OIDs. Additionally, modern layered communication architectures also require addressing and processing capabilities at several layers, for example, at the Data Link Layer, at the Network Layer, at the Transport (Protocol ID), and at the session/application layer. Naturally, there is also a desire for simplicity. Some argue that different identification schemes are required for different applications. For example, the information related to things such as books, medicine, and clothes may not require global identification because revocation lists are required (e.g., some objects may eventually be consumed and/or destroyed).

[1]The content ID, defined by the Content ID Forum, is an identifier that is typically attached to a content-based object. It can specify and distinguish digital content, being a complete set of attribute information about a content object stored as metadata including, among other aspects, the nature of the contents, rights-related information, and information about distribution.

An example of IDs for objects is the above-mentioned EPC used in the RFID/sensor context. An EPC is a number assigned to an RFID tag representative of an actual EPC. Their value is that they have been carefully characterized and categorized to embed certain meanings within their structure. Each number is encoded with a header, identifying the particular EPC version used for coding the entire EPC number. An EPC manager number is defined, allowing individual companies or organizations to be uniquely identified; an object class number is present, identifying objects used within this organization, such as product types. Finally, a serial number is characterized, allowing the unique identification of each individual object tagged by the organization (4). An EPC is a unique identification code that is generally thought of as the next generation of the traditional bar code. Like the bar code, EPC uses a numerical system for product identification, but its capabilities are much greater. An EPC is actually a number that can be associated with specific product information, such as date of manufacture and origin and destination of shipment. This provides significant advantages for businesses and consumers. The EPC is stored on an RFID tag, which transmits data when prompted by a signal emitted by a special reader. Note that EPC and RFID are not interchangeable—there are numerous RFID applications that have nothing to do with the EPC, such as E-Z Pass use at tollbooths (5).

In addition to OID, there may be a need for object naming. Domain name system (DNS) is one example of a mechanism for Internet-based naming; however, currently one only identifies the specific server in which the contents are stored; the data itself is not named. In the IoT context, some proponents have argued for the advantages of identifying information by name, not by node address. DNS is used to map the "human-friendly" host names of computers to their corresponding "machine-friendly" IP addresses. Hence, one is able, for example, to access the server (or large farm of servers) of CNN, Google, and so on, simply by the term www.cnn.com and so on. To some large degree, object name service (ONS) will also be important in the IoT to map the "thing-friendly" names of object which may belong to heterogeneous name spaces (e.g., EPC, uCode, and any other self-defined code) on different networks (e.g., TCP/IP network) into their corresponding "machine-friendly" addresses or other related information of another TCP/IP network (1). However, a "thing" or an object in an IoT world may be a lot more mundane and modest in scope/function (say, than CNN, Citibank, United Airlines, Ford), such that it does not need to have its own name, since very few people may be interested in that specific thing. For example, a large villa may have, for argument's safe, a dozen security sensors. While it is true that they could be named "Smith-villa-front-door sensor," "Smith-villa-front-gate sensor," "Smith-villa-back-door sensor," "Smith-villa-garage-door sensor," very few people besides Mr. Smith or Mr. Smith's security company will ever want to specifically identify these objects by name. Nonetheless, object naming service for IoT applications needs to be developed, at least for a set of applications.

For some applications, especially where there is a need for simple end-user visibility of a small set of objects (i.e., where the objects are few and discretely identifiable – a home's thermostat, a home's refrigerator, a home's lighting system, a pet of the owner), the object may be identified through Web Services (WSs). WSs provide

standard infrastructure for data exchange between two different distributed applications. Lightweight WS protocols are of interest; for example, the representational state transfer (REST) interface may be useful in this context. REST is a software architecture for distributed systems to implement WSs. REST is gaining popularity compared with more classical protocols such as simple object access protocol (SOAP) and web services description language (WSDL) due to its relative simplicity.

Given the potential pervasive nature of IoT objects and IoT applications (e.g., grid control, home control, traffic control, and medical monitoring), security and privacy in communications and services become absolutely critical. Security needs to be intrinsically included in protocol development, and not just be a catch-up afterthought. The plethora of heterogeneous devices now connected to the Internet, from traditional PCs and laptops, to smartphones and Bluetooth-enabled devices, to name just a few, aggravates the risk. Strong authentication, encryption while transmitting, and also encryptions for data at rest is ideal; however, the computational requirements for encryption can be significant. Furthermore, at the central/authenticating site, rapid authentication support is desirable; otherwise objects would not be able to authenticate in large-population environments.

In some IoT applications, as discussed in Chapter 3, there is a need to know the precise physical location of objects; thus, the challenge is how to cost-effectively obtain location information; methods that rely on GPS or cellular services may be too expensive for some applications. In some cases, objects move independently; in other cases, the objects move as the one group. Different tracking methods may be required to achieve efficient handling of tracking information. That is, if a group of objects is known to move as an ensemble (say, a myriad of sensors on a cruise ship; or, multiple medical monitors on an individual, as part of a medical body area network (MBAN) with one gateway controller), then one needs only to figure out where one object is, and the rest of the objects is then in the same relative position. Typically, there is a need to maintain ubiquitous and seamless communication while tracking the location of objects.

Capabilities for scalability are important in order to be able to support an IoT environment where there is a large population that is highly distributed. Solutions are necessary in the arena of distributed networking. For example, the IAB's October 2006 Routing and Addressing Workshop (RFC 4984) refocused interest in scalable routing and addressing architectures for the Internet. Among the many issues driving this renewed interest are concerns about the scalability of the routing system. Proposals have been made recently based on the "locator/identifier separation." The basic idea behind the separation is that the Internet architecture combines two functions, routing locators (where one is attached to the network) and identifiers (where one is located), in one number space: the IP address. Proponents of the separation architecture postulate that splitting these functions apart will yield several advantages, including improved scalability for the routing system. The separation aims to decouple locators and identifiers, thus allowing for efficient aggregation of the routing locator space and providing persistent identifiers in the identifier space. The locator/ID separation protocol (LISP) IETF Working Group (WG) has completed the first set of experimental RFCs describing the LISP. LISP requires no changes to

end-systems or to routers that do not directly participate in the LISP deployment. LISP aims for an incrementally deployable protocol. The LISP WG is working on deliverables for the 2012/2013 time frame that include (i) an architecture description, (ii) deployment models, (iii) a description of the impacts of LISP, (iv) LISP security threats and solutions, (v) allocation of end-point identifier (EID) space, (vi) alternate mapping system designs, and (vii) data models for management of LISP. The first three items (architecture, deployment models, and impacts) need to be completed first before other items can be submitted as RFCs (2). Shim6 (RFCs 5533 through 5535) is another example of possible interest. This protocol is a layer 3 shim for providing locator agility with failover capabilities for IPv6 nodes. Hosts that employ Shim6 use multiple IPv6 address prefixes and setup state with peer hosts. This state can later be used to failover to a different set of locators, should the original locators stop working. The Shim6 approach has a number of advantages, such as enabling small sites to be multihomed without requiring a provider-independent IPv6 address prefix for the site. However, the approach has also been criticized, for example, for the operational impacts that the use of multiple prefixes causes; at this time, there is no clear view on how well Shim6 works in practice, and implementation and deployment in select networks is needed to determine its true characteristics (3).

4.2 STRUCTURAL ASPECTS OF THE IoT

Some key structural related desiderata are highlighted in this section; these issues ultimately may determine the extent and/or rapidity of deployment of IoT services and technologies. This list is not exhaustive.

4.2.1 Environment Characteristics

As we have seen at various points in this text, most (but certainly not all) IoT/machine-to-machine (M2M) nodes have noteworthy design constraints, such as but not limited to the following (6):

- Low power (with the requirement that they will run potentially for years on batteries)
- Low cost (total device cost in single-digit dollars)
- Significantly more devices than in a LAN environment
- Severely limited code and RAM space (e.g., generally desirable to fit the required code—MAC, IP, and anything else needed to execute the embedded application—in, for example, 32K of flash memory, using 8-bit microprocessors)
- Unobtrusive but very different user interface for configuration (e.g., using gestures or interactions involving the physical world)
- Requirement for simple wireless communication technology. In particular, the IEEE 802.15.4 standard is very promising for the lower (physical and link) layers

TABLE 4.1 Properties and Requirements of M2M Applications

	ITS	e-Health	Surveillance	Smart Meters
Mobility	Vehicular	Pedestrian/ vehicular	None	None
Message size	Medium	Medium?	Large	Small (few kB)
Traffic pattern	Regular/ irregular	Regular/ irregular	Regular	Regular
Device density	High	Medium	Low	Very high (up to 10,000 per cell)
Latency requirements	Very high (few milliseconds)	Medium (seconds)	Medium (<200 ms)	Low (up to hours)
Power efficiency requirements	Low	High (battery power devices)	Low	High (battery-powered meters)
Reliability	High	High	Medium	High
Security requirements	Very high	Very high	Medium	High

Courtesy: A. Maeder, NEC Laboratories Europe.

4.2.2 Traffic Characteristics

The characteristics of IoT/M2M communication is different from other types of networks or applications. For example, cellular mobile networks are designed for human communication and communication is connection centric; it entails interactive communication between humans (voice, video), or data communication involving humans (web browsing, file downloads, and so on). It follows that cellular mobile networks are optimized for traffic characteristics of human-based communication and applications. Specifically, communication takes place with a certain length (sessions) and data volume; furthermore, communication takes place with a certain interaction frequency and patterns (talk-listen, download-reading, and so on) (7). On the other hand, in M2M the expectation is that there are many devices, there will be long idle intervals, transmission entails small messages, there may be relaxed delay requirements, and device energy efficiency is paramount. Table 4.1 depicts some key properties and requirements of M2M applications.

4.2.3 Scalability

While some applications (e.g., smart grid, home automation, and so on) may start out covering a small geographic area or a small community of users, as noted above, there invariably will be a desire over time for the service to expand, in order to make such service more cost-effective on a per-unit basis, or to have sufficient critical mass for developers to be motivated to invest resources to add capabilities to the service. When contemplating expansion, one wants to be able to build on previously deployed technology (systems, protocols), without having to scrap the system and start from scratch. Also, the efficiency of a larger system should be better than the efficiency of a smaller system. This is what is meant by scalability. The goal is to make sure that

capabilities such as addressing, communication, and service discovery, among others, are delivered efficiently in both small and large scale. There is a need for enough name space to support increasing populations of devices and new applications. In particular, note that IPv6 is an ideal component (but not the only one) to be employed to support scalability, both for a given application as it reaches more users and for use for a wide class of applications spanning many fields (as described in Chapter 3).

4.2.4 Interoperability

Because of the plethora of applications, technology suppliers, and stakeholders, it is desirable to develop and/or re-use a core set of common standards. To the degree possible, existing standards may prove advantageous to a rapid and cost-effective deployment of the technology. Product and service interoperability is of interest.

4.2.5 Security and Privacy

Unfortunately, security is chronically an after-thought when it comes to protocol development: almost invariably a protocol spec will have many pages of data format and operation procedures and only a short paragraph or two on security considerations. When IoT relates to electric power distribution, goods distribution, transport and traffic management, e-health, and other key applications, as noted earlier, it is critical to maintain system-wide confidentiality, identity integrity, and trustworthiness.

4.2.6 Open Architecture

The goal is to support a wide range of applications using a common infrastructure, preferably based on a service-oriented architecture (SOA) over an open service platform, and utilizing overly networks (these being logical networks defined on top of a physical infrastructure). In an SOA environment, objects expose their functionalities using a protocol such as SOAP or REST application programming interface (API). These devices may provide their functionality as a WS that can in turn be used by other entities (other devices or other business applications).

4.3 KEY IoT TECHNOLOGIES

There are a number of key supportive technologies that are needed for wide-scale deployment of IoT applications. This list is not exhaustive.

4.3.1 Device Intelligence

A key consideration relates to on-board intelligence. In order for the IoT to become a reality, the objects should be able to intelligently sense and interact with the environment, possibly store some passive or acquired data, and communicate with the world around them. Object-to-gateway device communication, or even direct object-to-object communication, is desirable. These intelligent capabilities are necessary to support the ubiquitous networking to provide seamlessly interconnection between

humans and objects. Some have called this mode of communication *Any Services, Any Time, Any Where, Any Devices, and Any Networks* (also known as "5-Any") (1). Pervasive computing (also known as ubiquitous computing) deals with the embedded ability to support logical processing as well with the ability to be in continuous range of a wireless gateway peer.

4.3.2 Communication Capabilities

As just noted, it is highly desirable for objects to support ubiquitous end-to-end communications; hence, another technological consideration relates to communication mechanisms. To achieve ubiquitous connectivity human-to-object and object-to-object communications, networking capabilities will need to be implemented in the objects ("things"). In particular, IP is considered to be key capability for IoT objects; furthermore, the entire TCP/IP Internet Suite is generally desirable. Self-configuring capabilities, especially how an IoT device can establish its connectivity automatically without human intervention, are also of interest. IPv6 auto-configuration and multihoming features are useful in this context, particularly the scope-based IPv6 addressing features.

While we have discussed objects that have sophisticated capabilities (IP support, IPv6 support, Web server capabilities, and so on) in the past few paragraphs, some applications, especially those using simple sensors and/or where is there a very large number of dispersed sensors and/or where there is limited remote energizing power, may have a need to support leaner protocols both at the network layer (e.g., route and/or topology management) and at the transport layer (e.g., using UDP). This may entail some extensions of existing networking protocols to achieve a level of simplicity and minimize power consumption. For constrained objects that do not have high levels of energizing power, memory, and/or computing, lightweight protocols that minimize energy consumption is a desiderata; however, one needs to keep in mind that these protocols may not have enough capabilities to support advanced applications. It should be noted that some existing applications may not even support the IP protocol (even IPv4) and the IP addressing scheme. Hence, there is a need to support heterogeneous (IP and non-IP) networking interfaces, at least in the short term. There may be a need for proxy gateways; such gateways would support multiple interfaces that have evolved from different heterogeneous networks. Interoperability among heterogeneous interfaces can facilitate commercial deployment.

4.3.3 Mobility Support

Yet another consideration relates to tracking and mobility support of mobile object (1). Mobility-enabled architectures and protocols are required. Some objects move independently, while others will move as one of group. Therefore, according to the moving feature, different tracking methods are required. It is important to provide ubiquitous and seamless communication among objects while tracking the location of objects. Mobile IPv6 (MIPv6) offers several capabilities that can address this requirement.

4.3.4 Device Power

A key consideration relates to the powering of the "thing," especially for mobile devices or for devices that otherwise would not have intrinsic power. M2M/IoT applications are almost invariably constrained by the following factors: devices have ultra-low-power capabilities, devices must be of low cost, and devices generally must have small physical size and be light. Specifically, efficient communication mechanisms are needed. A number of devices operate with a small battery, while other devices use a self-energizing energy source, for example a small solar cell array. Yet other devises are passive (e.g., passive RFID) and, thus, need to derive energy indirectly from the environment, such as an intercepting electric/magnetic field. The power requirement is driven by the need to operate for extended periods of time from small batteries or from energy-scavenger mechanisms. In general, wireless technologies require significant amounts of power; hence, the need for low energy (LE) wireless technologies, as discussed in Chapter 6. Batteries are critical to all sorts of products including laptops, pads, smartphones, and IoT objects. The so-called "coin batteries," also known as "button batteries," are typical in many IoT applications.

In recent years, battery technology has seen a doubling in performance approximately every 10 (some say 15) years. Unfortunately, battery technology does not follow Moore's Law which observes empirically that computer chips double in performance and drop their price 50% every 18 to 24 months.

Batteries convert chemical energy released in particular chemical reactions into electrical energy. Batteries have a positive and a negative electrode (the cathode and the anode), separated by an electrolyte. When the electrodes are connected to a closed circuit, a series of chemical reactions occurs such that at one end charged particles (ions) from the electrolyte flow to the anode, react, and free up electrons; at the other end, reactions at the cathode attract free electrons. Thus, electrons at the anode move to the cathode and the flow of electrons through the electric circuit creates an electric current—the electrolyte also prevents the electrons from taking the shortest direct path, instead forcing them through the attached circuit. In rechargeable batteries, the reactions are reversible, with the ions and electrons flowing back in the opposite direction during charging. Batteries can be classified into primary and secondary systems (8). Primary batteries are disposable batteries, that is, batteries that cannot be recharged, and their conversion of chemical energy into electrical energy is irreversible (the chemicals are consumed while the battery discharges). Secondary batteries can be recharged, and the electrode material is reconstituted using an electric charge, so that discharge process can be repeated a number of times during the lifecycle of the battery.

The most common primary systems are alkaline, lithium, and metal/air batteries. Among secondary batteries, lead acid, nickel/cadmium (NiCd), nickel/metal hybrid (NiMH), and lithium-ion (Liion)/lithium-polymer (Li-polymer) batteries dominate the market, but efforts are constantly being made to find new systems that can match or exceed the performance of existing systems, improve their safety, and reduce their cost.

Rechargeable Liion batteries have an anode comprising carbon (e.g., graphite), a metal oxide cathode, and an electrolyte containing lithium salt. It is relatively easy to peel ions from lithium metal. The widespread deployment of this battery technology is due to the fact that the resulting batteries are lightweight, have a high energy density, hold their charge better than other batteries, and they do not suffer from the "memory effect," where batteries hold less and less charge over time if they are not drained and then recharged completely. The technology became popular in the early 1990s, replacing the nickel cadmium predecessors. In recent years, manufacturers have improved battery performance by applying enhanced engineering, optimizing the structures, and/or adding new materials inside the battery to make them more efficient. While battery technology is evolving, Liion batteries will continue to be important for the foreseeable future.

Materials such as silicon and others are being studied as possible replacement of the graphite anodes in Liion batteries. Silicon is of interest because it is inexpensive, it is abundant, and, by weight, it can store 10 times more lithium ions than graphite; this implies that it could theoretically allow a 10-fold increase in performance. However, to be useful, researchers must overcome a problem: while graphite anodes hold their shape when they soak up lithium ions, silicon swells, causing silicon particles to become separated, quickly reducing the performance of the battery. There is work underway to address this challenge, for example, by developing rubbery conductive binders that stick to the silicon particles within the anode, stretching and shrinking as the battery is charged and discharged. Others are looking to develop Liion batteries with anodes containing silicon nanowires. We are also starting to see a research on totally different chemistries. One example is lithium–air batteries with anodes made of lightweight porous carbon. Oxygen from the air enters the porous carbon and reacts with the lithium ions in the electrolyte and electrons in the external circuit to form solid lithium oxide. Recharging causes the lithium compound to decompose, releasing the lithium ions and releasing the oxygen; calculations of the amount of energy involved in the chemical reactions involved suggest it could produce batteries that last three to five times as long as existing Liion batteries. Other researchers are also investigating lightweight lithium–sulfur packs, which have a life span of three times that of current Liion batteries. Research work is proceeding on these and other technologies (9). However, these advances represent only incremental improvements in performance. New materials may be needed to make a quantum leap forward. MIT's Materials Project has already identified four new materials with the potential to be used in batteries. For example, the use of magnesium metal anode could result into a three-fold energy density compared with the best Liion batteries; furthermore these magnesium-ion batteries hold the majority of their charge over 3000 charge cycles. The technology, however, is not ready for widespread commercialization.

Another approach is the fuel cell, which is a proven technology since they were used to power electronics of the Gemini and Apollo space missions. Fuel cells convert chemical energy into electricity by converting the chemical energy from a fuel (e.g., alcohol) into electricity through a chemical reaction with oxygen. Fuel cells have a high energy density: hydrogen contains nearly 150 times the energy of an equivalent weight of lithium. However, to be practical, they need to be small and have an easily

rechargeable reservoir for fuel. Microelectromechanical system (MEMS) technology is being investigated for this purpose. MEMSs are miniaturized mechanical devices that are already used in solar cells and flat-screen TVs. Currently, the technology is expensive because precious metals such as platinum and palladium are used; companies such as, but not limited to, NEC, Toshiba, and Apple are continuing substantive research in the field.

Some evolving technologies use small solar panels embedded in the screen of a smartphone or object; other systems may use kinetic devices that translate movement of objects into an electric current. Solar cells are an example of an energy harvester, but they are for low efficiency when converting ambient light into useful electrical energy. A 3 cm^2 solar cell (dimensions similar to the common CR2032 coin cell) yields only 12 μW.

There are a number of factors that must be considered in selecting the most suitable battery for a particular application; key considerations include (8):

- Operating voltage level
- Load current and profile
- Duty cycle—continuous or intermittent
- Service life
- Physical requirement
  - Size
  - Shape
  - Weight
- Environmental conditions
  - Temperature
  - Pressure
  - Humidity
  - Vibration
  - Shock
  - Pressure
- Safety and reliability
- Shelf life
- Maintenance and replacement
- Environmental impact and recycling capability
- Cost

4.3.5 Sensor Technology

A sensor network is an infrastructure comprising sensing (measuring), computing, and communication elements that gives the administrator the ability to instrument, observe, and react to events and phenomena in a specified environment. The administrator typically is some civil, government, commercial, or industrial entity.

Network(ed) sensor systems support a plethora of applications, not the least being Homeland Security. Typical applications include, but are not limited to, data collection, monitoring, surveillance, and medical telemetry. Sensors facilitate the instrumenting and controlling of factories, offices, homes, vehicles, cities, and the ambiance, especially as commercial off-the-shelf technology becomes available. With sensor network technology, specifically, with embedded networked sensing, ships, aircrafts, and buildings can "self-detect" structural faults (e.g., fatigue-induced cracks). Places of public assembly can be instrumented to detect airborne agents such as toxins and to trace the source of the contamination, should any be present (this also can be done for ground/underground situations). Earthquake-oriented sensors in buildings can locate potential survivors and can help assess structural damage; tsunami-alerting sensors can certainly prove useful for nations with extensive coastlines. Sensors also find extensive applicability in battlefield for reconnaissance and surveillance. In addition to sensing, one is often also interested in control and activation (13).

There are four basic components in a sensor network: (i) an assembly of distributed or localized sensors; (ii) an interconnecting network (usually, but not always, wireless-based); (iii) a central point of information clustering; and (iv) a set of computing resources at the central point (or beyond) to handle data correlation, event-trending, querying, and data mining. Because the interconnecting network is generally wireless, these systems are known as wireless sensor networks (WSNs).

In this context, the sensing and computation nodes are considered part of the sensor network; in fact, some of the computing may be done in the network itself. Because of the potentially large quantity of data collected, algorithmic methods for data management play an important role in sensor networks. The computation and communication infrastructure associated with sensor networks is often specific to this environment and rooted in the device- and application-based nature of these networks. For example, unlike most other settings, in-network processing is desirable in sensor networks; furthermore, node power (and/or battery life) is a key design consideration.

Sensors, the things or objects in this discussion, are active devices that measure some variable of the natural or man-made environment (e.g., a building, an assembly line, an industrial assemblage). Sensors in a WSN have a variety of purposes, functions, and capabilities. The radar networks used in air traffic control, the national electrical power grid, and the nation-wide weather stations deployed over a regular topographic mesh are all examples of early-deployment sensor networks. All of these systems, however, use specialized computers and communication protocols and are very expensive. Less expensive WSNs are now being planned for novel applications in physical security, healthcare, and commerce. The technology for sensing and control includes electric and magnetic field sensors; radio-wave frequency sensors; optical, electro-optic, and infrared sensors; radars; lasers; location/navigation sensors; seismic and pressure-wave sensors; environmental parameter sensors (e.g., wind, humidity, heat, and so on); and biochemical Homeland Security-oriented sensors.

Sensors can be described as "smart" inexpensive devices equipped with multiple on-board sensing elements: they are low cost, low power, untethered multifunctional nodes that are logically homed to a central sink node. Sensors are typically internetworked via a series of multihop short-distance low power wireless links

(particularly within a defined "sensor field"); they typically utilize the Internet or some other network for long-haul delivery of information to a point (or points) of final data aggregation and analysis. In general, within the "sensor field," WSNs employ contention-oriented random access channel sharing/transmission techniques that are now incorporated in the IEEE 802 family of standards; indeed, these techniques were developed in the late 1960s and 1970s expressly for wireless (not cabled) environments, and for large sets of dispersed nodes with limited channel-management intelligence. However, other channel management techniques are also available. Sensors are typically deployed in a high density manner and in large quantities: a WSN consists of densely distributed nodes that support sensing, signal processing, embedded computing, and connectivity; sensors are logically linked by self-organizing means (sensors that are deployed in short-hop point-to-point master-slave pair arrangements are also of interest). Wireless sensors typically transmit information to collecting (monitoring) stations that aggregate some or all of the information. WSNs have unique characteristics, such as, but not limited to, power constraints/limited battery life for the wireless sensors, redundant data, low duty cycle, and many-to-one flows. Consequently, new design methodologies are needed across a set of disciplines, including, but not limited to, information transport, network and operational management, confidentiality, integrity, availability, and in-network/local processing. In some cases, it is challenging to collect (extract) data from wireless nodes (WNs) because connectivity to/from the WNs may be intermittent due to a low battery status (e.g., if these are dependent on sun light to recharge), or other wireless sensor malfunction. Furthermore, a lightweight protocol stack is desired. Often, a very large number of client units (say 64K or more) need to be supported by the system and by the addressing apparatus.

Sensors span several orders of magnitude in physical size; they (or, at least some of their components) range from nanoscopic-scale devices to mesoscopic-scale devices at one end; and, from microscopic-scale devices to macroscopic-scale devices at the other end. Nanoscopic (also known as nanoscale) refers to objects or devices in the order of 1–100 nm in diameter; mesoscopic scale refers to objects between 100 and 10,000 nm in diameter; the microscopic scale ranges from 10 to 1000 microns; and the macroscopic scale is at the millimeter-to-meter range. At the low end of the scale, one finds, among others, biological sensors, small passive microsensors (such as "smart dust"), and "lab-on-a-chip" assemblies. At the other end of the scale, one finds platforms such as, but not limited to, identity tags, toll collection devices, controllable weather data collection sensors, bioterrorism sensors, radars, and undersea submarine traffic sensors based on sonars.[2] Some refer to the latest generation of sensors, especially the miniaturized ones that are directly embedded in some physical infrastructure, as "microsensors." Microsensors with on-board processing and wireless interfaces can be utilized to study and monitor a variety of phenomena and environments at close proximity.

Sensors may be passive and/or be self-powered; further along in the power-consumption chain, some sensors may require relatively low power from a battery

[2]While satellites can be used to support sensing, this book does not explicitly include them in the technical discussion.

or line feed. At the high end of the power-consumption chain, some sensors may require very high power feeds (e.g., for radars). Chemical-, physical-, acoustic-, and image-based sensors can be utilized to study ecosystems (e.g., in support of global parameters such as temperature, microorganism populations, and so on). Near-term commercial applications include, but are not limited to, industrial/building WSNs, appliance control (lighting and heating, ventilation, and air conditioning (HVAC)), automotive sensors and actuators, home automation and networking, automatic meter reading/load management (LM), consumer electronics/entertainment, and asset management. Commercial market segments include the following:

- Industrial monitoring and control
- Commercial building and control
- Process control
- Home automation
- Wireless automated meter reading (AMR)/ LM
- Metropolitan operations (traffic, automatic tolls, fire, and so on)
- Homeland Security applications: chemical, biological, radiological, and nuclear wireless sensors
- Military sensors
- Environmental (land, air, sea)/agricultural wireless sensors

Suppliers and products tend to cluster according to these categories.

Implementations of WSNs have to address a set of technical challenges; however, the move toward standardization will, in due course, minimize a number of these challenges by addressing the issues once and then resulting in off-the-shelf chipsets and components. One of the challenges of WSNs is the need for extended temporal operation of the sensing node in spite of a (typically) limited power supply (and/or battery life). In particular, the architecture of the radio, including the use of low power circuitry, must be properly selected. In practical terms, this implies low power consumption for transmission over low bandwidth channels and low power-consumption logic to pre-process and/or compress data. Energy-efficient wireless communications systems are being sought and are typical of WSNs. Low power consumption is a key factor in ensuring long operating horizons for non-power-fed systems (some systems can indeed be power-fed and/or relay on other power sources). Power efficiency in WSNs is generally accomplished in three ways:

(i) Low duty cycle operation
(ii) Local/in-network processing to reduce data volume (and, hence, transmission time)
(iii) Multihop networking (this reduces the requirement for long-range transmission since signal path loss is an inverse power with range/distance (e.g., 4)—each node in the sensor network can act as a repeater, thereby reducing the link range coverage required, and, in turn, the transmission power

Conventional wireless networks are generally designed with link ranges of the order of tens, hundreds, or thousands of miles. The reduced link range and the compressed data payload in WSNs result in characteristic link budgets that differ from conventional systems.

4.3.6 RFID Technology

RFIDs are electronic devices associated with objects ("things") that transmit their identity (usually a serial number) via radio links. The RFID space is large and well documented. Our discussion here is very limited by choice; the reader requiring more details is encouraged to seek out the literature on the topic.

RFID tags are devices that typically have a read-only chip that stores a unique number but has no processing capability. RFID tags have broad applications, including the rapid collection of data in commercial environments. For example, RFID and bar coding are nearly ubiquitous in the inventory process, providing both accuracy and speed of data collection. These technologies facilitate the global supply chain and impact all subsystems within that overall process, including material requirement planning (MRP), just in time (JIT), electronic data interchange (EDI), and electronic commerce (EC). RFIDs are also used in industrial environments, such as but not limited to dirty, wet, or harsh environments. The technology can also be used for identification of people or assets. Figure 4.1 depicts two illustrative examples of RFIDs. Figure 4.2 depicts the basic operation of an RFID system.

Contactless smart cards (SCs) are more sophisticated than RFID tags, being that they contain a microprocessor that enables (i) on-board computing, (ii) two-way

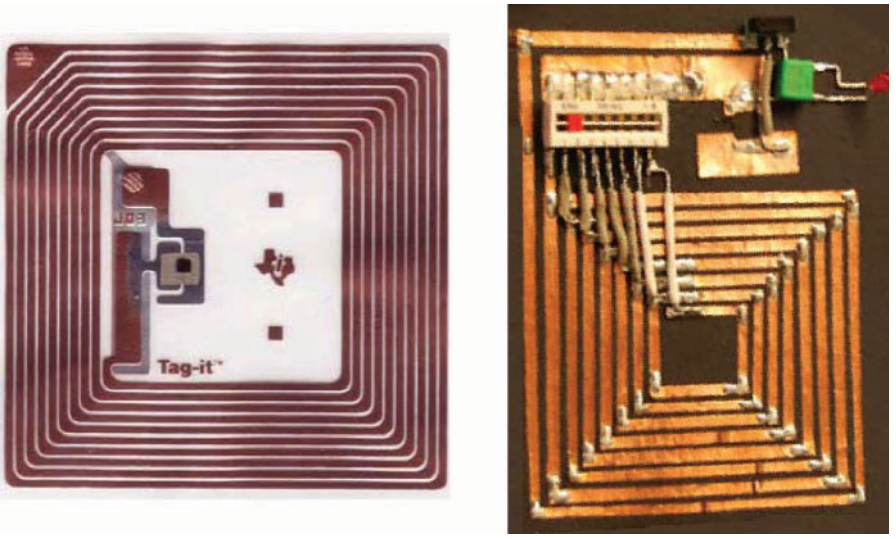


FIGURE 4.1 Illustrative examples of RFIDs.

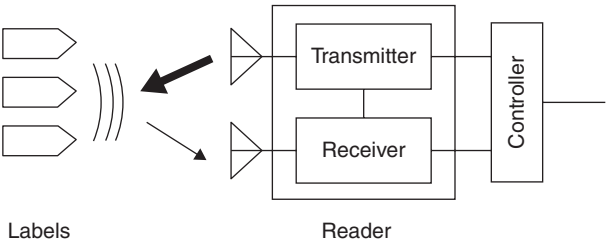


FIGURE 4.2 RFID reader operation.

communication including encryption, and (iii) storage of predefined and newly acquired information. Because of their more restricted capabilities, RFID tags are typically less expensive than SCs. When an RFID tag or contactless SC passes within a defined range, a reader generates electromagnetic waves; the tag's integrated antenna receives the signal and activates the chip in the tag/SC, and a wireless communications channel is set up between the reader and the tag enabling the transfer of pertinent data. Figure 4.3 provides a comparison between SCs and RFID tags.

RFID examples applicable to IoT include but are not limited to the following:

- Warehouse retailer automotive
- Grocery chain transportation
- Distribution center asset management
- Manufacturing
- Inventory management
- Warehousing and distribution
- Shop floor (production)
- Document tracking and asset management
- Industrial applications (e.g., time and attendance, shipping document tracking, receiving fixed assets)
- Retail applications

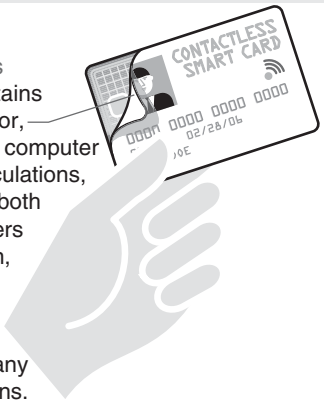
There are a number of standards for RFIDs. Some of the key ones include the following:

- The ISO 14443 standard describes components operating at 13.56 MHz frequency that embed a CPU; power consumption is about 10mW; data throughput is about 100 Kbps and the maximum working distance (from the reader) is around 10 cm.
- The ISO 15693 standard also describes components operating at 13.56 MHz frequency, but it enables working distances as high as 1 m, with a data throughput of a few Kbps.

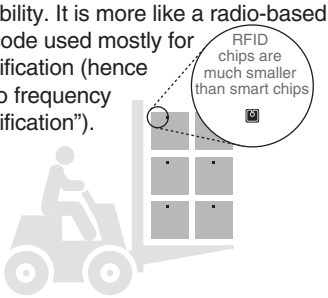
Overview: what happens in RF (radio frequency) communication

1. When a contactless smart card or an RFID tag passes within range, a reader sends out radio frequency electromagnetic waves.
2. The antenna, tuned to receive these waves, wakes up the chip in the smart card or tag.
3. A wireless communications channel is set up between the reader and the smart card or tag.

The contactless smart card contains a microprocessor, a small but real computer that makes calculations, communicates both ways, remembers new information, and actively uses these capabilities for security and many other applications.



RFID tags are devices that typically have a read-only chip that stores a unique number but has no processing capability. It is more like a radio-based bar code used mostly for identification (hence "radio frequency identification").



Characteristics of a contactless card

- Strong security capacities:
  - mutual authentication before providing access to information
  - access can be further protected via PIN or biometric
  - encryption to protect data on card during exchange
  - hardware and software protection to combat attacks or counterfeiting
- Hundreds of security features mean an individual's personal ID, financial details, payment transactions, transit fares or physical access privileges can be safely stored, managed, and exchanged
- Read and write memory capacity of 512 bytes and up, with very large memory storage possible
- Short-distance data exchange, typically two inches

Characteristics of an RFID tag

- Minimal security:
  - one-way authentication; card cannot protect itself
  - insufficient storage for biometrics
  - no on-chip calculations of new information
  - relies on static keys
- Single function; used to help machines identify objects to increase efficiency. Example: inventory control
- Small memory (92 bytes); often read-only
- Larger distance data exchange, typically several yards

Because of their more restricted capabilities, RFID tags are generally cheaper.

FIGURE 4.3 Comparison between contactless SCs and RFID tags. *Source:* Gemalto (used with Permission).

- The ISO 18000 standard defines parameters for air interface communications associated with frequency such as 135 KHz, 13.56 MHz, 2.45 GHz, 5.8 GHz, 860–960 MHz, and 433 MHz. The ISO 18000–6 standard uses the 860–960 MHz range and is the basis for the Class-1 Generation-2 UHF RFID, introduced by the EPCglobal Consortium.

As a side note, EPCglobal Inc. was created as a joint venture between GS1 (formerly EAN International) and GS1 US (formerly the Uniform Code Council, Inc.)—the same organizations entrusted to drive adoption of the barcode—to develop standards and to create a "visible" global supply chain. EPCglobal is a neutral, not-for-profit standards organization consisting of manufacturers, technology solution providers, and retailers. Many industries participate in the EPCglobal standards development process such as aerospace, apparel, chemical, consumer electronics, consumer goods, healthcare and life sciences, and transportation and logistics.

Typically, EPC codes used for active RFIDs or IP addresses are transmitted in clear form; however, some new protocols are now emerging that can provide strong privacy for the IoT. The host identity protocol (HIP) is one example; with this protocol, active RFIDs do not expose their identity in clear text, but protect the identity value (e.g., an EPC) using cryptographic procedures (10).

Table 4.2 based on material from Reference 11 provides a very basic listing of RFID concepts. An RFID system is logically comprising several layers, as follows: the tag layer, the air interface (also called media interface) layer, and the reader layer; additionally there are network, middleware, and application aspects. Some of the key aspects of the basic layers are as follows:

- Tag (device) layer: Architecture and EPCglobal Gen2 tag finite state machine
- Media interface layer: Frequency bands, antennas, read range, modulation, encoding, data rates
- Reader layer: Architecture, antenna configurations, Gen2 sessions, Gen2

The following is a list of key specifications supporting basic RFID operations:

- EPCglobal™: *EPC™ Tag Data Standards*
- EPCglobal™ (2004): *FMCG RFID Physical Requirements Document*
- EPCglobal™ (2004): *Class-1 Generation-2 UHF RFID Implementation Reference*
- EPCglobal™ (2005): *Radio-Frequency Identity Protocols, Class-1 Generation-2 UHF RFID, Protocol for Communications at 860 MHz–960 MHz*
- European Telecommunications Standards Institute (ETSI), EN 302 208: *Electromagnetic Compatibility and Radio Spectrum Matters (ERM)—Radio-Frequency Identification Equipment Operating in the Band 865 MHz to 868 MHz with Power Levels up to 2 W, Part 1 – Technical Characteristics and Test Methods*

TABLE 4.2 Basic RFID Concepts

Concept	Definition
Air interface	The complete communication link between an interrogator and a tag including the physical layer, collision arbitration algorithm, command and response structure, and data-coding methodology
Continuous wave (CW)	Typically a sinusoid at a given frequency, but more generally any interrogator waveform suitable for powering a passive tag without amplitude and/or phase modulation of sufficient magnitude to be interpreted by a tag as transmitted data
Cover-coding	A method by which an interrogator obscures information that it is transmitting to a tag. To cover-code data or a password, an interrogator first requests a random number from the tag. The interrogator then performs a bit-wise EXOR of the data or password with this random number and transmits the cover-coded (also called ciphertext) string to the tag. The tag uncovers the data or password by performing a bit-wise EXOR of the received cover-coded string with the original random number
EPC	A unique identifier for a physical object, unit load, location, or other identifiable entity playing a role in business operations. EPCs are assigned following rules designed to ensure uniqueness despite decentralized administration of code space, and to accommodate legacy coding schemes in common use. EPCs have multiple representations, including binary forms suitable for use on RFID tags, and text forms suitable for data exchange among enterprise information systems
EPCglobal architecture framework	A collection of interrelated standards ("EPCglobal Standards"), together with services operated by EPCglobal, its delegates, and others ("EPC Network Services"), all in service of a common goal of enhancing business flows and computer applications through the use of EPCs
Interrogator	A device that modulates/transmits and receives/demodulates a sufficient set of the electrical signals defined in the signaling layer to communicate with conformant tags, while conforming to all local radio regulations. A typical interrogator is a passive-backscatter, interrogator-talks-first (ITF), RFID system operating in the 860–960 MHz frequency range. An interrogator transmits information to a Tag by modulating an RF signal in the 860 MHz–960 MHz frequency range. The tag receives both information and operating energy from this RF signal. Tags are passive, meaning that they receive all of their operating energy from the interrogator's RF waveform. An interrogator receives information from a tag by transmitting a continuous-wave (CW) RF signal to the tag; the Tag responds by modulating the reflection coefficient of its antenna, thereby backscattering an information signal to the interrogator. The system is ITF, meaning that a tag modulates its antenna reflection coefficient with an information signal only after being directed to do so by an interrogator. Interrogators and tags are not required to talk simultaneously; rather, communications are half-duplex, meaning that interrogators talk and tags listen, or vice versa

(*continued*)

TABLE 4.2 (*Continued*)

Concept	Definition
Operating environment	A region within which an interrogator's RF transmissions are attenuated by less than 90dB. In free space, the operating environment is a sphere whose radius is approximately 1000 m, with the interrogator located at the center. In a building or other enclosure, the size and shape of the operating environment depends on factors such as the material properties and shape of the building and may be less than 1000 m in certain directions and greater than 1000 m in other directions
Operating procedure	Collectively, the set of functions and commands used by an interrogator to identify and modify tags (also known as the *tag-identification layer*)
Passive tag (or passive label)	A tag (or label) whose transceiver is powered by the RF field
Physical layer	The data coding and modulation waveforms used in interrogator-to-tag and tag-to-interrogator signaling
Singulation	Identifying an individual tag in a multiple-tag environment
Slotted random anticollision	An anticollision algorithm where tags load a random (or pseudo-random) number into a slot counter, decrement this slot counter based on interrogator commands, and reply to the interrogator when their slot counter reaches zero
Tag air interface	As defined in ISO 19762–3, a conductor-free medium, usually air, between a transponder and a reader/interrogator through which data communication is achieved by means of a modulated inductive or propagated electromagnetic field
Tag-identification layer	Collectively, the set of functions and commands used by an interrogator to identify and modify tags (also known as the *operating procedure*)

- European Telecommunications Standards Institute (ETSI), EN 302 208: *Electromagnetic Compatibility and Radio Spectrum Matters (ERM)—Radio-Frequency Identification Equipment Operating in the Band 865 MHz to 868 MHz with Power Levels up to 2 W, Part 2—Harmonized EN under article 3.2 of the R&TTE Directive*
- ISO/IEC Directives, Part 2: *Rules for the Structure and Drafting of International Standards*
- ISO/IEC 3309: *Information Technology—Telecommunications and Information Exchange Between Systems—High Level Data Link Control (HDLC) Procedures—Frame Structure*
- ISO/IEC 15961: *Information Technology, Automatic Identification and Data Capture—Radio Frequency Identification (RFID) for Item Management—Data Protocol: Application Interface*
- ISO/IEC 15962: *Information Technology, Automatic Identification and Data Capture Techniques—Radio Frequency Identification (RFID) for Item*

Management—Data Protocol: Data Encoding Rules and Logical Memory Functions

- ISO/IEC 15963: *Information Technology—Radiofrequency Identification for Item Management—Unique Identification for RF Tags*
- ISO/IEC 18000–1: *Information Technology—Radio Frequency Identification for Item Management—Part 1: Reference Architecture and Definition of Parameters to be Standardized*
- ISO/IEC 18000–6: *Information Technology Automatic Identification and Data Capture Techniques—Radio Frequency Identification for Item Management Air Interface—Part 6: Parameters for Air Interface Communications at 860–960 MHz*
- ISO/IEC 19762: *Information Technology AIDC Techniques—Harmonized Vocabulary—Part 3: Radio-Frequency Identification (RFID)*
- U.S. Code of Federal Regulations (CFR), Title 47, Chapter I, Part 15: *Radio-Frequency Devices, U.S. Federal Communications Commission*

Figure 4.4 depicts the set of relevant standards in the EPCglobal environment. In particular, the EPCglobal organization has defined an EPCglobal Architecture Framework in the document *The EPCglobal Architecture Framework, EPCglobal Final Version 1.4, December 2010*. The EPCglobal Architecture Framework is a

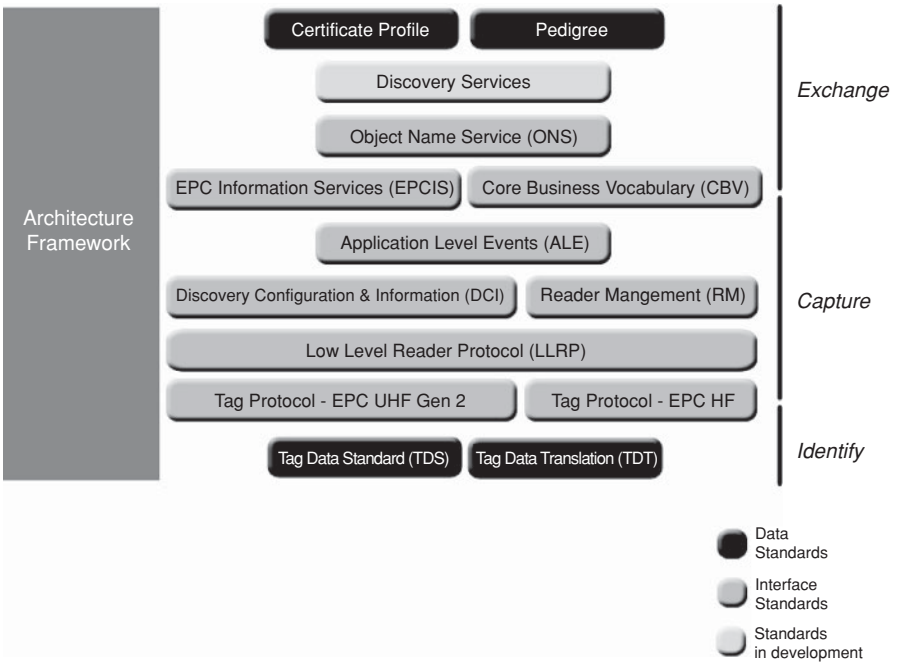


FIGURE 4.4 Standards that comprise the EPCglobal environment.

collection of interrelated standards ("EPCglobal Standards"), together with services operated by EPCglobal, its delegates, and others ("EPC Network Services"), all in service of a common goal of enhancing business flows and computer applications through the use of EPCs. It describes the collection of interrelated standards for hardware, software, and data interfaces, together with core services that are operated by EPCglobal and its delegates, all in service of a common goal of enhancing the supply chain through the use of EPCs. The architecture define core services that are operated by EPCglobal and its delegates, showing how the different components fit together to form a cohesive whole. It discusses (12):

- Individual hardware, software, and data interfaces are defined normatively by EPCglobal standards, or by standards produced by other standards bodies. EPCglobal standards are developed by the EPCglobal Community through the EPCglobal Standard Development Process (SDP). EPCglobal standards are normative, and implementations are subject to conformance and certification requirements. An example of an interface is the UHF Class-1 Gen-2 tag air interface, which specifies a radio-frequency communications protocol by which an RFID tag and an RFID reader device may interact. This interface is defined normatively by the UHF Class-1 Gen-2 tag air interface standard.
- The design of hardware and software components that implement EPCglobal standards are proprietary to the solution providers and end users that create such components. While EPCglobal standards provide normative guidance as to the behavior of interfaces between components, implementers are free to innovate in the design of components so long as they correctly implement the interface standards. An example of a component is an RFID tag that is the product of a specific tag manufacturer. This tag may comply with the UHF Class-1 Gen-2 tag air interface standard.
- A special case of components that implement EPCglobal standards are shared network services that are operated and deployed by EPCglobal itself (or by other organizations to which EPCglobal delegates responsibility), or by other third parties. These components are referred to as EPC network services and provide services to all end users. An example of an EPC Network Service is the ONS, which provides a logically centralized registry through which an EPC may be associated with information services. The ONS is logically operated by EPCglobal; from a deployment perspective this responsibility is delegated to a contractor of EPCglobal that operates the ONS "root" service, which in turn delegates responsibility for certain lookup operations to services operated by other organizations.

4.3.7 Satellite Technology

Due to its global reach and the ability to support mobility in all geographical environments (including Antarctica), satellite communications can play a critical role in

many broadly distributed M2M applications. This topic deserves more attention and development because it offers interesting commercial possibilities.

REFERENCES

1. Lee GM, Park J, Kong N, Crespi N. The Internet of Things – Concept and Problem Statement. July 2011. Internet Research Task Force, July 11, 2011, draft-lee-iot-problem-statement-02.txt.
2. Manderson T, Halpern JM. Locator/ID Separation Protocol (lisp). IETF Working Group.
3. Huston G, Lindqvist K. Site Multihoming by IPv6 Intermediation (shim6). IETF Working Group, 2010.
4. Lee GM, Choi JK, et al. Naming Architecture for Object to Object Communications. HIP Working Group, Internet Draft, March 8, 2010, draft-lee-object-naming-02.txt.
5. EPCglobal® Organization Web Site, http://www.gs1.org/epcglobal.
6. Mulligan G. IPv6 Over Low power WPAN (6lowpan). Description of Working Group, IETF, 2012, http://datatracker.ietf.org/wg/6lowpan/charter/, http://www.ietf.org/mail-archive/web/6lowpan/.
7. Maeder A. How to Deal with a Thousand Nodes: M2M Communication Over Cellular Networks. IEEE WoWMoM 2012 Panel, San Francisco, California, USA June 25–28, 2012.
8. Tidblad AA. The Future of Battery Technologies – Part I, Intertek White Paper, November 2009, icenter@intertek.com.
9. Fleming N. Smartphone Batteries: When will They Last Longer?. bbc.com online article, February 27, 2012.
10. Urien P, Lee GM, Pujolle G. HIP Support for RFIDs. HIP Research Group, Internet Draft, draft-irtf-hiprg-rfid-03, July 2011.
11. EPCglobal®, EPC™ Radio-Frequency Identity Protocols, Class-1 Generation-2 UHF RFID, Protocol for Communications at 860 MHz–960 MHz, Version 1.0.9, January 2005.
12. EPCglobal®, *The EPCglobal Architecture Framework*, EPCglobal Final Version 1.4, December 2010, Ken Traub Editor.
13. Minoli D, Sohraby K, Zanti T. *Wireless Sensor Networks*, Wiley 2007, New York, NY.