Unit-1

Introduction, Background and Initial Visions:

The term *Internet of Things* first came to attention when the Auto-ID Center launched their initial vision of the EPC network for automatically identifying and tracing the flow of goods in supply-chains, in Chicago in September 2003 (EPC Symposium 2003). Whereas the first mention of 'Internet of Things' appears in an Auto-ID Center paper about the Electronic Product Code by David Brock in 2001 (Brock 2001), increasing numbers of researchers and practitioners have followed this vision, as it is documented by books, conferences and symposia having Internet of Things in their titles.
The Internet of Things is a concept in which the virtual world of information technology integrates seamlessly with the real world of things. The real world

Becomes more accessible through computers and networked devices in business as well as everyday scenarios. With access to fine-grained information, management can start to move freely from macro to micro levels and will be able to measure, plan and act accordingly. However, the Internet of Things is more than a business tool for managing business processes more efficiently and more effectively – it will also enable a more convenient way of life.
Since the founders of the Auto-ID Center coined the term 'Internet of Things' (Santucci 2010), it has widely been used by researchers and practitioners to describe the combination of the real world with the virtual world of information technology (Fleisch and Mattern 2005, Bullinger and ten Hompel 2007, Floerkemeier et al. 2008) by means of automatic identification technologies, real-time locating systems, sensors and actuators.
Detection of the physical status of things through sensors, together with collection and processing of detailed data, allows immediate response to changes in the real world. This fully interactive and responsive network yields immense potential for citizens, consumers and business.
RFID is increasingly being deployed in applications across supply chains with readers that are distributed across factories, warehouses, and retail stores. Sensor technology is also being adopted in manufacturing and logistics in order to control processes and the quality of goods. In traditional RFID applications, such as access control and production automation, tags moved in closed-loop processes, and the RFID data was consumed only by a single client system. Accordingly, there was little need for exchange of data across organisational boundaries. In the same way that monolithic business information systems of the past have evolved into highly networked systems

that use the Internet extensively, open-loop RFID applications in networked environments represent a challenge that various stakeholders from industry are facing and partly solving.

Accessing real-time information through Information and Communication Technology (ICT) usage in the 'anytime, anywhere' manner, as suggested by the paradigm of the Internet of Things, calls for open, scalable, secure and standardised infrastructures which do not fully exist today. These have been developed and continue to be developed for example in working groups within the EPCglobal

community in order to gather user requirements and business cases to develop open global technical standards for improved visibility. Similarly, members of the Open Geospatial Consortium (OGC) are building a framework of open standards for exploiting Web-connected sensors and sensor systems of all types, including flood gauges, air pollution monitors, stress gauges on bridges, mobile heart monitors, webcams and satellite-borne earth imaging devices. Today's technology-centric instead of user-centric developments are some of the problems that hinder a broader and faster adoption. The arrival of NFC and RFID technology in the consumer market (e.g., Nabaztag.com, Touchatag.com) together with the availability of mobile Internet (e.g., Apple iPhone, HTC Touch) and scalable information sharing infrastructures (e.g., Twitter.com) opens an enormous space for end-user innovation and user-centric developments. People and things are getting closer. An open and holistic approach of a network of products and people has yet to be developed.

Most existing RFID-installations in production and logistics today can be considered as an Intranet of Things or Extranet of Things. Traditional communication means, such as EDIFACT, are used to communicate with

limited number of preferred partners. These early approaches need to be extended to support open Internet architectures.

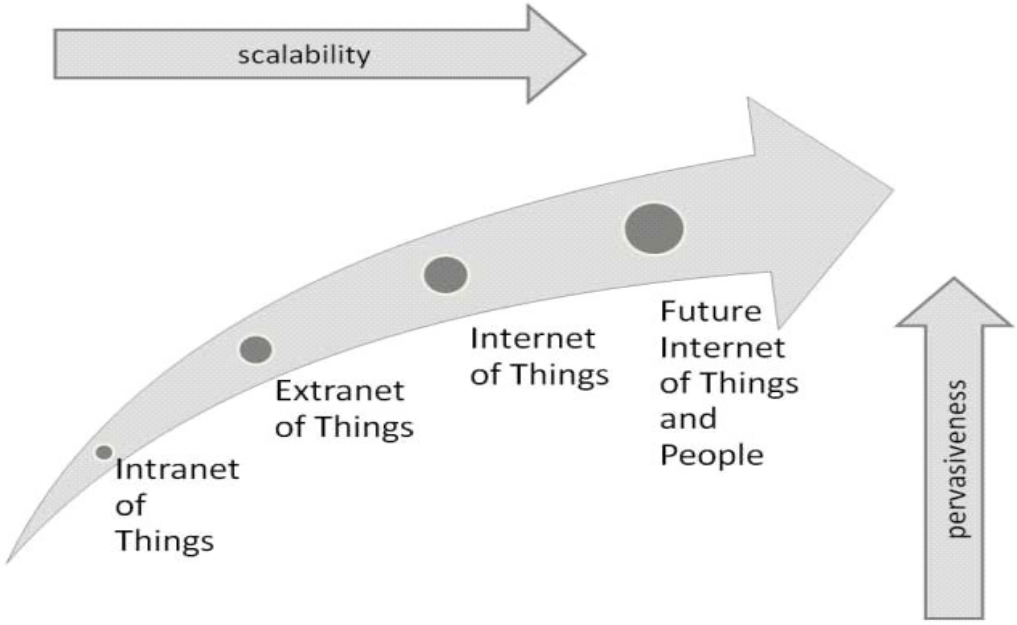


Fig. 1.1 A Phased Approach from the Intranet of Things to a Future Vision on the Internet of Things

Figure 1.1 shows a phased approach from the current Intranet / Extranet of Things to a future Internet of Things and People. While pervasiveness increases

through new applications and wider adoption, the scalability requirements of the Internet of Things have to be met.

Additionally, a solid business case and flexible mechanisms for balancing costs and benefits are missing in many of today's early implementations. The usability needs to be improved by providing flexible but simple devices and services to connect things and people. The Internet of Things can benefit from the latest developments and functionalities commonly referred to as Web 2.0 through provision of new intuitive user-centred and individually configurable and self-adapting smart products and services for the benefit of businesses and society. Whereas the successful examples of Web 2.0, such as Facebook or Twitter, connect people with data, this is achieved by proprietary

Application Programming Interfaces (APIs) that do not provide powerful data-sharing models capable of Business-to-Business (B2B) requirements, such as data management and analysis.

This chapter will focus on providing an overview of the Internet of Things and its future requirements. In section 1.2 we will provide a definition of the Internet of Things. Section 1.3 will provide a broad review of development projects and initiatives, whereas in section 1.4 we will highlight ten key requirements for the future Inconclusion and a further outlook towards future developments.

Definitions and Functional Requirements

The term Internet of Things is not well defined and has been used and misused as a buzzword in scientific research as well as marketing and sales strategies. Until today it remains difficult to come up with a clear definition of the Internet of Things. One definition has recently been formulated in the Strategic Research Agenda of the Cluster of European Research Projects on the Internet of Things (CERP-IoT 2009):

"Internet of Things (IoT) is an integrated part of Future Internet and could be defined as a dynamic global network infrastructure with self configuring capabilities based on standard and interoperable communication protocols where physical and virtual 'things' have identities, physical attributes, and virtual personalities and use intelligent interfaces, and are seamlessly integrated into the information network. In the IoT, 'things' are expected to become active participants in business, information and social processes where they are enabled to interact and communicate among themselves and with the environment by exchanging data and information 'sensed' about the environment, while reacting autonomously to the 'real/physical world' events and influencing it by running processes that trigger actions and create services with or without direct human intervention.

Interfaces in the form of services facilitate interactions with these 'smart things' over the Internet, query and change their state and any information associated with them, taking into account security and privacy issues."

While this definition lists the possible technical components of the Internet of Things, it still has three major shortcomings. Firstly, it lists components that have been mentioned before in relation to other visions such as pervasive or ubiquitous computing and therefore it is difficult to distinguish from these concepts. Secondly, it misses wider consideration of current developments and user- interactions in the

Internet commonly referred to as Web 2.0. Similar to the relationship between the World Wide Web (WWW) and the Internet, the addition of Web 2.0 functionality may be seen as a user-centric extension to the Internet of Things rather than an integral part of it. However, whereas the

development of the Internet began more than thirty years before the realisation of the WWW in the early 1990s, the Internet of Things is already being influenced by Web 2.0 functionality right from the beginning.

Both technology developments have been happening in parallel rather than consecutively. Thirdly, it does not provide a reason why or how the Internet of Things will be a self-sustainable and successful concept for the future. Self-sustainability encompasses viability, including a dynamic global network infrastructure with self-configuring capabilities based on standards and interoperable communication protocols as well as openness for future extensions, ideas, and technologies. Economic success may never have been a part of a definition for the Internet or other technical network infrastructures. Nevertheless, we consider it a valid consideration within a holistic definition approach as economic success and adoption is just as important as technical sustainability in a forward-looking statement.

For the purposes of differentiation it may be best to consider what the Internet of Things is not – or at least not exclusively. A corresponding blog discussion has been started by Tomas Sánchez López (Sánchez López 2010). He considers that the Internet of Things is not only:

- *ubiquitous / pervasive computing*, which does not imply the usage of objects nor does it require a global Internet infrastructure
- the *Internet Protocol (IP)*, as many objects in the Internet of Things will not be able to run an Internet Protocol
- a *communication technology*, as this represents only a partial functional requirement in the Internet of Things similar to the role of communication technology in the Internet
- an *embedded device*, as RFID tags or Wireless Sensor Networks (WSN) may be part of the Internet of Things, but stand-alone they miss the back-end information infrastructures and in the case of WSN the standards to relate to 'things'
- the *application*, just as Google or Facebook could not be used in the early 90's to describe the possibilities offered by Internet or WWW

With these negations in mind it is easier to differentiate the Internet of Things. Consequently, this implies that most publications claiming to address the Internet of Things are not really covering the real essence of the Internet of Things. We suggest two more negations. The Internet of Things is *not the Internet of People*

not the Intranet or Extranet of Things. Therefore, applications that provide only access to a small group of stakeholders (e.g., few companies) should not be considered to represent the full scope of the Internet of Things. However, all fields of research that have been mentioned above overlap partially with the Internet of Things (Figure 1.2).

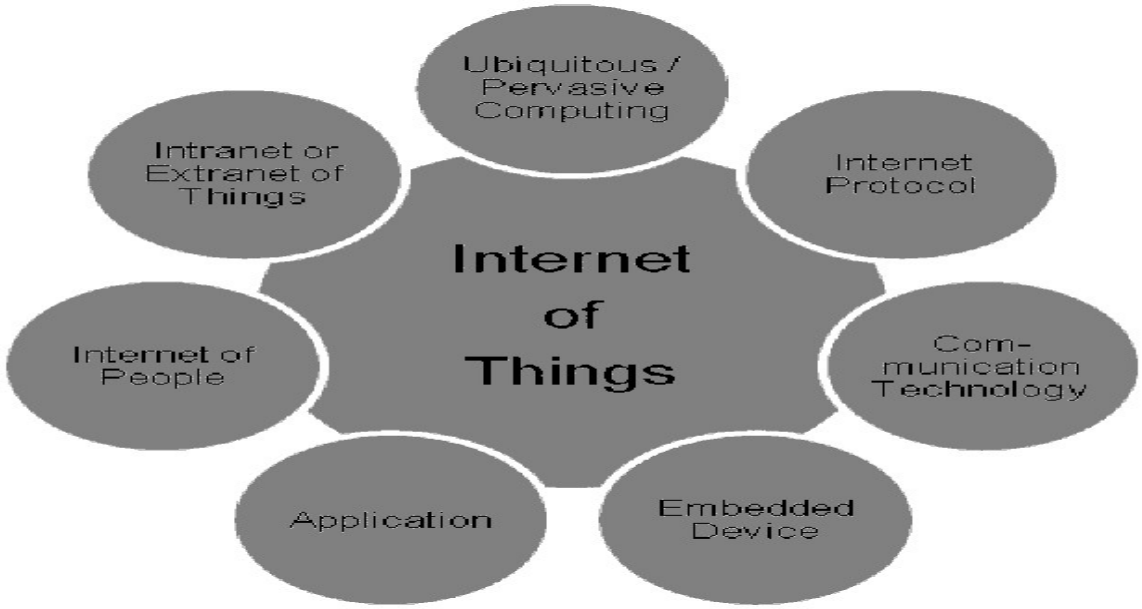


Fig. 1.2 Overlaps of the Internet of Things with Other Fields of Research

The second problem in the CERP-IoT definition is the missing Web 2.0 integration. One could argue that the Web 2.0 is exemplified only by certain types of applications in the Internet of People, which again is not equal to the Internet of Things. However, the Web 2.0 has changed usage of the WWW by

providing more intuitive interfaces for user interaction, social networking and publication of user-generated content, without requiring fundamental changes to the design and existing standards of the internet. The primary advantage of Web 2.0 technology has been the use of intuitive interfaces to enable web contributions by end-users irrespective of their technical expertise. The interaction between things and people will be one core issue in the future Web of Things. End-user product ratings and usage instructions provide a valuable set of information on things. Unfortunately today this information is very much scattered across the WWW and there is no direct link to a product identifier.

Thirdly, the reason for success is missing in the above CERP-IoT definition. Maybe a definition on the Internet of Things does not require a benefit statement – the Internet of Things itself surely does, if it is ever to become a reality. Initially, most applications of Auto-ID technologies were internal or closed-loop applications rather than applications across company boundaries. The main reason is the missing benefit for the individual participants. While benefits can be easily calculated across supply chains or product life cycles, input data to cost- benefitanalysis is most often based on "educated guessing" (Gille and Strüker 2008, Laubacher et al. 2006) rather than on hard facts.
Another approach towards a definition of the Internet of Things can be derived from logistics where it is common to ask for the *right product* in the *right quantity* at the *right time* at the *right place* in the *right*

Condition and at the *right price*. In this analogy the *right product* relates to accurate and appropriate information about a uniquely identifiable physical object as well as its form, fit and function. This includes the usage of Auto-ID and appropriate sensor information or any other kind of linked information to the object that can be accessed through the Internet of Things. The *right quantity* can be achieved through high granularity of information combined with filtering and intelligent processing. The *right time* does not necessarily mean anytime, but more precisely 'when needed'. It may be sufficient to receive information about an object only once a day or only in the case of a status change. Consequently, right-time does not equal real-time, a term that is mentioned quite often in relation to the Internet of Things.

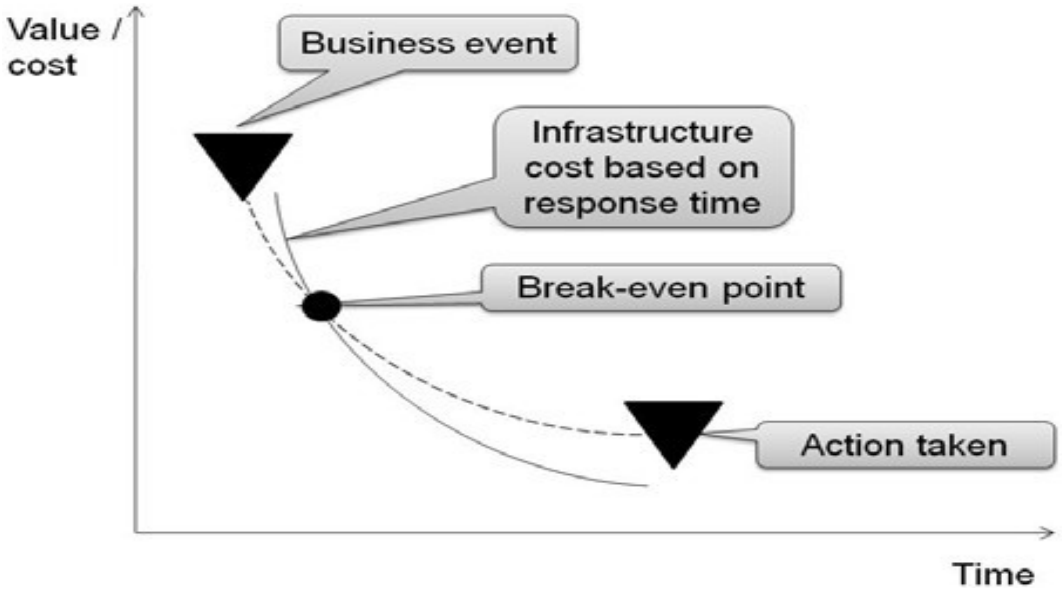


Fig. 1.3 Infrastructure cost vs. response time (based on Hackathorn 2004)

In general, real-time access to data is desirable to reduce the latency between a business event and a corresponding action; the ability to achieve such a reduction is also referred to as agility. Unfortunately, real-time capability is linked to high infrastructure cost (Figure 1.3).

Similarly, the information availability at *right place* does not imply any place - but rather, where the information is needed or consumed (which may not necessarily be the same place as where it is generated). If information is not generated and consumed in the same place and if either of these places have unreliable or intermittent network connectivity, then effective data synchronisation protocols and caching techniques may be necessary to ensure availability of information at the right place. Again, the cost of any place availability has to be seen in relation to its profit potential. But as mobile devices are more and more ubiquitous, there will evidently be an opportunity to access information in the Internet of Things at any place at a reasonable price. The *right information* condition is met if it can be utilised with a minimum effort. This includes human readable information for human interaction as well as semantically and syntactically enriched machine-readable information, which may in turn require transformation of low-level raw data (possibly from multiple sources) into meaningful information and may even require some pattern recognition and further analysis to identify correlations and trends in the generated data. The *right price* is not automatically the lowest price, but instead it is a price between the costs for information provisioning and the achievable market price. Information provisioning costs include labour costs as well as infrastructure costs.

A minimalist approach towards a definition may include nothing more than *things*, the *Internet* and *a connection in between*. *Things* are any identifiable physical object independent of the technology that is

used for identification or providing status information of the objects and its surroundings. *Internet* in this case refers to everything that goes beyond an extranet, thus requiring access to information for more than a small group of people or businesses. A closed loop application consequently has to be regarded as an *Extranet of Things*. The *Internet* acts as a storage and communication infrastructure that holds a virtual representation of *things* linking relevant information with the object.

Combining the different approaches we can conclude that the future Internet of Things links uniquely identifiable things to their virtual representations in the Internet containing or linking to additional information on their identity, status, location or any other business, social or privately

relevant information at a financial or non-financial pay-off that exceeds the efforts of information provisioning and offers information access to non-predefined participants. The provided accurate and appropriate information may be accessed in the right quantity and condition, at the right time and place at the right price. The Internet of Things is not synonymous with ubiquitous / pervasive computing, the Internet Protocol (IP), communication technology, embedded devices, its applications, the Internet of People or the Intranet / Extranet of Things, yet it combines aspects and technologies of all of these approache

Opportunities and Motivation:

Even though there are numerous projects and developments concerning certain aspects of the Internet of Things, an open and accessible infrastructure for a wider adoption of the Internet of Things is missing. A more generic approach towards a future development schedule is needed. While technologies are important building blocks, they are not enough to embrace the large research spectrum that needs to be addressed. The following five subject guidelines may be used to trigger successful and sustainable contributions to the Internet of Things.

1. **Envision** – A vision of the Internet of Things needs to provide holistic scenarios focusing on private, social and business benefits. Experimen- tally-driven, participative research approaches will be needed to allow involvement of different stakeholders for identification of requirements, usability testing, evaluation and active participation. Mechanisms are needed for empowering citizens to fully participate and innovate in the Internet of Things, in order to provide a new multi-directional communi- cation infrastructure for researchers, industries and citizens. This user- centric concept maybe referred to as the 'Web of Things' as it provides intuitive graphical user interfaces that include functionalities familiar to Web 2.0 applications.

2. **Extend** – To leverage state-of-the-art developments and accepted technologies, existing architectures, such as the EPCglobal Network, should be utilised and extended by adding new

functionalities to support diverse means of identification (RFID, barcode, 2D-code), sensors, actuators, intelligent devices and other information sources (e.g. user- generated content, commercial databases) within an open framework. The value of product-related data needs to be increased through semantic enrichment. Extending existing approaches will allow utilisation of prior efforts and investments and allow a phased approach towards the Internet of Things. Disruptive new approaches should be avoided unless they provide substantial new benefits or build on existing work. It should be noted that this approach does not exclude integration of other heterogeneous technologies, but it promotes the usage of a single core architecture.

3. **Enable** – It is crucial to solve today's adoption challenges. There is still a lot of research needed on technical challenges that too often are considered to be solved (especially by researchers and practitioners lacking the technical knowledge). Privacy, security and confidentiality are key factors to provide a trustworthy Internet of Things. New mechanisms for sharing costs and benefits to enable the creation of opportunities for new market entrants are needed.

4. **Excite** – New stakeholders need to be excited to contribute to the future Internet of Things. Ease of participation, collaboration and generation of benefits are major requirements to excite new entrants to the Internet of Things. Open frameworks and end-user programming environments may

empower citizens to create cost-free as well as billable micro services, such as a product guides and reviews.

5. **Evaluate –** New approaches need to be discussed with a large variety of stakeholders and verified in industry pilots and user-centric environments. A good example for the future Internet of Things is the informed and ethical consumer who requires product-related data (e.g., country of origin, ingredients, dynamic best-before date, carbon- footprint) and who is willing to add information to the Internet of Things. Other popular examples include public user-centric scenarios that build on the concept of Smart Cities and Smart Homes. Furthermore, we need to evaluate the Internet of Things in a philosophical context as things will become social actors in a networked environment.

A Possible Architecture for the Future Internet of Things

While it is quite obvious that there are and will be numerous approaches towards the Internet of Things, thus leading to a creative variety of applications in the Internet of Things, we favour an architectural approach that is based on extensions to a successful standardised open architecture – the

EPCglobal Network. The EPCglobal Network is widely accepted and has gained the biggest support from IT companies that have adopted the standardised interfaces into their own applications. Numerous products have been developed and certified (EPCglobal 2010). Therefore, the EPCglobal Network provides a solid foundation, despite the fact that it is still under development.

However, the Internet of Things requires a more holistic architecture as described before. This can build on the same design principles as the EPCglobal Architecture Framework (EPCglobal 2007). These include layering of standards, separation of data models and interfaces, provision of extension mechanisms, specification of data models and interfaces, initially in a neutral abstract manner (e.g., using UML), then with provision of specific transport bindings (e.g., web services) and schema bindings (e.g., XML).

A future Internet of Things has to integrate stakeholders who will be affected by the Internet of Things, such as citizens, small and medium enterprises, governmental institutions and policy makers, to meet and match key societal and economic needs. Applications that recognise and improve the fundamental qualities of life for users, businesses, society and the environment are needed.

The foundation will need to provide open architectures, protocols and technologies for new classes of

smart Internet-/Web-based public and business applications. Social platforms to share experience and personalized insights will be integrated with business-centric applications. Discovery and retrieval of useful and relevant information beyond personal expectations will be achieved though engineering for serendipity. Users shall be empowered to access more information about things (e.g., Where has an item been produced? – Who owned it previously

- What was it used for?) instantly at their fingertips, subject to compliance with privacy regulations. Mash-ups and end-user programming will enable people to contribute to the Internet of Things with data, presentation and functionality. Things-generated 'physical world' content from Auto-ID, sensors, actuators or meshed networks shall be aggregated and combined with information and events from 'virtual worlds', such as business databases and social platforms, and processed based on new business intelligence concepts. Results will be displayed in a user-centred design, including intuitive interfaces and Web 2.0 functionalities. Direct action on the physical world will be supported through Internet of Things machine-interfaces and introduction of agile strategies. Buying decisions will be supported through the access to relevant information as needed. Agile strategies in this context refer to real-time management and execution capability under consideration of conflicting optimisation values (e.g., shipment size).

Information sharing will be rewarded through incentives, including transparent, open billing interfaces between numerous stakeholders, thus transforming the Internet of Things from a cost-focused infrastructure to a benefit-focused infrastructure to accelerate business innovation. Distributed data ownership across the object life cycle will be addressed by integrated billing. Information will be as easily tradable as products and services. The gap between distributed intelligence concepts (e.g., autonomous logistics) and the Internet of Things will be overcome through integration of open interfaces, protocols and lookup services as well as information services on mobile devices, acting as a mediator among decentralize information systems. Openness, scalability and security will be addressed as an integral part of the core architecture. Openness includes social (e.g., governance, privacy), organizational (e.g., industries) and technical (e.g., infrastructures, identifiers) dimensions. The integration and interoperability with mainstream business software platforms will be enhanced and its functionality will be extended through real-time analytics and business intelligence.

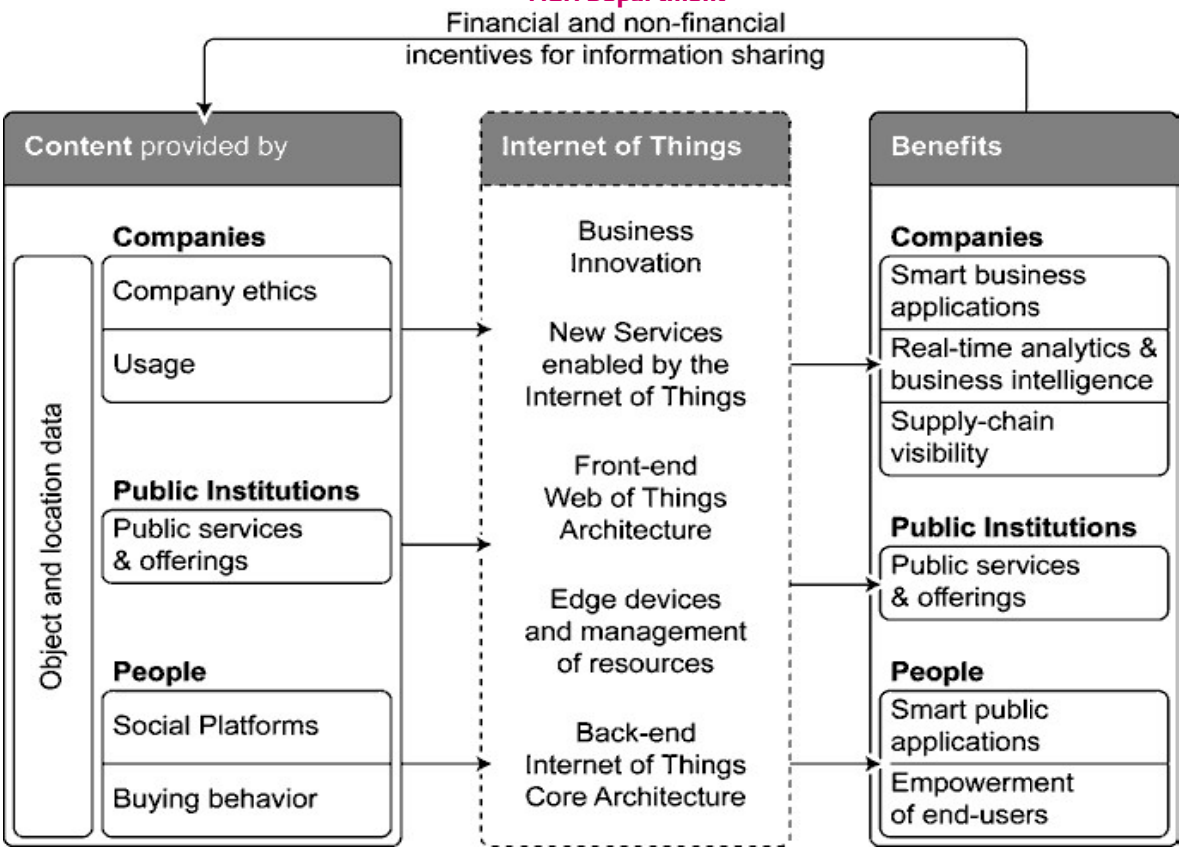


Fig. 1.4 A Holistic Internet of Things Scenario Including Companies, Public Institutions and People

Figure 1.4 shows one possible scenario that includes content providers (producers) and content users (consumers) that utilise the Internet of Things and share benefits. Company data includes for example product and usage data as well as company ethics that may influence buying behaviour. Public institutions as well

as people will be able to contribute content. New services and business innovation will be enabled by an enhanced Internet of Things infrastructure including edge devices and back-end services as well as front-end user-interfaces. Companies, public institutions and people will be able to access data for their own benefits and financial as well as non-financial benefit compensation will further add to a fast adoption process of the Internet of Things.

Key goals for a future Internet of Things architecture to achieve are:

- An open, scalable, flexible and secure infrastructure for the Internet of Things and People

- A user-centric, customisable 'Web of Things' including interaction possibilities for the benefit of society

- New dynamic business concepts for the Internet of Things including flexible billing and incentive capabilities to promote information sharing

The EPCglobal Network architecture is currently only one aspect of the broader Internet of Things. However, if openness, scalability and security can be assured, the EPCglobal Network could be the most promising and comprehensive architecture in the Internet of Things. The availability of free, open standards and free open source implementations for the EPCglobal Network architecture may play a significant enabling role in its development, alongside complementary technologies and standards, such as Open Geospatial Consortium (OGC) Sensor Web Enablement. Other extensions, such as support for multiple identifier schemes, federated discovery services, actuator integration and software agents for decentralised data processing and decision rendering, could further extend the functionality of the EPCglobal Network.

The vision of the future Internet of Things includes extended Internet of Things Information Services based on the EPC Information Services. The extensions are necessary to provide a broader support for other identifiers than the EPC, additional static and dynamic data, actuator support, software agent integration, integration of non-IP devices and offline-capabilities. In detail, the vision includes the following components:

- *Extended static data support* – The EPCglobal Network today is based on the EPC. The EPC is not a single identifier scheme but a framework supporting multiple identifier schemes including GS1 identifiers such as Serialised Global Trade Identification Number (SGTIN), Serial Shipping Container Code (SSCC), and Global Returnable Asset Identifier (GRAI). This framework is not limited to GS1 identifiers; EPC formats are also defined for unique identifier constructs specified by the US Department of Defense. In

- principle, other ap- proaches such as the Uniform Resource Names (URNs) could be used to sup- port identifiers based on ISO 15962 and even identifiers based on Uniform Re- source Locators (URLs) could be included, since they are a subset of Uniform Resource Identifiers (URIs). There is a need to support all things that carry a unique ID, because changing an

established identifier scheme in an industry can cost millions of Euro and should be compared to the efforts involved for changing databases in the last millennium to make them year 2000 compliant. There have been and continue to be approaches to transform existing estab- lished identification schemes into a format that is compatible with the EP- Cglobal Network, as well as EPCglobal standards such as Tag Data Standard (TDS) and Tag Data Translation (TDT) that enable two-way translation be- tween an EPC representation and an existing legacy representation. Additional structured data in barcodes (e.g., for best-before-date) may need to be sup- ported to fully integrate existing optical identification techniques and to exploit the user

memory capabilities of RFID tags, as well as facilitating stock rotation, product recalls, etc. An open, universal identifier translation framework would enable all things that carry a unique ID to be part of the Internet of Things. However, until everything carries a unique ID, the Internet of Things may also need to support objects identified by a classID (productID) and attributes.

- *Integration of dynamic data* – In order to bring the real and the virtual world closer together there is a need to sense environmental conditions as well as the status of devices. A standardized sensor interface to the Internet of Things would help to minimise costs and foster implementation. Sensors are

- key com- ponents of the next generation of internet services because they empower bot- tom-up interaction with things by enabling the gathering of information about their state or condition within the real world. The state of the things can be used to feed services at the infrastructure layer, transforming everyday things into true enablers of the Internet of Things.

- *Support for non-IP devices* – Non-IP devices offer only limited capability. They can be integrated in the Internet of Things through gateways that take care of the computational overhead required to share physical devices over the Internet, while also providing advanced functionality that are not available on the devices themselves.

- *Integration of an actuator interface* – Actuator integration into the Internet of Things will allow standardised communication with machines executing deci- sions either rendered by humans

or software-agents on their behalf. Actuators complement bidirectional interaction processes by providing

- the means for ser- vices and users to influence the state of things. The combination of sensors and actuators and their integration in the core Internet of Things infrastructure is an indispensable feature and needs to be considered at all layers of the architec- ture.
- *Optional integration of software agents* – The complexity of global supply networks will require more decentralised and automated decision making. Software-agents have been researched broadly but have not yet gained consid- erable acceptance in industries. The reason for this may be the lack of stan- dardisation. A standardised interface in the Internet of Things would help to boost the usage of software agents. Smart objects in the Internet of Things need to execute intelligent algorithms to be able to discard irrelevant data, interact with other things in an efficient way, raise warnings about their state or the state of their environment, and take informed decisions and actions on behalf of human end-users to eliminate or assist control / management activities by hu- mans. Additionally, software agents may help to increase scalability and ro- bustness in the Internet of Things (Uckelmann et al. 2010). In a holistic sce- nario we imagine things to host a certain infrastructure subset of the Internet of Things.

- These things may not always be connected to the Internet. Therefore, we envision a certain degree of smart characteristics and autonomy.
- *Extended, federated discovery services* – The EPCglobal Network today does not yet provide ratified standards for federated discovery services, although a technical standard for discovery services is currently under development. At the time of writing, the only lookup service currently

- provided by EPCglobal is the ONS, which only holds class-level records pointing to authoritative informa- tion. This is currently operated under contract by VeriSign Corp. under the on- sepc.com domain. The existing ONS implementation is distributed across mul- tiple servers globally. Nevertheless, there are political concerns that the ONS is defined under the .com Top-Level-Domain, which is under the authority of the US Department of Commerce and that the ONS service is operated only by one American company. This has led to political discussions on governance in the Internet of Things, resulting in national focused approaches in China and Europe (Muguet 2009). Federated discovery services are needed to enable open governance, scalability and choice of lookup service in the Internet of Things.

- *Data-synchronization for offline support* – The EPCglobal Network requires online connection to access data related to the identified product. In certain cases online-connectivity cannot be assured. Data-synchronization is needed to support mobile scenarios and decentralised decision making.

- *Interface to federated billing services* – In order to enable competition between billing service providers, a standardized interface to these services is needed. This billing interface will enable balancing of costs and benefits as well as new business models and revenue generation opportunities for business and citizens based on micro-trading of information in the Internet of Things.

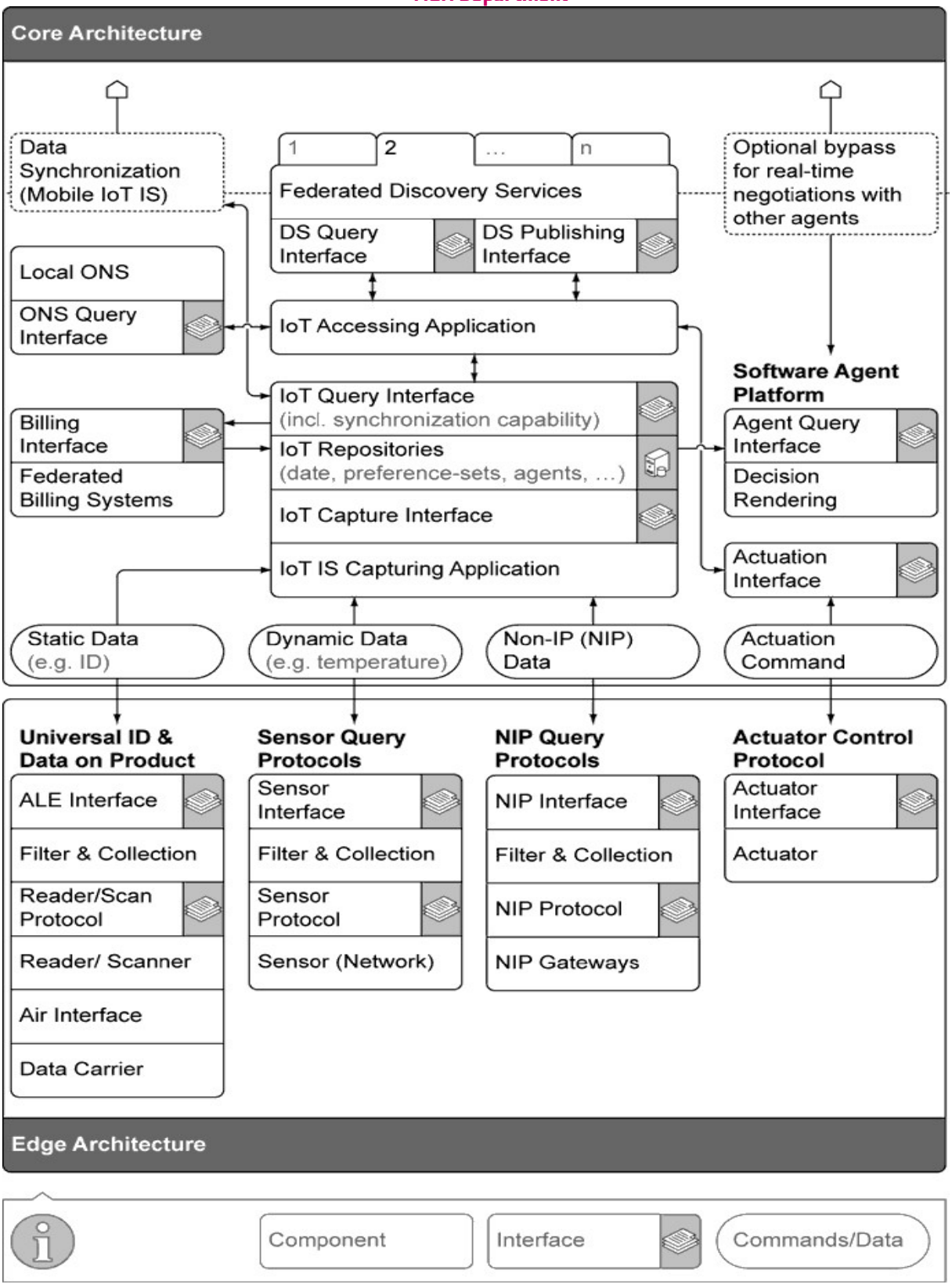

Core Architecture
Data Synchronization (Mobile IoT IS)
1
2
...
n
Federated Discovery Services
DS Query Interface
DS Publishing Interface
Optional bypass for real-time negotiations with other agents
Local ONS
ONS Query Interface
IoT Accessing Application
Software Agent Platform
IoT Query Interface (incl. synchronization capability)
Billing Interface
Federated Billing Systems
IoT Repositories (date, preference-sets, agents, ...)
Agent Query Interface
Decision Rendering
IoT Capture Interface
Actuation Interface
IoT IS Capturing Application
Static Data (e.g. ID)
Dynamic Data (e.g. temperature)
Non-IP (NIP) Data
Actuation Command
Universal ID & Data on Product
Sensor Query Protocols
NIP Query Protocols
Actuator Control Protocol
ALE Interface
Sensor Interface
NIP Interface
Actuator Interface
Filter & Collection
Filter & Collection
Filter & Collection
Actuator
Reader/Scan Protocol
Sensor Protocol
NIP Protocol
Reader/ Scanner
Sensor (Network)
NIP Gateways
Air Interface
Data Carrier
Edge Architecture
Component
Interface
Commands/Data

Fig. 1.5 An Extended EPCglobal Architecture Towards a Future Internet of Things

Hefe1 the integration of sensors, actuators and software agents connected to the Internet of Things Information Service (IoT IS) is shown. Parts of this infrastructure may be mobile and disconnected, thus requiring means for synchronisation of data and logic.

Accessibility of information will be enabled through federated discovery services, which will support open governance and choice of lookup service in the Internet of Things. In the Internet of Things, human beings, software systems and smart things will have a strong need for technologies supporting them in the search and discovery of the many distributed resources available, including information repositories, sensors, actuators, etc. These search and discovery services will rely upon mechanisms for universal authentication and access control, at the desired level of granularity, through which resource owners can precisely control the criteria that determine whether their resources may be discovered by others.

iot: A Web 3.0 View:

The Internet (network) and the web (application) are two sides of a coin. The Internet was invented by Vinton Cerf in 1973, and the invention of the web in 1989 was credited to Tim Berners-Lee and later caught worldwide attention by Marc Andreessen's Mosaic web browser in 1992. The Internet (hardware) is the infrastructure and the web (software) is the application everybody uses. Just like the Internet revolution, in the Internet of Things, web-based applications and soft- ware (the supporting data representation and middleware) are the keys. McKinsey [36] s Ioummarized the key application functional- ities of T systems:

1,Information and analysis

- Tracking behavior
- Enhanced situational awareness
- Sensor-driven decision analytics

2.Automation andcontrol

Process optimization

- Optimized resource consumption
- Complex autonomous systems

According to Harbor Research, the web-based applications, systems, and networked services of smart systems or IoT are expanding more rapidly than the hardware and infrastructure [37]. This means the software (middleware and web-based integrated applications) market will play a pivotal role in the IoT business.

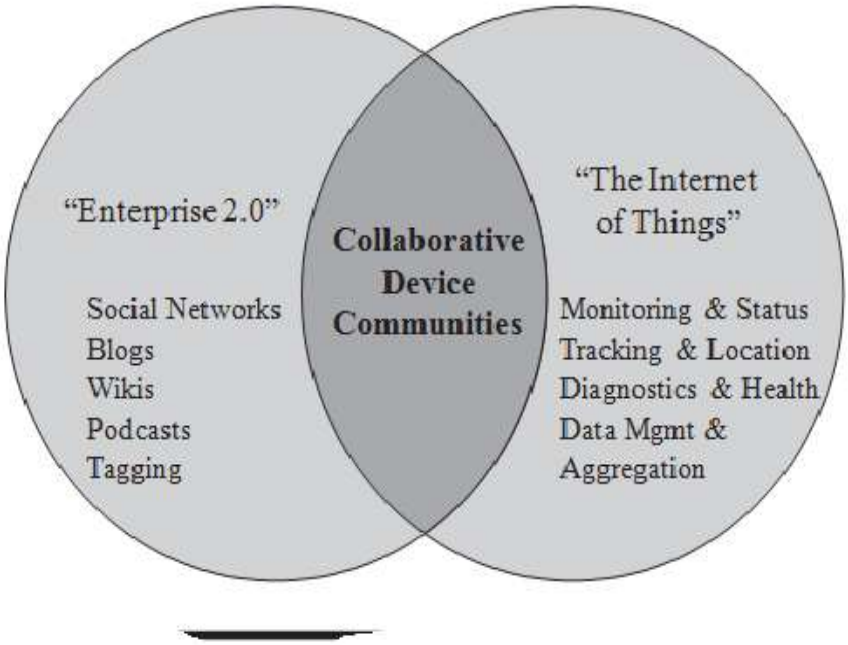
As is well known, Web 1.0 is about publishing and push- ing content to the users. It's mostly a unidirectional flow of information. The shift from Web 1.0 to Web 2.0 can be seen as a result of technological refinements as well as the behavior change of those who use the World Wide Web, from publish- ing to participation, from web content as the outcome of large up-front investment to an ongoing and interactive process. Web 2.0 is about two-way flow of information and is associ- ated with web applications that facilitate participatory infor- mation sharing, interoperability, user-centered design, and collaboration. Example applications of Web 2.0 include blogs, social networking services (SNSs), wikis, mashups, folkson- omies, video-sharing sites, massive multiplayer online role-playing games, virtual reality, and so on.

Enterprise 2.0 is the use of Web 2.0 technologies within an organization to enable or streamline business processes while enhancing collaboration (Figure 1.8). It is the extension of Web 2.0 into enterprise applications. IoT technologies and applications can be integrated into Enterprise 2.0 for enter- prises that need to monitor and control equipment and facili-ties and integrate with their ERP and CRM back office systems.

Definitions of Web 3.0 vary greatly. Many believe that its most important

Features are Semantic Web and personaliza- tion; some argued that Web 3.0 is where the *computer* is gen- erating new information rather than the human.

The term Semantic Web was coined by Tim Berners-Lee, the inventor of the World Wide Web. He defines the Semantic Web as "a web of data that can be processed directly and indi- rectly by machines." Humans are capable of using the web to carry out tasks such as reserving a library book or searching



networks Summary After decades of fast-paced development, telecom worldwide now basically satisfy the need for man-to-man

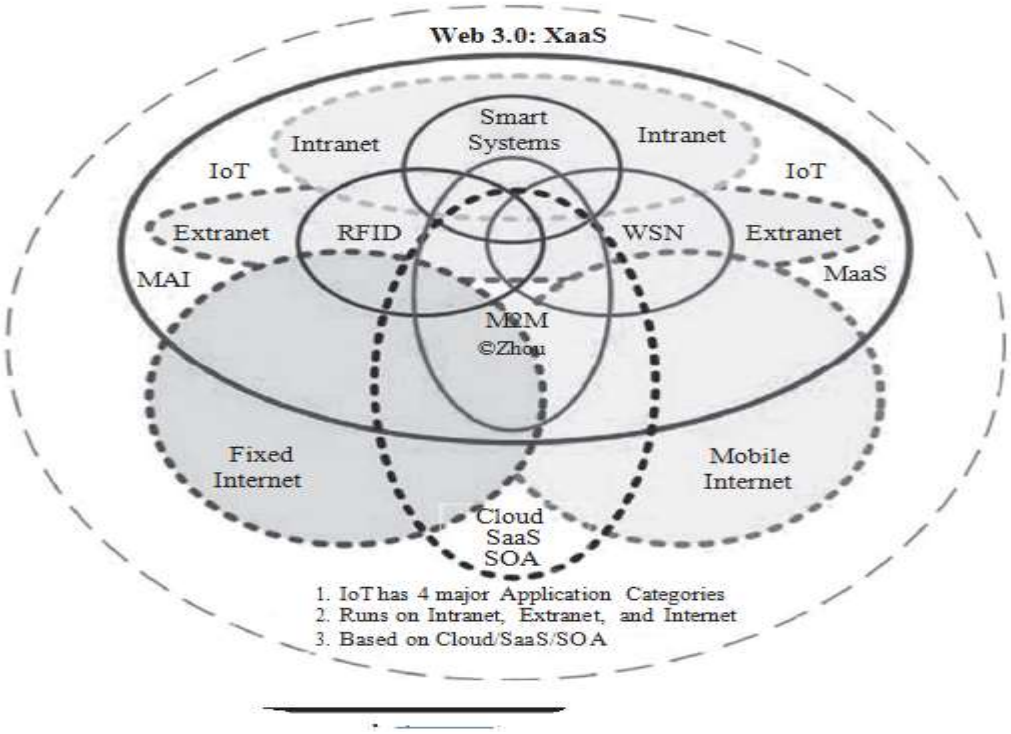


Figure 1.9 Web 3.0: the internet of things.

Communication anywhere and at any time. However, new demand has arisen for machine-to-machine and machine- to-man, or the Internet of Things, communications. The development of these M2M technologies has attracted greater attention in recent times in light of the "smart Earth" and "Sensing China" concepts proposed by the American and Chinese governments and other parts of the world such as the European Union following the global financial crisis.

According to Forrester Research, by 2020 machine-to-machine data exchange will be 30 times greater than the number of exchanges between people. M2M or IoT is therefore consid- ered the next trillion-dollar segment of the international tele- com market.

The physical world itself is becoming a connected infor- mation system. In the world of the Internet of Things, sen- sors and actuators embedded in physical objects are linked through wired and wireless networks that connect the Internet. These information systems churn out huge volumes of data that flow to computers for analysis. When objects can both sense the environment and communicate, they become tools for understanding the complexity of the real world and responding to it swiftly.

The Internet of Things and related concepts, terms, and phrases and their potentially vast scope of applications as well as their impacts on business and social life were described in this chapter. The definitions of IoT were described and the author also gave his own definition and understanding, which will be the foundation of the book.

In the next chapter, a more detailed, panoramic view of IoT applications will be introduced and a few concrete vertical applications will be described in greater detail.

for a low price for a DVD. However, machines cannot accom- plish all of these tasks without human direction, because web pages are designed to be read by people, not machines. The Semantic Web is a vision of information that can be readily interpreted by machines, so machines can perform more of the tedious work involved in finding, combining, and acting upon information on the Web.
Some consider the Semantic Web an unrealizable abstrac- tion and see Web 3.0 as the return of experts and authorities to the Web. I share the same thought. If there is no tangible difference but only a conceptual one, the concept of Semantic Web–based Web 3.0 doesn't stand on solid ground. Rather, the Web 3.0 of machine-generated data is more practical, makes more sense, and is possible to implement.
While Web 3.0 arguments are not yet settled, some people have started talking about Web 4.0 [30], the ubiquitous Web.
A fundamental difference between the Internet of People (Web 1.0 and Web 2.0) and the Internet of Things is that in the former, data are generated by people (keyed in by hand, photographed by hand, etc.); in the latter, data are generated by machines, not humans. This difference makes it enough to start a new version of the World Wide Web, or Web 3.0.
The data are generated by things and consumed by people and machines via SaaS or XaaS (Everything as a Service), and this model constitutes the basis of Web 3.0 as depicted in Figure 1.9 [74]. We choose to use the term Web 3.0 instead of Web 4.0 based on the concept of machine-generated data in addition to the Semantic Web, which seems to not have much substance up to now. It is too much of a jump to go to Web 4.0.

Four Pillars of iot:

the Horizontal, Verticals, and Four Pillars

Applications of the Internet of Things (IoT) have spread across an enormously large number of industry sectors, and some technologies have been used for decades as described in the previous chapter. The development of the vertical applications in these sectors is unbalanced. It is very important to sort out those vertical applications and identify common underpin- ning technologies that can be used across the board, so that interconnecting, interrelating, and synergized grand integration and new creative, disruptive applications can be achieved.

One of the common characteristics of the Internet of Things is that objects in a IoT world have to be instrumented (step 3 in Figure 3.1), interconnected (steps 2 and 1), before anything can be intelligently processed and used anywhere, anytime, anyway, and anyhow (steps 1 and 2), which are the 5A and 3I [180] characteristics.

Another common feature that IoT brought to information and communications technology (ICT) systems is a fund- mental change in the way information is generated, from

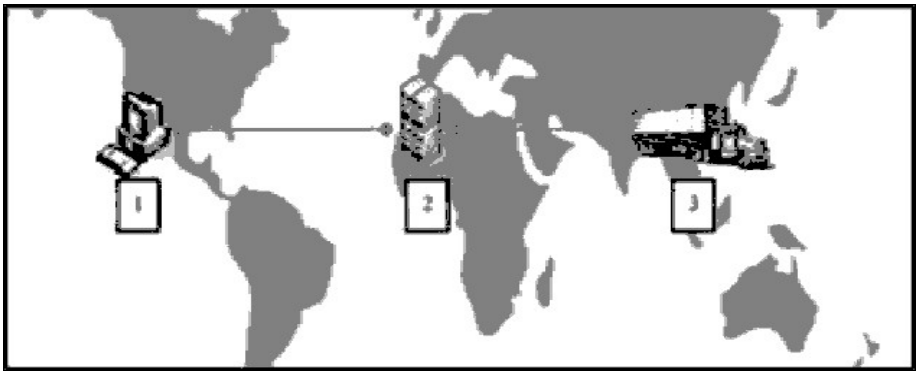


Figure 3.1 3i and 5A.

mostly manual input to massively machine-generated without human intervention.

To achieve such 5A (anything, anywhere, anytime, anyway, anyhow) and 3I (instrumented, interconnected, and intelligent) capabilities, some common, horizontal, general-purpose tech- nologies, standards, and platforms, especially middleware platforms based on common data representations just like the three-tiered application server middleware, HTML language, and HTTP protocol in the Internet/web arena, have to be estab- lished to support various vertical applications cost effectively, and new applications can be added to the platform unlimitedly.

Most of the vertical applications of IoT utilize common technologies from the networking level and middleware plat- form to the application level, such as standard wired and wireless networks, DBMS, security framework, web-based three-tiered middleware, multitenant PaaS (platform as a service), SOA (service-oriented architecture) interfaces, and so on. Those common technologies can be consolidated into a general-purpose, scalable framework and platform to better serve the vertical applications as demonstrated.

Service-management platforms (SMPs) are the key to entry into the machine-to-machine (M2M) market. They allow for the essential connectivity management, intelligent rate-plan management, and customer self-service capability that are today's fundamental prerequisites for providing a successful, managed M2M service. Consequently, with its acquisition of Telenor Connexion's M2M SMP technology and the staff related to the platform's development, Ericsson has taken a decisive step into the market. Ericsson has built a horizontal platform for the 50 billion M2M market's vertical telemetric, medical, utilities, and government applications [203].

Telenor Objects was formed in July 2009 by researchers and developers in Telenor Norway and Telenor R&I. The two entities had individually been working on piloting managed M2M services since 2007, with an RFID (radio-frequency identification) focus in Telenor Norway, and a focus on trace-and- track initiatives in Telenor R&I. Telenor Objects [104] aims to provide a layered and horizontal architecture for connecting devices and applications. The company's platform, dubbed Shepherd, adheres to ETSI's standardization initiative on con- nected objects and provides a device library as well as a set of enablers to device and application providers. In addition,

Shepherd includes a range of operational management services. As a driver for connecting devices to the Internet of Things, Telenor Objects is a founding member of coosproject.org (Connected Objects Operating System), a general-purpose, modular, pluggable, and distributable open source middleware platform in Java, designed for connecting service and device objects that communicate via messages and enabling monitor- ing and management. (The targeting devices totaled 2.675 tril- lion according to Telenor Objects and Harbor Research's Intelligent Device Hierarchy at http://www.harborresearch. com/_literature_32606/News.htm.) The initiative is among several newly established steps by Telenor into the open source and open innovation sphere.

The key benefits of horizontal standard-based platforms will be faster and less costly application development and more highly functional, robust, and secure applications.

Similar to the market benefit of third-party apps (e.g., Apple's application store) running on smart phone platforms, M2M applications developed on horizontal [183] platforms will be able to make easier use of underlying technologies and services. Application developers will not have to pull together the entire value chain or have expertise in esoteric skill sets. This will dramatically increase the rate of innovation in the industry in addition to creating more cross-linkages between various M2M applications

In an issue of the M2M (now *Connected World*) magazine's cover story in 2007 [50], editorial director Peggy Smedley intro- duced a graphic that encapsulates the ever-expanding M2M landscape. The graphic covers the "six pillars" of M2M tech- nology, representing market segments that involve network- ing physical assets and integrating machine data into business systems. The six pillars of M2M are as follows:

- Remote monitoring is a generic term most often repre- senting supervisory control, data acquisition, and automa- tion of industrial assets.
- RFID is a data-collection technology that uses electronic tags for storing data.
- A sensor network monitors physical or environmental conditions, with sensor nodes acting cooperatively to form/maintain the network.
- The term *smart service* refers to the process of network- ing equipment and monitoring it at a customer's site so that it can be maintained and serviced more effectively.
- Telematics is the integration of telecommunications and informatics, but most often it refers to tracking, naviga- tion, and entertainment applications in vehicles.
- Telemetry [185] is usually associated with industrial-, medical-, and wildlife-tracking applications that transmit small amounts of wireless data.

However, there is plenty of overlap among the pillars in this graphic. Pick any application of M2M and chances are it fits into more than one of the six pillars. Take fleet management as an example. It is certainly remote monitoring. It can be considered a smart service depending on who's doing the monitoring. It may have elements of telematics. It fits the technical definition of telemetry. And, there may even be RFID tags or a sensor network onboard.
In this book, a four-pillar graphic is introduced for the broader IoT universe. The four pillars of IoT are M2M, RFID, WSNs (wireless sensor networks), and SCADA (supervisory control and data acquisition):

- M2M uses devices (such as an in-vehicle gadget) to cap- ture events (such as an engine disorder), via a network (mostly cellular wireless networks, sometimes wired or hybrid) connection to a central server (software program), that translates the captured events into meaningful infor- mation (alert failure to be fixed).
- RFID uses radio waves to transfer data from an electronic tag attached to an object to a central system through a reader for the purpose of identifying and tracking the object.
- A WSN consists of spatially distributed autonomous sen- sors to monitor physical or environmental conditions, such as temperature, pressure, motion, or pollutants, and to cooperatively pass their data through the network, mostly short-range wireless mesh networks, sometimes wired or hybrid, to a main location. (Methley et al. [62] reports on the overlaps or coverage differences when WSN was compared with M2M and RFID; SCADA or smart system was not mentioned in the report.)

SCADA is an autonomous system based on closed-loop control theory or a smart system or a CPS that connects, monitors, and controls equipment via the network (mostly wired short-range networks, a.k.a., field buses, sometimes wireless or hybrid) in a facility such as a plant or a building.

The term *SCADA* was picked as one of the pillars of IoT over the terms *smart system* and *CPS*. CPS [28] is more of an aca- demic term, and EPoSS defines *smart system* as "miniaturized devices that incorporate functions of sensing, actuation, and control" [22]. Both of these can be considered parts of the extended scope of SCADA or ICS (industrial control system) under the IoT umbrella.

Smart systems evolved from microsystems. They combine technologies and components from microsystems (miniatur- ized electric, mechanical, optical, and fluid devices) with knowledge, technology, and functionality from disciplines like biology, chemistry, nano sciences, and cognitive sciences.

However, Harbor Research [32] defines smart systems as a new generation of systems architecture (hardware, software, network technologies, and managed services) that provides real-time awareness based on inputs from machines, people, video streams, maps, news feeds, sensors, and more that integrate people, processes, and knowledge to enable col- lective awareness and decision making. Based on this defi- nition, a smart system is close to an industrial automation system, a facility management system, or a building manage- ment system.

Harbor Research's definition is close to what a SCADA sys- tem covers. Due to the difference of the definitions of Harbor and EPoSS, SCADA is chosen as one of the four pillars.

There is much less overlap between these four pillars com- pared with those of the six-pillar categorizations of M2M. The clear categories of the four pillars and the distinct networking technologies are shown in Table 3.1 and Figure 3.2.

table 3.1 Four Pillars of iot and their Relevance to networks

Four Pillars and Networks	*Short-Range Wireless*	*Long-Range Wireless*	*Short-Range Wired*	*Long-Range Wired*
RFID	Yes	Some	No	Some
WSN	Yes	Some	No	Some
M2M	Some	Yes	No	Some
SCADA	Some	Some	Yes	Yes

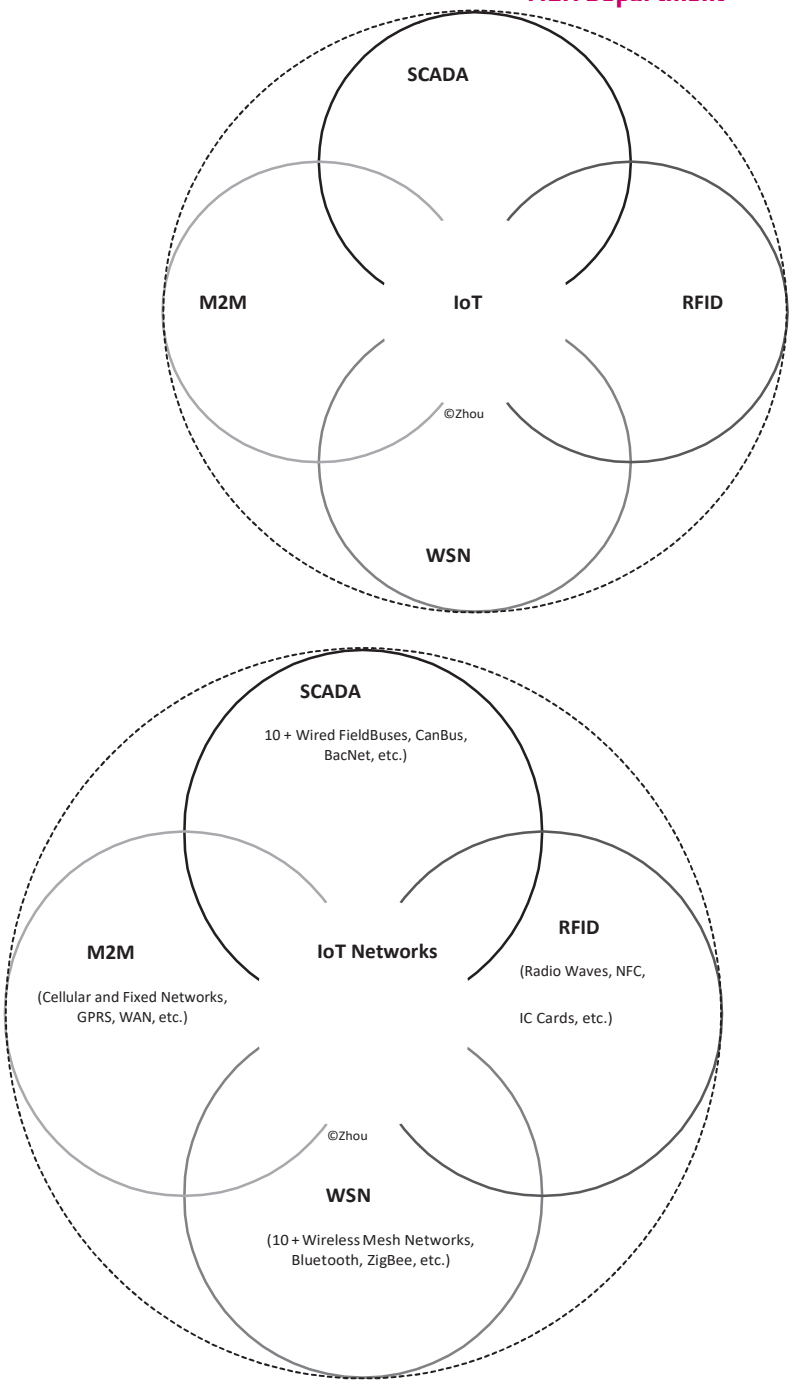


Figure 3.2 the four pillars of iot paradigms and related networks.

The Strategy Analytics research firm also categorized the IoT networks as wired (stationary) and wireless (mobile), and compared their market value and ease of integration as early as 2004 [204]. IoT is the glue that fastens the four pillars through a common set of best practices, networking methodology, and middleware platform. This enables the user to connect all of their physical assets with a common infrastructure and a consistent methodol- ogy for gathering machine data and figuring out what it means. Take away the glue, and end users are left with multiple appli- cation platforms and network accounts. The true power of the Internet of Things occurs when it is working behind the scenes (just like Mark Weiser said about ubiquitous computing) and sharing a common platform, which can't happen if companies have to manage multiple, independent systems.

M2M: the internet of Devices:

Although the rest of the world may not agree, in the United States, *machine-to-machine* is a more popular term than the *Internet of Things*, thanks perhaps to *M2M Magazine*'s efforts since 2004. Two of the six pillars, remote monitoring and smart service, are features or functions of an IoT sys- tem rather than pillars. Conceptually, the terms M2M, RFID, and WSN are similar, but when the underlying communi- cation network is taken into consideration, they are quite different segments. In this book, the term M2M is restricted to refer to device connectivity technologies, products, and services relevant to the cellular wireless networks operated by telco companies. In fact, most of the M2M market research reports assume M2M modules are simply just cellular modules.

Table 3.2 showcases the major applications. However, there is overlap between M2M and the consumer electronics applications. The con- sumer electronics offerings include the following (as opposite to the traditional M2M offerings shown in Table 3.2):

- Personal navigation devices
- eReaders
- Digital picture frames
- People-tracking devices
- Pet-tracking devices
- Home security monitors
- Personal medical devices

ABI Research forecasts that the M2M market is expected to reach more than 85 million connections globally by 2012, and more than 200 million by 2014, with a total market valuation of approximately $57 billion, with utilities (automatic meter reading, telemetry) and automotive (telematics) the clear winners. In fact, it has been assumed that M2M comprises telemetric and telemetry [42]. However, Analysys Mason predicts telemetry (utilities, etc.) will outperform telemetric .

iSuppli's research depicts the worldwide cellular M2M module market by vertical applications in millions of dollars and the market shares of major vendors [206]. Juniper Research

table 3.2 Application Areas for Cellular M2M

Industry	*Example Application*	*Benefits*
Medical	Wireless medical device	Remote patient monitoring
Security	Home alarm and surveillance	Real-time remote security and surveillance
Utility	Smart metering	Energy, water, and gas conservation
Manufacturing	Industrial automation	Productivity and cost savings
Automotive	Tracking vehicles	Security against theft
Transport	Traffic systems	Traffic control for efficiency
Advertising and public messaging	Billboard	Remote management of advertising displays
Kiosk	Vending	Remote machine management for efficiency and cost savings
Telematics	Fleet management	Efficiency and cost savings
Payment systems	Mobile transaction terminals	Mobile vending and efficiency
Industrial automation	Over-the-air diagnosis and upgrades	Remote device management for time savings and reduced costs

estimates there will be approximately 412 million M2M mobile connected devices in the marketplace by 2014 [207].

The number of cellular M2M devices surpassed the number of mobile phones for the first time in Europe in 2010, just a few months later than the time predicted by e-Principles in 2003. According to Beecham Research in August 2011, Cisco recently announced dedicated routers for the M2M market, stating that it believes M2M will become an important mass market. This is just the latest announcement of a series of recent initiatives in the M2M market, both in the United States and in Europe.

In April 2011, Ericsson announced the acquisition of longtime M2M platform provider Telenor Connexion, while in July Telia Sonera announced that it had signed a coopera- tion agreement with France Telecom-Orange and Deutsche Telekom to increase the quality of service and interoperability for M2M services. In May 2011, T-Mobile USA announced that it had cast off its M2M operational business to longtime service partner Raco Wireless, although in July T-Mobile USA struck a partnership with asset protection provider Contain and Asset Protection Products LLC to help reduce operating costs of $7 billion in the US rent-to-own (RTO) sector.
Those and other initiatives signal that the M2M market is deemed ready to truly become a mass market, and players from hardware providers to M2M specialists passing through Telco operators and system integrators [208] are trying to position themselves to reap the benefits.
While the executive-level comments and business unit launches from AT&T and Verizon signal a highly promising vision for the future, the reality of the M2M market is differ- ent and less optimistic as seen by other analysts such as Berg Insights. A comparison of analyst projections for the M2M market points to a market of about 100 million unit shipments for 2012 [38]. Strategy Analytics identifies five key barriers to scaling the global M2M market [275]:

- Lack of a low-cost local access media that can be implemented on a global basis
- The fragmented nature of both the technology vendors and the solutions they provide
- Lack of any single killer application that can consolidate the market and drive demand forward
- The increased costs associated with development and inte- gration because of the complex nature of M2M solutions
- Management's inability to express the benefits of M2M in anything other than cost savings, rather than exploit- ing and encouraging the service enablement capacity of mobile M2M

Figure 3.3 shows the typical architecture of an M2M system from BiTX. The integration middleware at the server side is the brain of the entire system.
Cellular networks were designed for circuit-switched voice.

While they do a perfectly adequate job for regular, packet- switched data such as email and web browsing, they do not have the requisite functionality for M2M applications. For example, the normal OSS (operation support system) and BSS (business support system) are not designed for low-cost, mass handling of huge amounts of similar subscriptions. That led to the development of service enablement middleware platforms by specialized service providers (Table 3.3).

Figure 3.3 BitX M2M architecture based on middleware.

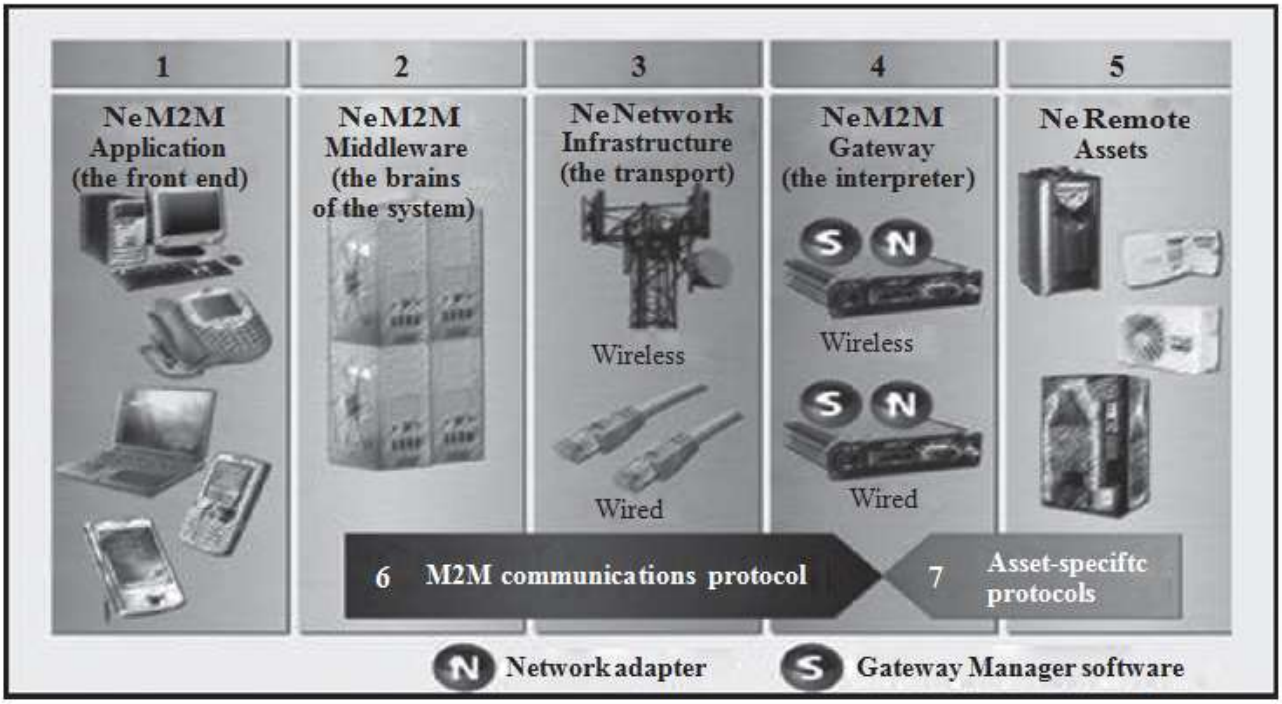


table 3.3 M2M Service enablement Middleware

Vertical Applications Applications to connect to and communicate with objects tailored for specific verticals. Must be done in partnership with industry.

Service Enablement Middleware (APIs over Internet) Reduce complexities with regard to fragmented connectivity, device standards, application information protocols, etc., and device management. Build on and extend connectivity.
Connectivity (ADSL, SMS, USSD, GSM, GPRS, UMTS, HSPA, WiFi, Satellite, Zigbee, RFID, Bluetooth, etc.) Connectivity tailored for object communication with regards to business model, service level, SIM provisioning, billing, etc.

Service enablement is a middleware layer that facilitates the creation of applications. You can think of it as an operat- ing system that the software developers write to this layer via application programming interface (APIs). A significant per-
centage of the functionality of the middleware comes from the charging, mediation, service management, and network man- agement solutions that are being deployed in next-generation networks. These components have functionality that is similar and in some ways superior to that of regular M2M middle- ware platforms.

Table 3.4 shows the value chain of M2M business, which can be separated into two parts: the first relating to devices and the second to application development and service delivery.

The broad intersection between these two parts represents the means by which devices are procured and integrated into M2M solutions and services. Both MNOs (mobile network operators), with some operators taking a more active role than others, and MVNOs (mobile virtual network operators, as shown in the table), subject to having their devices certified on a host opera- tor's network, are trying to be M2M service providers.

The M2M device market share of chipset vendors including TI, Infineon, ST-Ericsson, Qualcomm, and others, and module vendors including Enfora, Infone, Kyocera, Murata, Mobicom, Novatel, Panasonic, Semco, Siemens, Sierra Cellular, Simcom, Telit, Wavecom, and others

As MNOs become more directly involved with M2M appli- cation service providers (ASPs), some MNOs such as Sprint, AT&T, Verizon Wireless, China Mobile, China Telecom, China Unicom, Orange, Rogers Communications, Telenor, Telefonica, NTT DOCOMO, and others are actively deploying M2M-based services. Many are deploying key network elements, specifi- cally mobile packet gateways (e.g., Gateway GPRS Support Node [GGSN], Packet Data Serving Node [PDSN], Home Location Register [HLR], etc.), specifically for their M2M opera- tions, separate from their general mobile data infrastructure.

Key benefits of doing this are that it simplifies internal busi- ness operations and optimizes use of the network.

Likewise, MVNOs active in the M2M market are also increasingly deploying mobile packet gateways and similar equipment to interconnect with their MNO partners' radio infrastructure. (ABI Research classifies MVNOs who have deployed HLRs and mobile packet gateways as "MMOs" i.e., M2M Mobile Operators, Aeris Communications, Jasper Wireless, Numerex, Kore Telematics, Wyless, Qualcomm nPhase, Wireless Maingate, etc., are examples of MMOs.) The benefits to the MVNO for doing this include the ability to cre- ate new service offerings independently of their MNO partners and to enable quicker provisioning and diagnostic capabilities to their ASP customers. MMOs and ASPs are called M2M partners of MNOs. They could use only the connectivity services of an MNO or other services such as rating and charging. Amazon eReaders, M2M DataSmart, FleetMatics, TeloGis, and others are examples of ASPs. Jasper Wireless is an example that uses less services of MNOs in some applications, because it's also an MMO.

As more and more MNOs start to enter into the M2M market directly, such as Telenor Objects, etc., some ASPs and MMOs are forced to become mobile virtual network enablers (MVNEs), that is, MNO or MVNO enablers for M2M. For example, Jasper Wireless is an MVNE of some of AT&T's M2M businesses.

There is virtually no MVNO in existence in China because there is no regulation allowing such a business or service; the Big Three state-owned telcos, China Mobile, China Unicom, and China Telecom, dominate the market. Based on the flag- ship product ezM2M Middleware Platform for IoT applications, built at THTF Co., Ltd. (the second largest system integrator of China) led by the author, THTF has successful established a joint venture with China Mobile to construct the M2M Platform for China Mobile's M2M/IoT base in ChongQing serving nationwide users for all vertical applications.

RFiD: the internet of objects:

The term *Internet of Things* was first used by Kevin Ashton, co-founder and executive director of the Auto-ID Center, when he was doing RFID-related research at Massachusetts Institute of Technology in 1999. The Auto-ID lab is a research federa- tion in the field of networked RFID and emerging sensing technologies, consisting of seven research universities located on four different continents chosen by the former Auto-ID Center to design the architecture for the Internet of Things together with EPC global. The technology they have developed is at the heart of a proposal sponsored by EPC global and sup- ported by GS1, GS1 US, Walmart, Hewlett-Packard, and others to use RFID and the electronic product code (EPC) in the identification of items in the supply chain for companies.

An RFID tag is a simplified, low-cost, disposable contact- less smartcard. RFID tags include a chip that stores a static number (ID) and attributes of the tagged object and an antenna that enables the chip to transmit the store number to a reader. When the tag comes within the range of the appropriate RF reader, the tag is powered by the reader's RF

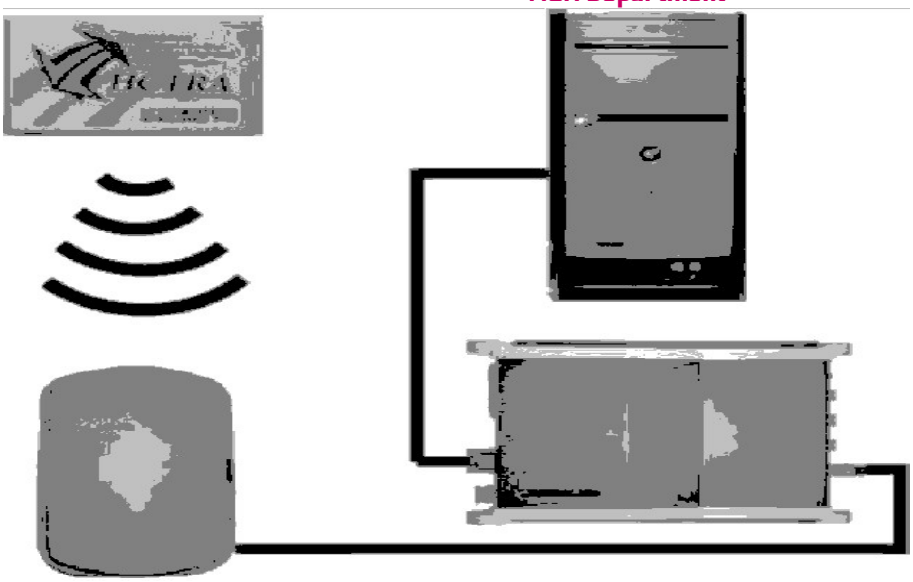


Figure 3.4 RFiD system components. (From erick C. Jones and Christopher A. Chung, *RFID in Logistics: A Practical Introduction,* Boca Raton, FL: CRC Press, 2008.)

field and transmits its ID and attributes to the reader. The contactless smartcard provides similar capabilities but stores more data.

An RFID system involves hardware known as readers and tags, as well as RFID software or RFID middleware (Figure 3.4). RFID tags can be active, passive, or semipassive. Passive RFID does not use a battery, while an active has an on-board battery that always broadcasts its signal. A semipas- sive RFID has a small battery on board that is activated when in the presence of a RFID reader.

The RFID technology is different from the other three tech- nologies of IoT in the sense that it tags on an "unintelligent" object such as a pallet or an animal (an early experiment with RFID implants was conducted by British professor of cybernet- ics Kevin Warwick, who implanted a chip in his arm in 1998) to make it an instrumented [180] intelligent object for monitor- ing and tracking, while the other three (M2M, WSN, and Smart Systems) simply connect "intelligent" electronic devices.

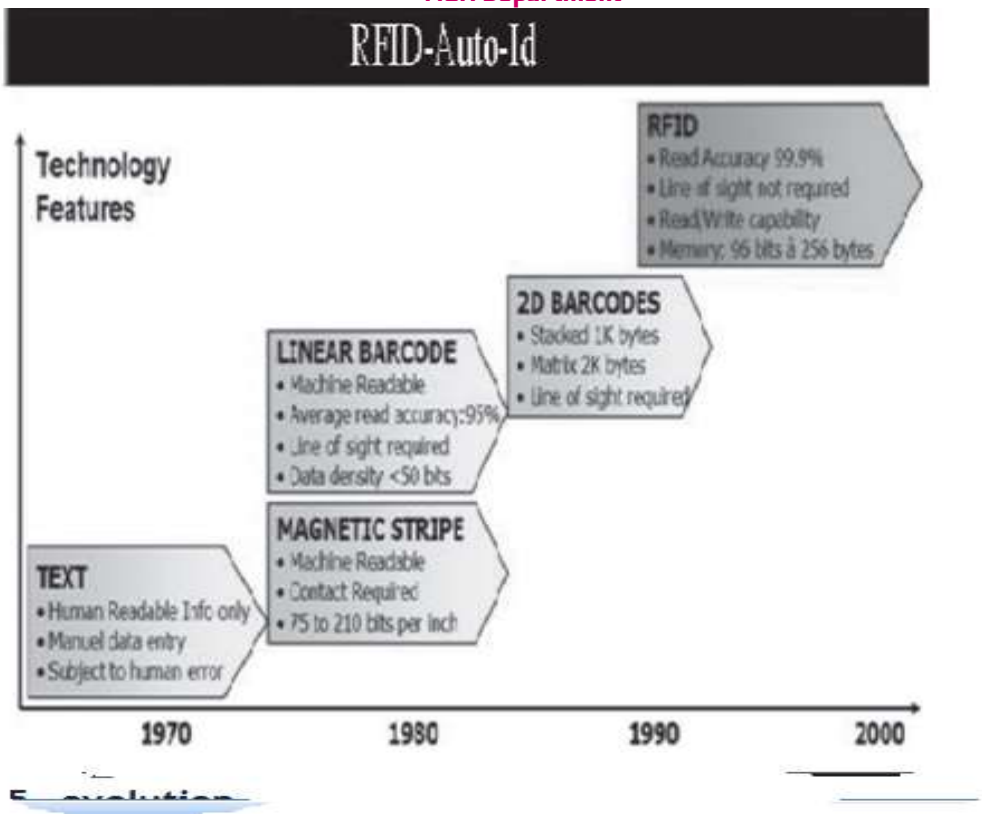


Figure 3.5 evolution of identifications.

Mario Cardullo's passive radio transponder device in 1973 was the first true ancestor of modern RFID. For object or article identifications, text and then barcodes were widely used before RFID tags come into being (Figure 3.5).

UPC (universal product code) of UCC (Uniform Code Council, later called GS1 US) was widely used in the United States and Canada for tracking trade items in stores (Figure 3.6). EAN (European article number), developed after UPC, was used in Europe. EAN International is now called GS1. All the numbers encoded in UPC and EAN (as well as EAN/UCC-13, EAN/ UCC-14, EAN-8, etc.) bar codes are known as global trade item numbers (GTIN). GS1, GS1 US, and Auto-ID labs joined forces to form EPCglobal in 2003 (which means the United States and Europe share the EPC standard; however, UID

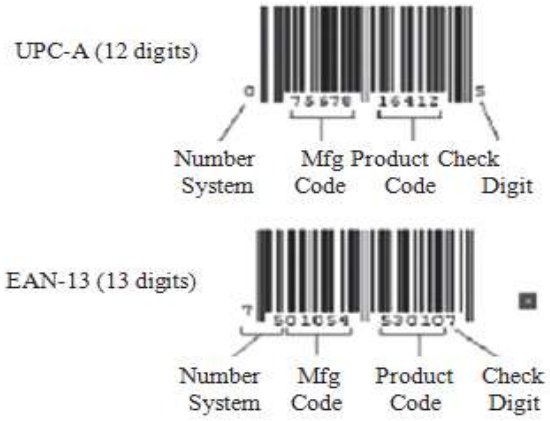


Figure 3.6 Bar code formats. (From James B. Ayers and Mary Ann odegaard, *Retail Supply Chain Management*, new York: Auerbach Publications, 2008.)

[ubiquitous ID] is used in Japan). EPCglobal is an organization set up to achieve worldwide adoption and standardization of EPC technology. The main focus of the group currently is to create both a worldwide standard for RFID and the use of the Internet to share data via the EPCglobal Network™. The automotive industry has been using the technology in manufacturing for decades. Pharmaceutical companies are already adopting the technology to combat counterfeiting.

The Department of Homeland Security has been looking to leverage RFID along with other sensor networks to secure supply chains and ensure port and border security. Many major businesses already use RFID for better asset visibility and management. But the RFID technology and applications became widely used after the industry mandates started in 2004. Walmart and the U.S. Department of Defense (DOD) along with some other major retailers required their suppli- ers to begin RFID tagging pallets and cases shipped into their distribution centers in 2005 (http://www.controlelectric.com/RFID/Wal-Mart_DOD_Mandates.html). The mandates impacted some 200,000 suppliers globally. That year was also when the ITU published the Internet of Things report. Many companies

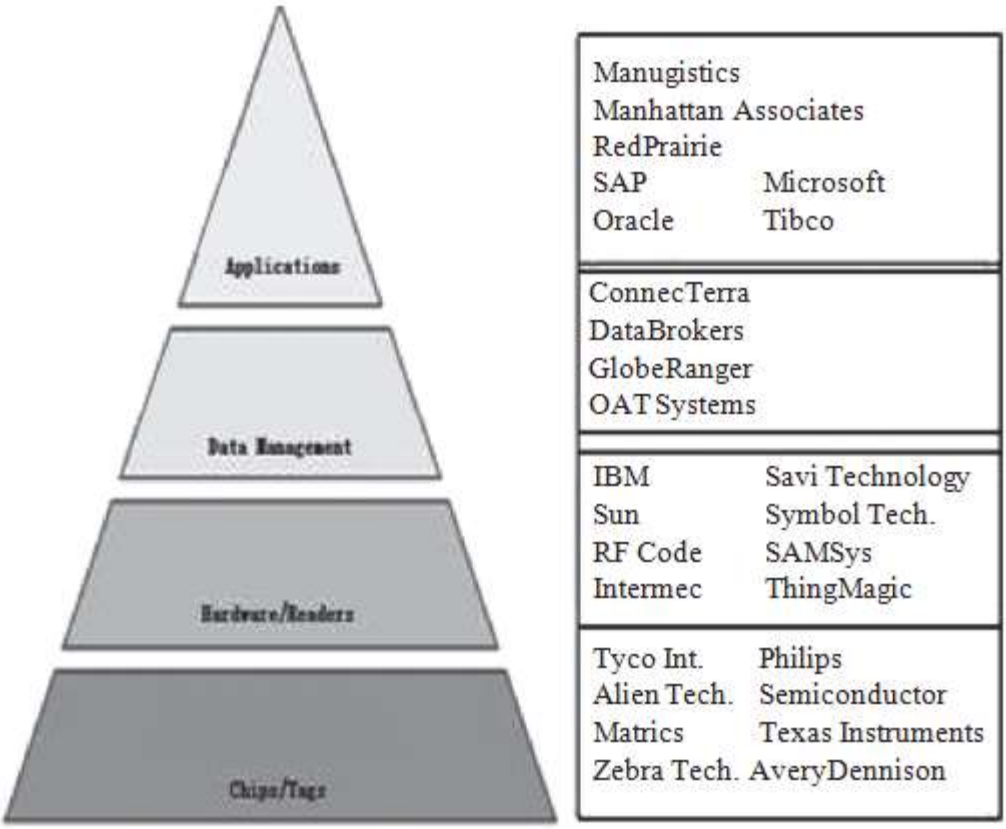


Figure 3.7 RFiD value chain and vendors.

worldwide have since started to aggressively invest and build RFID technologies and products. Figure 3.7 shows a list of RFID vendors and solutions introduced in 2004.

The International Organization for Standardization asserts jurisdiction over the air interface for RFID through standards- in-development ISO 18000-1 through 18000-7. These are rep- resented in the United States by American National Standards Institute and the Federal Communications Commission. The frequencies available are shown in Table 3.5.

The Auto-ID concept is that the data will be stored on the Internet or the EPCglobal network, and the EPC stored in the tag is used as an index to locate the data. This introduces several standards as shown in the EPCglobal architecture framework [51], which is a collection of interrelated standards for hardware, software, and data interfaces, together with core services that are operated by EPCglobal and its delegates.

All the software specifications from the Auto-ID Center are written in and for Java. Java-based middleware plays

table 3.5 RFiD Frequency Ranges

RFID	*Key Applications*	*Standard*
125 kHz (LF)	Inexpensive passive RFID tags for identifying animals	ISO 18000-2
13.56 MHz (HF)	Inexpensive passive RFID tags for identifying objects; library book identification, clothes identification, etc.	ISO 14443
400 MHz (UHF)	For remote control for vehicle center locking systems	ISO 18000-7
868 MHz, 915 MHz, and 922 MHz (UHF)	For active and passive RFID for logistics in Europe, the United States, and Australia, respectively	Auto-ID Class 0 Auto-ID Class 1 ISO 18000-6
2.45 GHz (MW)	An ISM band used for active and passive RFID tags; e.g., with temperature sensors or GPS localization	ISO 18000-4
5.8 GHz (MW)	Used for long-reading range passive and active RFID tags for vehicle identification, highway toll collection	ISO 18000-5

an important and pivotal role in the implementation of the EPCglobal architecture framework, especially the applica- tion level events (ALE) and EPC information services (EPCIS). That's why middleware and software giants such as IBM, Oracle, Microsoft, and SAP all have large investments in RFID and developed complete RFID solution stacks.
The ONS (object naming service) is an authoritative direc- tory service just like the DNS (domain name service) for the Internet that routes requests for information about EPCs between a requesting party and the product manufacturer, via a variety of existing or new network- or Internet-based information resources. That's why EPCglobal has worked with VeriSign to provide such a service in addition to VeriSign's

DNS. VeriSign has operated the authoritative root directory for the EPCglobal Network since 2005. Although companies have successfully implemented internal RFID solutions that have captured efficiencies within the enterprise, the greatest promise of the EPCglobal Network is the ability to extend the benefits across trading-partner boundaries via the Internet to realize the IoT vision. It is not hard to imagine that RFID can be used in almost all industry segments and the benefits it will bring. There are many estimates of the RFID market size. IDTechEx predicts that the total market of RFID will be around

US$27 billion worldwide in 2018. The market size of China will be around US$1.7 billion in 2014 per iSuppli reports [210]. The RFID market was more than US$3 billion in 2008 in China when the issuing of RFID-based national ID cards for each citizen reached its peak.

In a contactless smart card, using NFC (near field com- munication) technologies, the chip communicates with the card reader through an induction technology similar to that of RFID. These cards require close proximity to an antenna to complete a transaction. They are often used when transac- tions must be processed quickly or hands-free, such as on mass transit systems, where a smart card (ticket) can be used without even removing it from a wallet. Figure 3.8 shows the RFID-based ticket and the ezM2M middleware-based applica- tion system the author's team built for the Beijing Olympic Games in 2008.

Mobile payment or mobile wallet is an alternative payment method that has been well adopted in many parts of Europe and Asia. Juniper Research forecasts that the combined mar- ket for all types of mobile payments is expected to reach more than $600 billion globally by 2013. RFID/NFC technologies have been used for mobile payments in China by its big three telco companies as well as China UnionPay, whose UnionPay cards can be used in 104 countries and regions around the world.

Figure 3.8 example of RFiD application.

WSN: the internet of transducers:

As defined in the first section, WSN is more for sensing and information-collecting purposes. Other networks include BSN (body sensor network [56]), VSN (visual or video sensor network [54,55]), vehicular sensor networks (V2V, V2I), underwater (acoustic) sensor networks (UW-ASN), urban/ social/participatory sensor networks, interplanetary sensor net- works, fieldbus networks (categorized as SCADA systems, the good oldies in the buildings and plants are getting wireless/ mobile capabilities and scaling up), and others.

BSN is a term used to describe the application of wear- able computing devices to enable wireless communication between several miniaturized body-sensor units and a single body central unit worn on the human body to transmit vital signs and motion readings to medical practitioners or caregivers

(Figure 3.9). Applications of BSN are expected to appearprimar- ily in the healthcare domain, especially for continuous monitor- ing and logging of vital parameters for patients suffering from chronic maladies such as diabetes, asthma, and heart attacks.

Visual sensor networks are based on several diverse research fields, including image/vision processing, communica- tion and networking, and distributed and embedded system

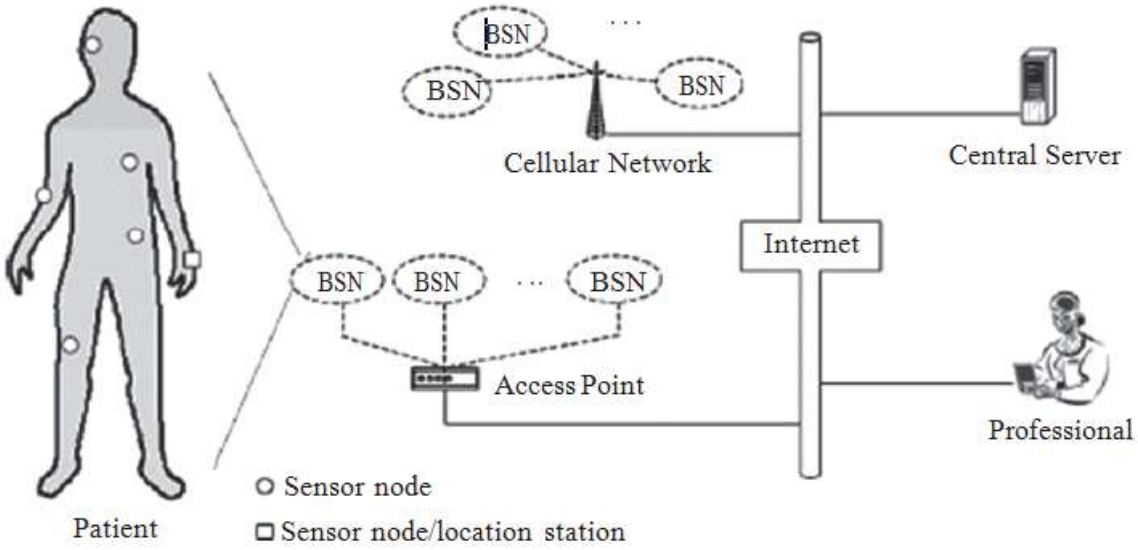


Figure 3.9 Body sensor networks. (From Hui Chen and Yang Xiao (eds.), *Mobile Telemedicine: A Computing and Networking Perspective*, new York: Auerbach Publications, 2008.)

processing. Applications include surveillance, environmental monitoring, smart homes, virtual reality, and others.

With the development of WSN, recent technological advances have led to the emergence of distributed wireless sensor and actuator networks (WSANs) that are capable of observing the physical world, processing the data, making decisions based on the observations, and performing appropri- ate actions. These networks can be an integral part of systems such as battlefield surveillance and microclimate control in buildings; nuclear, biological and chemical attack detection; home automation; and environmental monitoring.

The extended scope of WSN is the USN, or ubiquitous sensor network, a network of intelligent sensors that could one day become ubiquitous [53]. This USN is also a unified "invisible," "pervasive," or "ambient intelligent" Internet
of Things.

The development of WSNs was motivated by military appli- cations such as battlefield surveillance. The WSN is built of nodes—from a few to several hundred or even thousands—each node connected to one (or sometimes several) sensors. Each such sensor network node has typically several parts: a radio

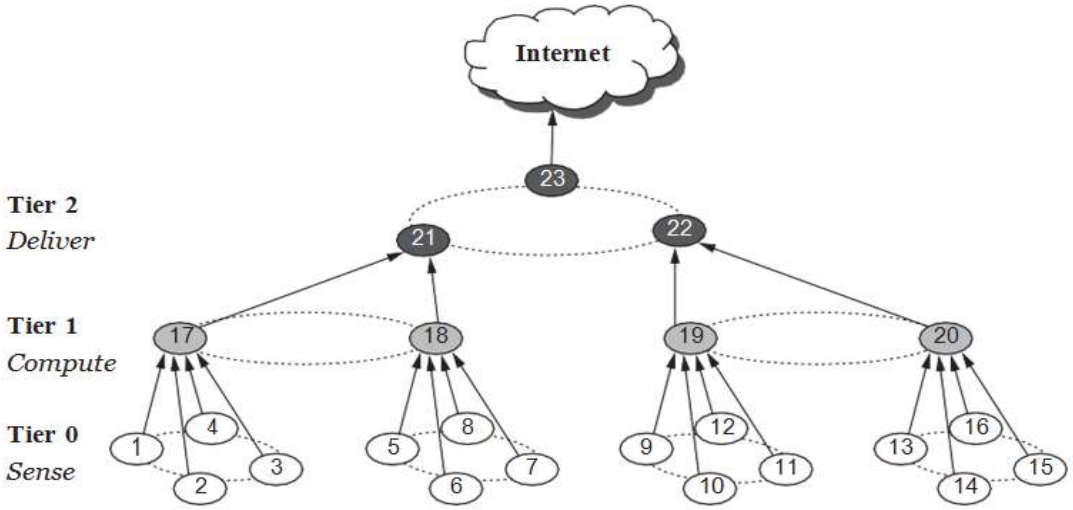


Figure 3.10 Sensor network architecture. (From Mark Yarvis and Wei Ye, "tiered Architectures in Sensor networks," in Mohammad ilyas and imad Mahgoub (eds.), *Handbook of Sensor Networks: Compact Wireless and Wired Sensing Systems*, Boca Raton, FL: CRC Press, 2004.)

transceiver with an antenna, a microcontroller, an electronic circuit for interfacing with the sensors, and an energy source, usually a battery or an embedded form of energy harvesting.

The architecture of a typical sensor network is shown in Figure 3.10. The topology of the WSNs can vary from a simple star network to an advanced multihop mesh network with a gateway sensor (sink) node connected (e.g., via a cellular M2M module) with a remote central server.

- Sensor node: sense target events, gather sensor readings, manipulate information, send them to gateway via radio link
- Base station/sink: communicate with sensor nodes and user/operator
- Operator/user: task manager, send query

Routing is required for reliable data transmission in a WSN mesh network. Routing protocols are distributed and reactive: nodes in the system start looking for a route only when they have application data to transmit. Ad hoc on-demand distance vector (AODV) and dynamic source routing (DSR) are fre- quently used routing algorithms.

The U.S. DOD, which operates the largest and most complex supply chain in the world, awarded in January 2009 a contract for $429 million in DASH7 infrastructure. This represents a major development in terms of global adoption of an ultra-low- power WSN technology based on a single global standard [72].

WSN is currently an active research area with limited mission-critical uses. IT giants such as IBM and Microsoft have invested in WSN research for a long time with little commer- cial success. Currently there is no common WSN platform.

Some designs such as Berkeley Motes and their clones have broader user and developer communities. However, many research labs and commercial companies prefer to develop and produce their own devices. Since there is no true killer application for WSNs that would drive the costs down, it is often more convenient and even less expensive to build your own WSN devices than to buy commercially available ones.

Some of the existing WSN platforms are summarized in Table 3.6. Most of the device designs are still in the research stage.

According to ID TechEx, the price per WSN node was about $30 in 2011. In the future (10 years), a functionally equivalent "smart dust" sensor node is expected to be available for use with cost per node less than $1.

Energy is the scarcest resource of WSN nodes, and it deter- mines the lifetime of WSNs. WSNs are meant to be deployed in large numbers in various environments, including remote and hostile regions, with ad hoc communications as key. For this reason, algorithms and protocols need to address the fol- lowing issues:

- Lifetime maximization
- Robustness and fault tolerance
- Self-configuration

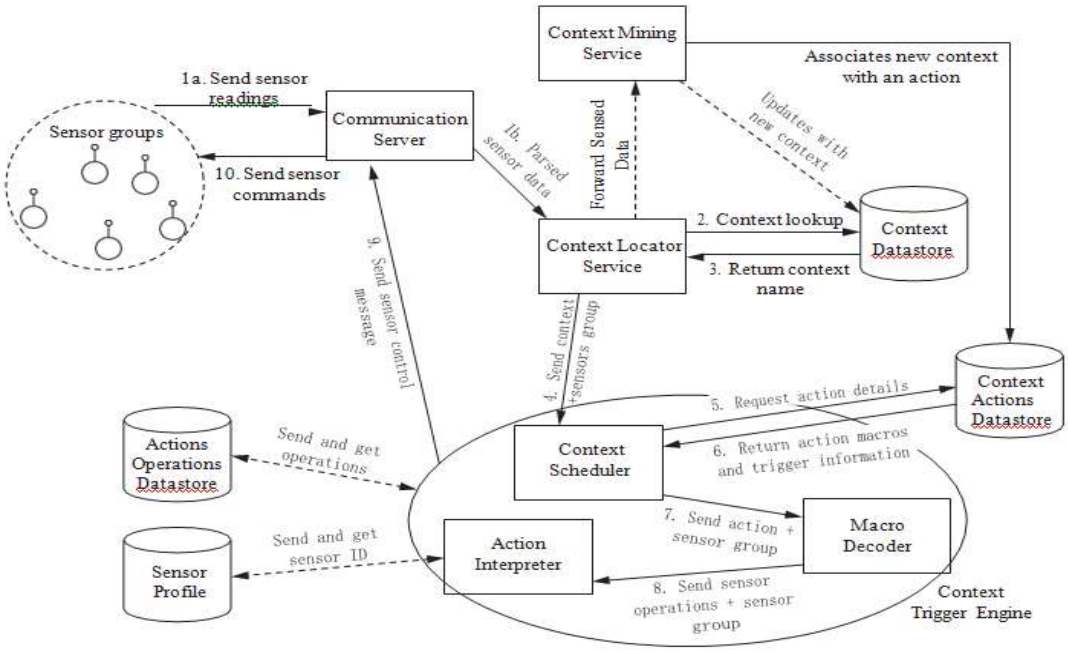


Figure 3.11 Context-aware system based on WSn. (From Seng Loke, *Context-Aware Pervasive Systems: Architectures for a New Breed of Applications*, new York: Auerbach Publications, 2007.)

high-level requirements from pervasive computing applica- tions and the underlying operation of WSNs. Middleware for WSN, the middle-level primitives between the software and the hardware, can help bridge the gap and remove impedi- ments. Middleware can help build context-aware IoT systems as shown in Figure 3.11.

Mobile sensor networks (MSNs) are WSNs in which nodes can move under their own control or under the control of the environment. Mobile networked systems combine the most advanced concepts in perception, communication, and control to create computational systems capable of interact- ing in meaningful ways with the physical environment, thus extending the individual capabilities of each network compo- nent and network user to encompass a much wider area and range of data. A key difference between a mobile WSN and a static WSN is how information is distributed over the net- work. Under static nodes, a new task or data can be flooded across the network in a very predictable way. Under mobil- ity this kind of flooding is more complex, depending on the mobility model of the nodes in the system. The proliferation of commodity smartphones that can provide location estimates using a variety of sensors—GPS, WiFi real-time locating sys- tems (RTLS), or cellular triangulation—opens up the attractive possibility of using position samples from drivers' phones to monitor traffic delays at a fine spatiotemporal granularity. MSN systems such as vTrack [58] of the MIT CarTel group have been built to monitor traffic delays and change routes.

According to IDTechEx research in the new report "Wireless Sensor Networks 2011–2021" [211], the WSN market is expected to grow rapidly from $0.45 billion in 2011 to $2 billion in 2021. These figures refer to WSN defined as wireless mesh networks, that is, self-healing and self-organizing. WSNs

will eventually enable the automatic monitoring of forest fires, avalanches, hurricanes, failure of country-wide utility equipment, traffic, hospitals, and much more over wide areas, something previously impossible. More humble killer applications already exist such as automating meter readings in buildings, and manufac- ture and process control automation.

The United States dominates (72 percent, according to IDTechEx, of all countries worldwide) the development and use of WSN partly because of the heavier funding available. The U.S. WSN industry sits astride the computer industry thanks to companies such as Microsoft and IBM, and WSN is regarded as a next wave of computing, so U.S. industry is

particularly interested in participating. Add to that the fact that the U.S. military, deeply interested in WSN, spends more than all other military forces combined, and creating and funding start-ups is particularly easy in the United States, and you can see why the United States is ahead at present.

SCADA: the internet of Controllers:

For more than a decade, many in the building industry have been envisioning a day when building automation systems (BAS) would become fully integrated with communication and human interface practices and standards widely employed for information technology systems. Not long ago, building automation graphical interfaces (shown in Figure 3.12; the part on the right is the human–machine interface the author's team built for the super-energy-efficiency building at QingHua University) employed almost no web-browser techniques and technologies; now, web approaches are the basis of many such packages. How close we are to a complete convergence of BAS and IT is difficult to tell, but it is not too much of a stretch to say that when the convergence is complete, there may be nothing to distinguish one from the other [59].

SCADA (supervisory, control and data acquisition) systems, as the core technology of the controls–IT convergence, will evolve and take the center stage. By their very nature, SCADA, low-data-rate (LDR), and M2M/IoT [129] services are closely related and largely overlapped in technologies and deployment approaches, as per GII Research [60]. Also, WSN is considered a new computing paradigm that emerged from the fusion of the SCADA systems and ad hoc networks technologies [61].

The advent of the Internet of Things will no doubt speed up the controls–IT convergence and make control systems and IT systems inseparable and indistinguishable from each other.

SCADA was generally referring to industrial control sys- tems (ICSs): computer systems that monitor and control industrial, infrastructure, or facility-based processes, as described below:

- Industrial processes include those of manufacturing, pro- duction, power generation, fabrication, and refining, and may run in continuous, batch, repetitive, or discrete modes.

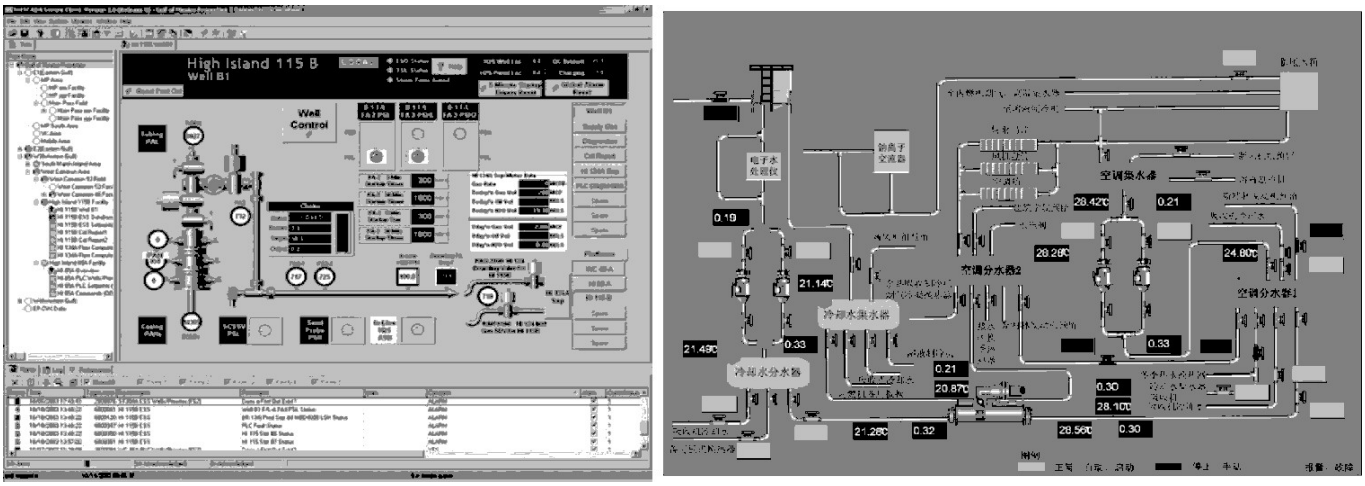


Figure 3.12 examples of SCADA graphics and animations.

- Infrastructure processes may be public or private and include water treatment and distribution, wastewater collection and treatment, oil and gas pipelines, electrical power transmission and distribution, wind farms, civil defense siren systems, and large transportation systems.
- Facility processes occur in both public and private facili-

ties, including buildings, airports, ships, and space sta- tions. They monitor and control HVAC, access, and energy consumption using PLCs (programmable logic controllers) and DCSs (distributed control systems) via the OPC (OLE for process control) middleware.

An existing SCADA system usually consists of the following subsystems (Figure 3.13):

- A human–machine interface (HMI), which is the appara- tus that presents process data to a human operator, and through this, the human operator monitors and controls the process.
- Remote terminal units (RTUs) connect to sensors in the process, convert sensor signals to digital data, and send digital data to the supervisory system.
- PLCs are used as field devices because they are more eco-

nomical, versatile, flexible, and configurable than special- purpose RTUs.

- DCSs; as communication infrastructures with higher capacity become available, the difference between SCADA and DCS will fade. SCADA is combining the traditional DCS and SCADA.
- As mentioned before, M2M (telemetry), WSN, smart systems, CPS, and others all have overlaps of scope with SCADA, but the extended scope of SCADA is bigger under the IoT umbrella.

A SCADA system could be a layer between the top-layer business systems such as ERP, WMS (warehouse management

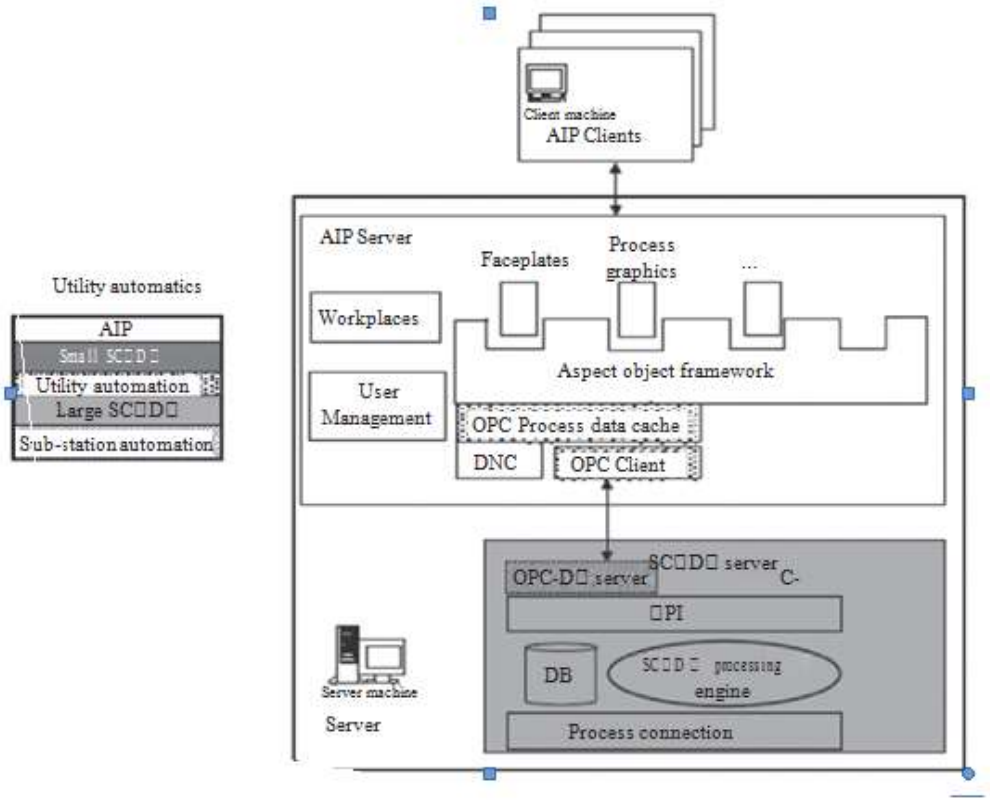


Figure 3.13 Components of a SCADA system. (From Yauheni Veryha and Peter Bort, "industrial it-Based network Management," in Richard Zurawski (ed.), *The Industrial Information Technology Handbook*, Boca Raton, FL: CRC Press, 2005.)

system), SCM, CRM, EAM (enterprise asset management), PIMS (plant information management system), EMI (enterprise man- ufacturing intelligence), LIMS (laboratory information manage- ment system), and other applications and the lower layer DCS, PLC, RTU, MES (manufacturing execution

system), SIS (super- visory information system in plant level), and other systems as exemplified in Figure 3.14.

A traditional SCADA system is a client/server system. New technological developments have turned C/S SCADA systems into middleware-backed, web-based, three-tiered open sys- tems with SOA capabilities.

Figure 3.15 showcases a typical SCADA middleware or platform architecture. Examples of such platforms include

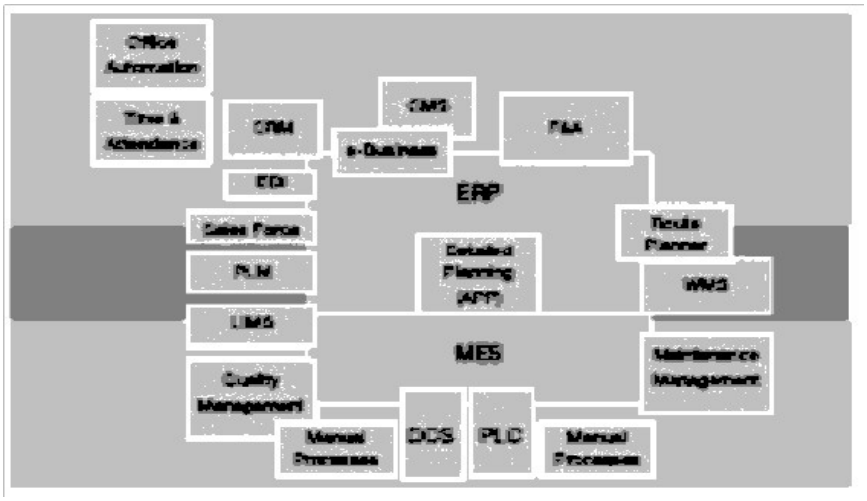


Figure 3.14 SCADA sits in the center.

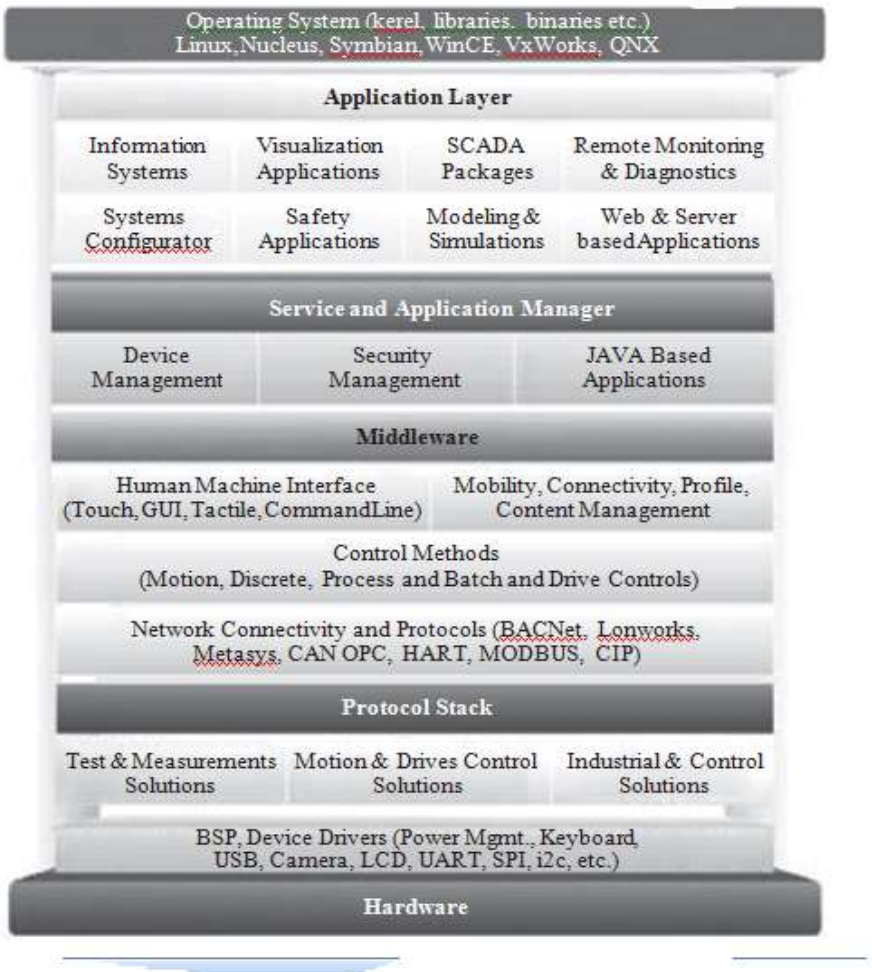


Figure 3.15 Middleware-based SCADA systems

(Invensys) Wonderware's ArchestrA TM, (Honeywell) Tridium's Niagara Framework TM (a Java EE–based platform), THTF's ezM2M Middleware for IoT, various implementations of the OPC UA framework standard, and the list goes on.

SCADA systems allow the automation of complex industrial processes where human control is impractical. However, with all the raw data and real-time updates pouring in, it can be difficult to decipher what is going on and how to respond. All the on-screen numbers, flashing lights, and blaring alarms still leave you in the dark. The solution is an integrated controls–IT convergence system

IP video technology has become one of the hottest trends in the automation industry today, especially since automation and surveillance systems have both migrated to IP-based applications. Moreover, the integration of IP surveillance software with automation systems is gaining popularity and momentum, and integrating real-time visual

surveillance systems [100] with SCADA systems via IP video technology is now both a viable and an affordable solution for system integrators.

Many industries are using SCADA as a core technology to link the geographically separated facilities and support new business processes in response to changing industry dynamics.
As examples, the worldwide oil and gas industry SCADA market was $850 million in 2007 and is forecast to be over
$1.3 billion in 2012; the worldwide market for electric power SCADA was $1.629 trillion in 2008 and is forecast to be over
$2.125 trillion in 2013; and the worldwide water and waste- water industry SCADA market was $212 million in 2006 and is forecast to be over $275 million in 2011, all according to ARC Advisory Group studies.
In 2010, Chinese government and industry leaders stated that a "unified strong and smart grid" [166] system is going to be built across the country by 2020. SCADA sales will increase as part of this initiative and overall IoT development.

Supported by intelligent field devices, expanded com- munications networks, and improved compatibility with IT, especially the Internet and web technologies, SCADA can now provide a wealth of information and knowledge as a means to modify business processes and enable the creation of new SCADA-based IoT applications.

- ## DCM: Device, Connect, and Manage

The first issue that the Internet of Things (IoT) ecosystem needs to address is the long and fragmented value chain that characterizes the industry. This results in numerous supplier– buyer interfaces, adding costs and time to the launch of any new product offering.
Just like the blind men and the elephant story and people's understanding of the four pillars or the six pillars mentioned before, the IoT is still different things to different people, even though introduced more than a decade ago. However, there
is one thing most people agree with: IoT (or machine-to- machine, M2M; wireless sensor networks, WSN; supervisory control and data acquisition, SCADA; radio-frequency identi- fication, RFID; etc.) systems all have three layers. Figure 4.1 is an example IoT application of an intelligent nuclear power plant IoT system [63] of Datang Telcom in China. More exam- ples of the three-layer architecture of IoT can be found at European Telecommunications Standards Institute (ETSI)'s website .

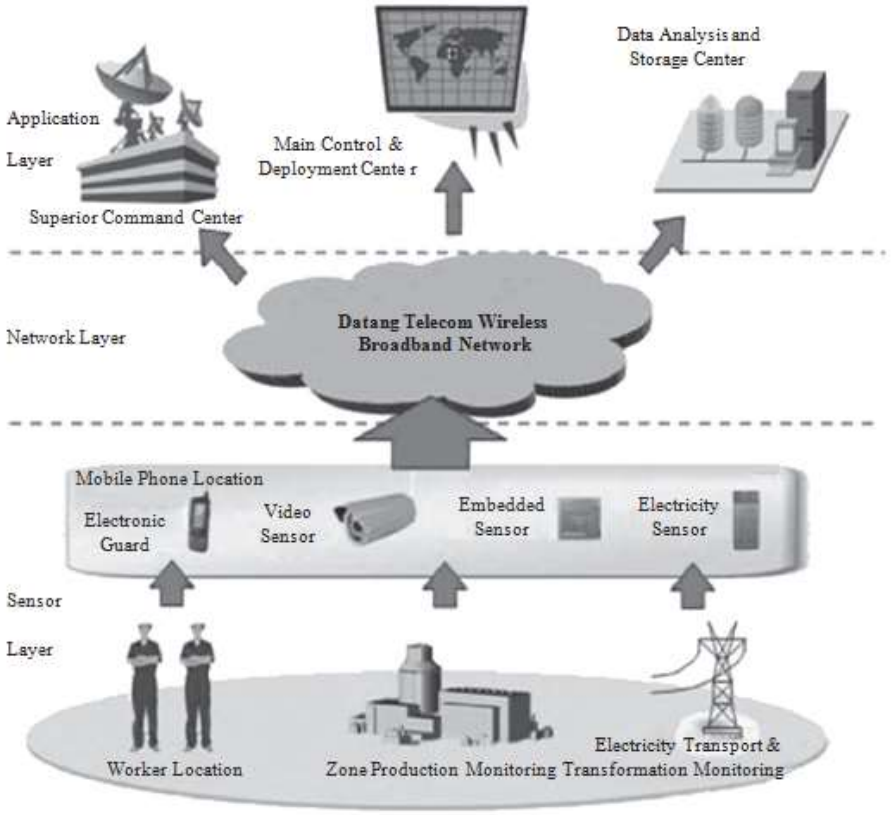


Figure 4.1 examples of three-layer architecture of iot.

The author has proposed the concept and acronym DCM (device, connect, and manage [74]) as a corporate stretegy or slogan for TongFang Co. Ltd. The board of the company announced financing of 500 million Chinese renminbi (RMB) (or US$78.5 million) for the development of the IoT/DCM business in 2005. Numerex created a better acronym called DNA^{TM} (devices, networks, and applications) [213] in 2008 (Figure 4.2).
The three-layer DCM classification is more about the IoT value chain than its system architecture at runtime. For system architecture, some (e.g., one of Numerex's and IBM's reports) have divided the IoT system into as many as nine layers, from bottom to top: devices, connectivity, data collection,

Figure 4.2 DCM (DnA) of iot.

communication, device management, data rules, administra- tion, applications, and integration. While large companies such as IBM, Oracle, Microsoft, and others have comprehensive solutions, products, and services that cover almost the entire value chain, startups or smaller players in the IoT sector should focus on providing products or services in no more than two components or areas in the value chain.
The following sections discuss the three DCM components.

Device: things that talk:

According to the IoT definitions and descriptions in the pre- vious chapters, devices or assets can be categorized as two groups: those that have *inherent intelligence* such as electric meters or heating, ventilation, and air-conditioning (HVAC) controllers, and those that are inert and must be *enabled* to become smart devices (e.g., RFID tagged) such as furniture or animals that can be electronically tracked and monitored— things that "talk."
Just as Paul Saffo [214], a technological forecaster and strate- gist, described in an interview in 2002:

This is the Cambrian explosion of communica- tions. We are seeing a radical species divergence of different kinds of devices and different types of things that want to talk, from your washing machine having an Internet connection and being able to scream for help if it is broken, to your car having a wireless connection for data telemetry back to the manufacturer. Today, voice communications is way below 1% of the total communications traffic on this planet. That's why people are giving voice away for free. So that means that we're going to see a whole zoo of new kinds of devices that have to talk. It's going to become a world of smartifacts, or intelligent objects. This stuff is so cheap, we're putting chips in everything, anything with a chip inside can be con- nected into the Internet of Things.
Devices that perform an input function are commonly called sensors because they "sense" a physical change in some characteristic that changes in response to some excitation,
for example, heat or force, and convert that into an electrical signal. Devices that perform an output function are generally called actuators and are used to control some external device, for example, movement. Both sensors and actuators are collec- tively known as transducers because they are used to convert energy of one kind into energy of another kind. For example, a microphone (input device) converts sound waves into electrical signals for the amplifier to amplify, and a loudspeaker (output device) converts the electrical signals back into sound waves.

A *sensor* (also called a detector) is a device that responds to a physical stimulus, measures the physical stimulus quantity, and converts it into a signal, usually electrical, which can be read by an observer or by an instrument.

Based on this definition, a sensor is basically an electrical device. It could be an M2M terminal, an RFID reader, or a SCADA meter. Sensors are particularly useful for making in-situ measurements (things that talk) such as in industrial process control or medical applications. A sensor can be very small and itself can be a tractable device; however, when a train or an aircraft is instrumented with a small sensor, the entire aircraft becomes one tractable device.

The sensor itself, if not connected, is not part of the IoT or WSN value chain. This is like a central processing unit (CPU), which is not part of the web or social networking services, even though they are somewhat related. Some sensors do not generate electrical signals; for example, a mercury-in-glass thermometer converts the measured temperature into expan- sion and contraction of a liquid, which can be read on a cali- brated glass tube. However, it's important to understand the types and shapes of the ubiquitous sensors if you are into IoT, just as an architect should know what concrete and cement are as well as their differences. Figure 4.3 showcases a few sample sensors.

Some of the existing sensors and their types are listed in Table 4.1. The size of the overall sensor market is difficult to estimate. A number of research reports on the market size of

Figure 4.3 examples of sensors.

different sensor sectors are on http://www.sensorsportal.com. For example, the global automotive sensor market, including silicon-based sensors, grew by 9.7 percent in 2006 to $10.1 bil- lion and is forecast by Strategy Analytics to reach $17.1 billion by 2013 as vehicle systems such as powertrain control, safety, and convenience features become more advanced and require more sensors. IC Insights estimates that the wireless sensors and transmitters market will surpass $1.8 billion by 2012. The CMOS image sensor market alone is projected to be $8.3 bil- lion by 2014.

Microelectromechanical systems (MEMS) is the technology of very small mechanical devices driven by electricity. It merges at the nanoscale into nanoelectromechanical systems (NEMS) and nanotechnology. MEMS are also referred to as micromachines in Japan, or microsystems technology in Europe. MEMS can be a sensor or actuator, or a transducer.
Energy harvesting (also known as power harvesting or energy scavenging) is the process by which energy is derived from external sources (e.g., solar power, thermal energy, wind energy, salinity gradients, and kinetic energy), captured, and stored for small wireless autonomous devices, like those used in wear- able electronics and WSNs. Energy-harvesting devices or sen- sors have a very long historical connection to the water wheel, windmills, and waste heat. Before batteries (Volta, 1799) and the dynamo (Faraday, 1831), those energy-harvesting devices were the only ways to get any useful power. The following are options for energy harvesting:

- RF, used for RFID tag energy broadcasting and harvesting
- Solar, a well-known clean energy
- Thermoelectric, used in watches
- Vibrations, used in (kinetic) watches
- Human input, home utility (piezoelectric) switches

Today, there is an accelerated interest in the information and communications technology (ICT) community for powering ubiquitously deployed sensor networks, mobile electronics, electric vehicles, and so on. Many things become possible as this technolgy improves.

• Connect: Via Pervasive networks:

The communications layer is the foundational infrastructure of IoT. There are two major communication technologies: wireless and wired (or wireline). Each category has broadband and narrowband, packet and circuit switched, as well as short-range and long-range communications. The penetration and traffic of U.S. wireless data subscribers in 2013 will reach the same level of broadband wired household usage in 2008 [215]. The mobile Internet is catching up quickly, thanks to the development of the Internet of Things and the flexibility of wireless communications.
Today's communications environment is a complex mix of wired and wireless networks employing circuit-switched (CS) and packet-switched (PS) technology. Developments are taking place in all four

sectors and there is competition between dif- ferent stakeholders, fixed mobile convergence (FMC) being an obvious example. We therefore have a communications environ- ment that is complex [64]. We need a next-generation network (NGN), which has more than the ability to transition between circuit- and packet-switched networks. The general idea behind the NGN is that one network transports all information and services (voice, data, and all sorts of media such as video) by encapsulating these into packets, similar to those used on the Internet. NGNs are commonly built around Internet protocol, and therefore the term all-IP is also sometimes used to describe the transformation toward NGN. For example, the 3GPP long- term evolution (LTE) is a standard for wireless communication of high-speed data. It is based upon GSM/EDGE and UMTS/ HSPA network technologies. One of the most important features of LTE is that it will be an all-IP flat network architecture includ- ing end-to-end QoS, provisions for low-latency communications.

With the growing abundance of embedded IoT systems comes
the increased pressure at the edge of the network: multiple access methods must be accommodated, implying the need for a com- mon underlying converged core IP/MPLS (multi-protocol label switching) network. A high-level graphic view of next-generation all-IP networking is described by Emmerson [64]. The connectiv- ity domain enables broadband access, both wired and wireless.

It also includes the transport and aggregation network. This part of the all-IP network supports various access technologies using copper lines, optical fiber, and air as transmission media.

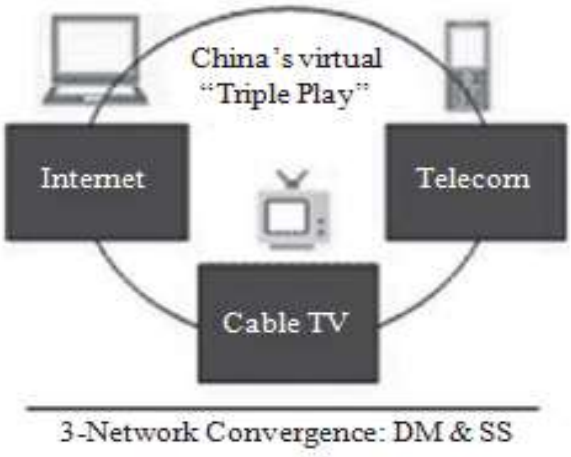


Figure 4.4 triple network convergence.

The Chinese government has been actively pushing for the convergence of the country's three big networks—the Internet, telecom networks, and TV broadcasting networks—via

various measures, most notably through the Triple Network Convergence Plan (Figure 4.4) it laid out early in 2010.

While the Triple Network Convergence Plan reiterates many government policies set out previously, one area that is expected to have significant effects on the market is the government's step to grant permission for TV broadcasting firms and telecom carriers to enter and do business in each other's realms. Local scholars estimate that triple network convergence will induce investment and consumption to 700 billion RMB (about US$103 billion), leading to widespread concern over the policy's effect on the development of related industries and various parties.

The fusion of the three networks is expected to start from business- or policy-level convergence, to application-level con- vergence, and finally to technological-level convergence, when the all-IP NGN vision is implemented. At that time, many good things will happen; for example, ubiquitous M2M devices can be used as cell phones, so no SIM card will be required for making a phone call.

There is no doubt that if all-IP is a reality, it will give the Internet of Things a huge lift and make the IoT dream come true much easier and faster. As an example, in the building

automation industry, all-IP networking will simplify the inte- gration work enormously, without having to deal with various field bus network protocols, OLE for process control (OPC) middleware, and so on.

Internet Protocol version 6 (IPv6) is a version of the Internet protocol that is designed to succeed Internet Protocol version 4 (IPv4). The Internet operates by transferring data in small packets that are independently routed across networks as specified by the Internet protocol. Since 1981, IPv4 has been the publicly used IP, and it is currently the foundation for most Internet communications. The Internet's growth has created a need for more addresses than IPv4 has (32 bits).

IPv6 allows for vastly more numerical addresses (128 bits), but switching from IPv4 to IPv6 may be a difficult process [216].

The Internet world is getting ready for the big change from IPv4 to IPv6. After the change, everything, every duct on the planet, could have a fixed IP address, which would have an enor- mously huge impact on the Internet of Things on all aspects.

However, as a side note, countries such as the United States are not eager to make the change from IPv4 to IPv6 compared with countries such as China and India, because more IPv4 addresses were allocated to the United States and Europe. It's rumored that a university such as Massachusetts Institute of Technology received more IPv4 address allocation than the entire country of China or India. That's why countries such

as China have developed other protocols such as IPv9 in an effort to get more IP addresses [65].

When talking about IoT, wireless communications is the topic most of the times, because three (M2M, RFID, and WSN) of the four IoT pillars are based on wireless. However, most of the systems in industrial automation, building automation, and so forth are built using SCADA technology on wired short-range field bus and long-range TCP/IP networks. The development of the Internet of Things, for the time being, should cover both wired and wireless networks, just as Axeda, the device relation management software product and service

provider, did in its product and service portfolio before or after the all-IP convergence and IPv6.

- **Wired Networks:**

Wired networks for IoT can be categorized as short-range field bus–based access networks, mostly for SCADA applications, and IP-based networks, for M2M and SCADA applications.
The IP-based networks are widely used and their protocol stack is well known, as shown in Figure 4.5, together with telephony SS7 and cable TV DOCSIS (data-over-cable service interface specification) protocols, the triple (Internet, tele- phony, and cable TV) networks convergence plan candidates. SS7 (Signaling System 7) is a critical component of modern telecommunications systems (PSTN, xDSL, GPRS, etc.). Every call in every network is dependent on SS7. Likewise, every mobile phone user is dependent on SS7 to allow inter-network roaming. SS7, a form of packet switching, is also the "glue" that sticks together circuit-switched (traditional) networks with Internet protocol–based networks.
DOCSIS is an international standard that permits the addi- tion of high-speed data transfer to an existing cable TV

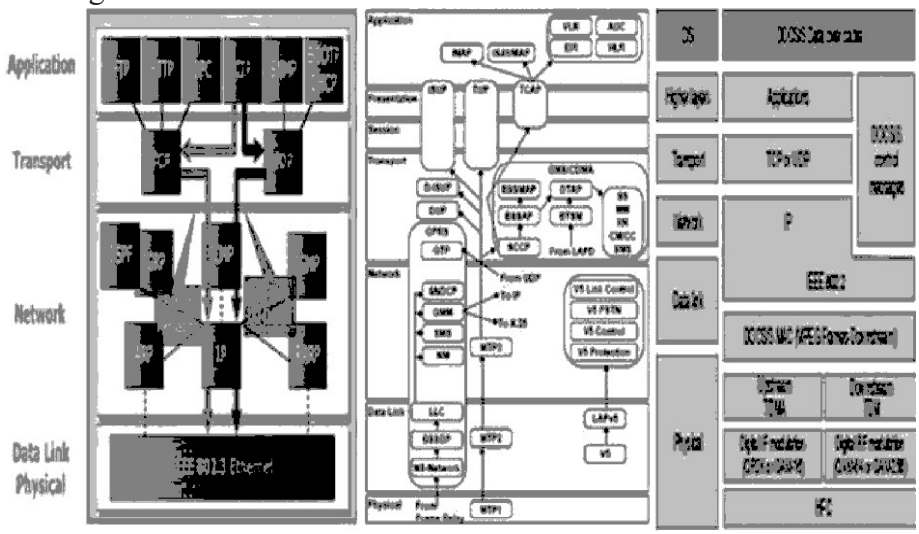


Figure 4.5 Protocol stacks of the "three networks."

system. It is employed by many cable television operators to provide Internet access over their existing HFC (hybrid fiber- coaxial) infrastructure.
A complex automated industrial system, such as a manu- facturing assembly line, usually needs an organized hierarchy of controller systems to function. In this hierarchy [217,218], there is usually a

SCADA/HMI (Human–Machine Interface) at the top, where an operator can monitor or operate the system. This is typically linked to a middle layer of programmable logic controllers (PLC) via a non-time-critical communications system (e.g., Ethernet). At the bottom of the control chain is the field bus (could run on top of a different power line communi- cations network too) that links the PLCs to the IoT device com- ponents that actually do the work, such as sensors, actuators, electric motors, console lights, switches, valves, and contactors.

More details on field bus and its relevance to IoT are described here because this information is currently often neglected in most of the materials about IoT. Field bus is the name of a fam- ily of industrial computer network protocols used for real-time distributed control, now standardized as IEC 61158. The IEC 61158 standard includes eight different protocol sets called types:

- Type 1 Foundation field bus H1
- Type 2 ControlNet
- Type 3 PROFIBUS
- Type 4 P-Net
- Type 5 FOUNDATION field bus HSE (high-speed Ethernet)
- Type 6 SwiftNet (a protocol developed for Boeing, since withdrawn)
- Type 7 WorldFIP
- Type 8 Interbus

There is a wide variety of concurring standards. Table 4.2 provides a comprehensive list of wired field bus standards or protocols used with SCADA systems for industrial automation.

- ***Wireless Networks***

Just like the wired networks, wireless networks for IoT can be categorized as follows:

- Short-range (including near field communication [NFC], usually narrowband, and wireless PAN, LAN, and MAN) mesh networks, RFID, WiFi, WiMax, and so on;
- Long-range (via cellular networks, wireless WAN, pseudo-long-range) GSM, CDMA, WCDMA, and other networks, as well as satellite communication.

Short-range wireless mesh networks are the fundamental com- munication techniques of WSN and RFID. Long-range cellular networks are the foundation networks for M2M.

Radio spectrum refers to the part of the electromagnetic spectrum corresponding to radio frequencies: lower than 300 GHz (or wavelengths longer than about 1 mm). Different parts of the radio spectrum are used for different applications. The so-called sweet spot at ultra-high frequency concentrated most of the widely used frequencies. Radio spectrum are typically government regulated, and in some cases, are sold or licensed to operators of

private radio transmission systems, for example, cellular telephone operators or broadcast television stations.

There are as many wireless standards as wired network protocols (Table 4.2). Before 2000, there were about five or six concurring standards, which lasted for a longer time than today's standard. Nowadays, there are more than 15 concur- ring wireless standards [220] and new ones keep coming, with each and every one's life span shorter than those before.

Wireless communications standards can also be categorized as standards for cellular communications networks (such as GSM, CDMA, HSPA, LTE, etc.) and wireless connectivity networks (such as Bluetooth, Wifi, WiMax).

Communications standards are evolving rapidly. With the advent of the Internet of Things, it is expected that new stan- dards will appear with even higher frequency and in larger numbers, due to requirements on wireless network improve- ments for machine-type communications (MTC) [66,189]. MTC is expected to be one of the major drivers of wireless commu- nications standards in the next decade. The ETSI now has a technical committee exclusively focused on M2M; the Chinese Communications Standards Association is currently exploring the definition of M2M standards for China; and the Geneva- headquartered International Telecommunications Union (ITU) is working on "mobile wireless access systems providing

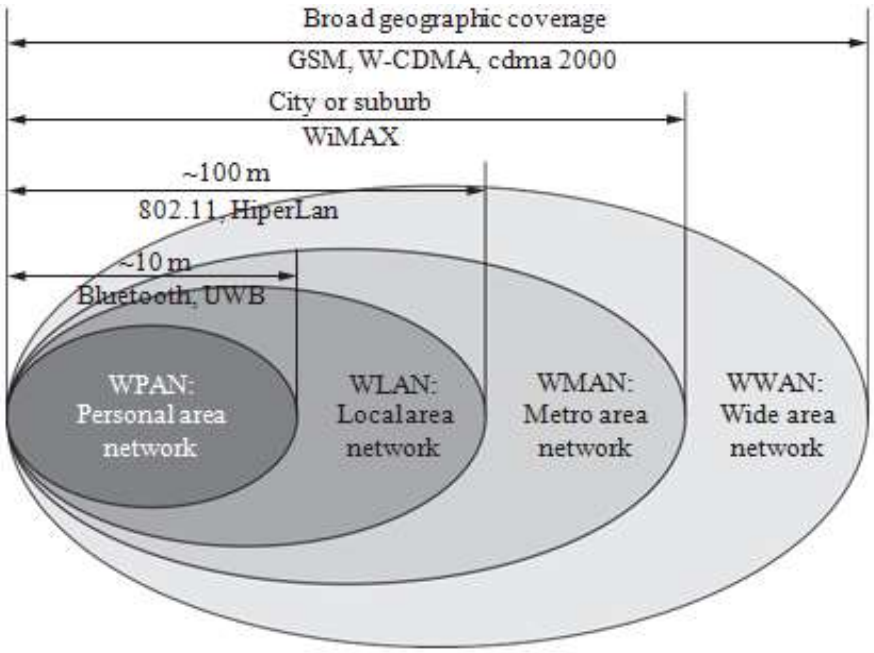


Figure 4.6 Short- and long-range wireless networks.

telecommunications for a large number of ubiquitous sen- sors or actuators scattered over wide areas in the land mobile service," which are at the center of the M2M ecosystem. The

U.S. Telecommunications Industry Association (TIA) has also launched a new engineering committee centered on smart device communications (TIA TR-50).
Figure 4.6 shows the spectrum of wireless communications standards from short-range to long-range. RFID and NFC are parts of WPAN.
Short-range wireless sensor networks can also be treated as access networks [221] to IP-based Internet for many vertical applications such as building automation and others.
Wireless communications can be via RF, microwave (long- range line of sight via highly directional antennas, or short- range), or infrared (short-range, consumer IR devices such as remote controls). Some of the standards that have not been discussed previously are as follows:

- 6LowPAN (IPv6 over low power wireless personal area networks): a working group of IETF
- BSN (body sensor network): IEEE 802.15.6
- Broadband fixed access: LMDS, AIDAAS, HiperMAN
- DASH7: active RFID standard
- DECT (digital enhanced cordless telecommunications): cordless telephony
- EnOcean: low-power, typically battery-less, proprietary wireless technology
- HomeIR: wireless IR home networking
- HomeRF: wireless RF home networking
- IEEE 1451: a set of smart transducer interface standards by the IEEE
- InfiNET: from home automation industry leader Crestron
- INSTEON: dual-mesh technology from SmartLabs
- IrDA: from Infrared Data Association
- ISA100.11a: an open wireless networking technology standard developed by the International Society of Automation (ISA)
- Land mobile radio or professional mobile radio: TETRA, P25, OpenSky, EDACS, DMR, dPMR, etc.
- ONE-NET: open-source standard for wireless networking
- OSIAN: open-source IPv6 automation network
- TransferJet: a new type of close-proximity wireless transfer technology by touching (or bringing very close together) two electronic devices; allows high-speed exchange of data
- Wavenis: a proprietary technology by Coronis Systems in 2001. In 2008, the Wavenis Open Standard Alliance Wavenis-OSA was created to manage and govern the technology moving forward.

Apart from new standards emerging from MTC improve- ments, other new technologies and standards can also help in advancing the Internet of Things:

- Orthogonal frequency division multiplexing (OFDM) and orthogonal frequency division multiple access (OFDMA): two different variants of the same broadband wireless air interface. LTE is an OFDMA-based technol- ogy standardized in 3GPP. OFDM technologies typically occupy nomadic, fixed, and one-way transmission stan- dards, ranging from TV transmission to Wi-Fi as well as fixed WiMAX and newer multicast wireless systems like Qualcomm's Forward Link Only.
- Ad hoc sensor network: a short-lived network of two or more mobile devices connected to each other without the help of intervening infrastructure. In contrast to a fixed wireless network, an ad hoc network can be deployed in remote geographical locations and requires minimal setup and administration costs. The integration of an ad hoc network with a bigger network such as the Internet or a wireless infrastructure network increases the coverage area and application domain of the ad hoc network.
- Software defined radio (SDR): SDR is the result of an evolutionary process from purely hardware-based equip- ment to fully software-based equipment. All functions, modes, and applications, such as transmit frequencies, modulation type, and other RF parameters, can be con- figured and reconfigured by software (SW) defines all waveform properties, cryptography, and applications, is reprogrammable, and may be upgraded in the field with new capabilities;
- Cognitive radio (CR): CR is a form of wireless communica- tion in which a transceiver can intelligently detect which communication channels are in use and which are not, and instantly move into vacant channels while avoiding occupied ones. This optimizes the use of available RF spec- trum while minimizing interference to other users. SDR is a required basic platform on which to build a CR. SDR and CR extend the software and middleware capabilities a step further into the communicating devices and increase the ubiquity, versatility, and smartness of devices in the Internet of Things.

- **Satellite IoT:**

A communications satellite (COMSAT) is a specialized wireless transponder in space, receiving radio waves from one location and transmitting them to another (also known as a *bent pipe*). Hundreds of commercial satellites are in operation around the world. These satellites are used for such diverse purposes as wide-area network communications (to ships, vehicles, planes, as well as hand-held terminals and phones), weather forecast- ing, television and radio broadcasting, amateur radio commu- nications, Internet access, and the global positioning system (GPS). Satellites have many important uses other than commu- nications; for example, weather reports rely on satellite infor- mation, and GPS works because of a linked set of satellites.

Satellite communications are especially important for transpor- tation, aviation, maritime, and military use.
Modern communications satellites use a variety of orbits:

- GEO: Geostationary Earth Orbit, 120 satellites maximum, examples include Inmarsat (4 + 5 Satellites)
- MEO: Medium Earth Orbit, examples include the GPS satellite constellations
- LEO: Low (polar and nonpolar) Earth Orbit (theoretically unlimited); examples are Iridium (66 satellites; rent for global Iridium satellite phones is as low as $24.95 per week shown on the company's website), ORBCOMM (30 satellites), Globalstar (48), ICO (10+2), Ellips0 (17), Teledesic (288 satellites); constellations of satellites required for coverage
- ELI: Elliptical Orbit
- Molniya Orbit and HAPs (high-altitude platforms)

The satellite industry is a subset of the telecommunications and space industries. According to a SIA (Satellite Industry Association) report [67], the worldwide revenue of the satellite industry was $168.1 billion in 2010.
It's obvious that satellite technologies (other than positioning- oriented global navigation satellite system or GPS, which will be discussed in Chapter 6 of the book) can be used for IoT applications (such as M2M, SCADA, and telemetry) just like cel- lular networks, with better coverage in remote areas. When people think of M2M communication, they usually think of cellular networks. For vehicles that move in urban areas or on major highways, cellular coverage is usually good enough, but what about construction equipment at remote locations, agricultural equipment, or ships? That's where satel- lite communication comes into play.

• Manage: to Create new Business Value

The previously described first two stages of the DCM model show the processes and venues of how the information is cap- tured from various types of devices and how this information is aggregated via various gateways and transported across access networks and the core backbone to the central servers. The machine-generated information comes in large volumes much bigger and faster than information generated by humans; how- ever, much of the data are of low value or even noises, which must be filtered out by middleware at the edge as described before in the RFID sections. And then those preprocessed
data are transformed into high-value information via a cogni- tive application platform, most of the times a high-performance cloud computing (or high-throughput computing) platform.

In the current customer-driven, technology-based envi- ronment, it is no longer enough to offer a service or product and expect it to satisfy your customers. Even if you have the best customer service in the industry, you have to be able to extend out your offerings to meet current demand to keep the customers satisfied. The Internet of Things brings enormous possibilities and potentials for creating new business value and generating new revenue ecosystems with data processing and managing rules that combine intelligence from remote assets unreachable before with your intelligent enterprise systems. With IoT, more and more areas of the real world become part of the ICT world, as shown in http://consen.org/node/9 from the IoSS (Internet Architecture for Optimization Sensing Systems) project in Europe. Disruptive applications beyond current imagination will appear. Smart grid, connected car, fleet control, mobile surveillance, and remote monitoring are listed as the top five disruptive applications out of a total of 65 identified, according to reports from the Boston Consulting Group. All of the top five are IoT applications. For example, with the wide use of telematics, things like total vehicle life cycle management, refined used car price estimate, Pay as

Figure 4.7 iPhone M2M application.

You Drive insurance policy, neighbor-to-neighbor car-sharing business such as those provided by startup RelayRides become possible, and the list goes on and on.

Let's take a look again at the typical capabilities of an M2M platform and how they support the business of a mobile operator or an M2M enabler/partner. With those functions and roles (as shown in http://machine2twomachine.files .wordpress.com/2011/08/fig-16.jpg[265]), both the mobile oper- ator and the M2M partner can attain additional revenue by offering advanced services to their M2M partners (Figure 4.7).

For example, the M2M platform and the fleet management sys- tem the author's team built for China Mobile utilize its existing Operation Support System/Business Support System (OSS/BSS) for SIM card issuing, billing, and other services, and China Mobile collects the revenue from the customers and shares it with us. China Unicom has also built and operates a telemat- ics service support platform on top of their OSS/BSS, aiming to provide foundation services to a variety of TSPs (telematics service providers).

M2M applications that can be linked inside the network to people's existing mobile subscriptions offer mobile operators enormous advantages in the competitive M2M marketplace. Using smartphones as connected portable navigation devices is such an example of potentially great market growth oppor- tunity. The application stores' model of Apple and Google Android has turned smartphones into M2M terminals. One example is the application from Portman Electronics Ltd.'s IES iPhone M2M Tracking System. It is a real time GPS/GSM/GPRS tracking service. Another example of nonoperator vendor is SeeControl, who empowers you to use sensors, GPS trackers, barcode scanners, RFID, and smart web forms to collect asset data from anywhere and manage business processes.

In the industrial automation scenario, the layering of the value chain components or subsystems looks as depicted in Figure 4.8. One of the major issues has been or still is that

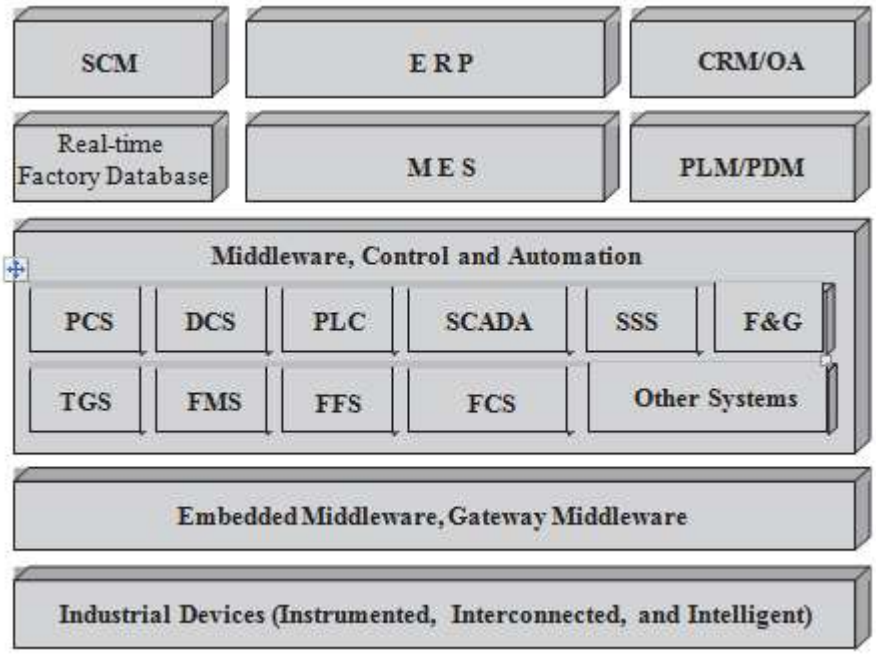


Figure 4.8 the industrial automation stack. FCS = Field Bus Control System; DCS = Distributed Control System; PLC = Programmable Logic Controller; SCADA = Supervisory Control and Data Acquisition; tMS = tank Management System; FMS = Flow Metering System;
F&G = Fire and Gas; SSS = Safety Shutdown System; FFS = Firefighting System; MeS = Manufacturing execution System; eRP = enterprise Resource Planning.

most of these subsystems are not integrated; operators have to deal with various subsystem interfaces to run the operation.
Sometimes, the factory database has to be manually keyed in to the IT database. When you want to expand and/or integrate the entire plant, you will need a solution provider with the expertise to provide the solution for you.
Before the advent of IoT or perhaps at the same time, people found out that efficient plant operations require the total integration of the field devices to the subsystems, then the integration of subsystems into a single centralized SCADA system that provides a single user interface or HMI. This is also where the new IoT system fits and sits. On top of this, those subsystems are further integrated into the MES and ERP as well as SCM, WMS, and other systems. All of those happen within an enterprise, it's an Intranet of Things ecosystem.
The vision of IoT augmented with advances in software technologies and methodologies such as SOA (service-oriented architecture), SaaS (software as a service), cloud computing, and others is causing a paradigm shift where devices can offer more advanced access to their functionality and business intel- ligence. As such, event-based information can be acquired, and then processed on-device and in-network. This capabil- ity provides new ground for approaches that can be more dynamic and highly sophisticated and that can take advantage of the available context. Cross-layer collaboration is expected to be a key issue in such a highly dynamic and heteroge- neous infrastructure such as the Real World Internet (RWI) or IoT [68]. Device relation management and intelligent device management are some of those cross-layer M2M paradigms or product concepts proposed by Axeda and Questra a few years ago, and now those products and services are serving more and more customers.
As mentioned before, the three layers of DCM are not the run-time architecture of an IoT system, but a gross classifica- tion of the IoT value chain. For an MNO or network operator

in general, IoT system architecture consists of the following seven layers [70], and the focus is on network infrastructure and service capabilities similar to those provided by telcos' existing Business and Operations Support System (BOSS).

- M2M applications
- M2M service capabilities
- Core network
- Access network
- M2M gateway
- M2M LAN
- M2M devices

For other parties in the IoT value chain, the diagrams from ETSI and Digi International [69] demonstrated more generic IoT system architectures. The key is to have a single common platform that can be used for all kinds of vertical applications of IoT. The data-collecting layer of IoT, from the last mile WSNs and the

gateway, to the access networks, and finally to the core network, can be distributed and replicated (and, of course, there may be cross-layer connections and Intranet of Things systems which are treated as subsystems). However, the layers above the core network should be highly integrated and centralized on top of a single common (platform as a service) PaaS + SaaS plat- form agnostic of and accommodating the variations of the con- nectivity including the IaaS (infrastructure as a service) layers.

The discussion of the PaaS + SaaS middleware layer is the focus of this book, which will be covered in more detail in the followed chapters.

In China, companies such as Datang, ZTE, and Huawei have also done extensive research on IoT/M2M because the Internet of Things is highly visible in the Chinese government and many grants have been allocated to sponsor such research activities. The sample architecture diagrams in Figure 4.9 are from Datang and ZTE.

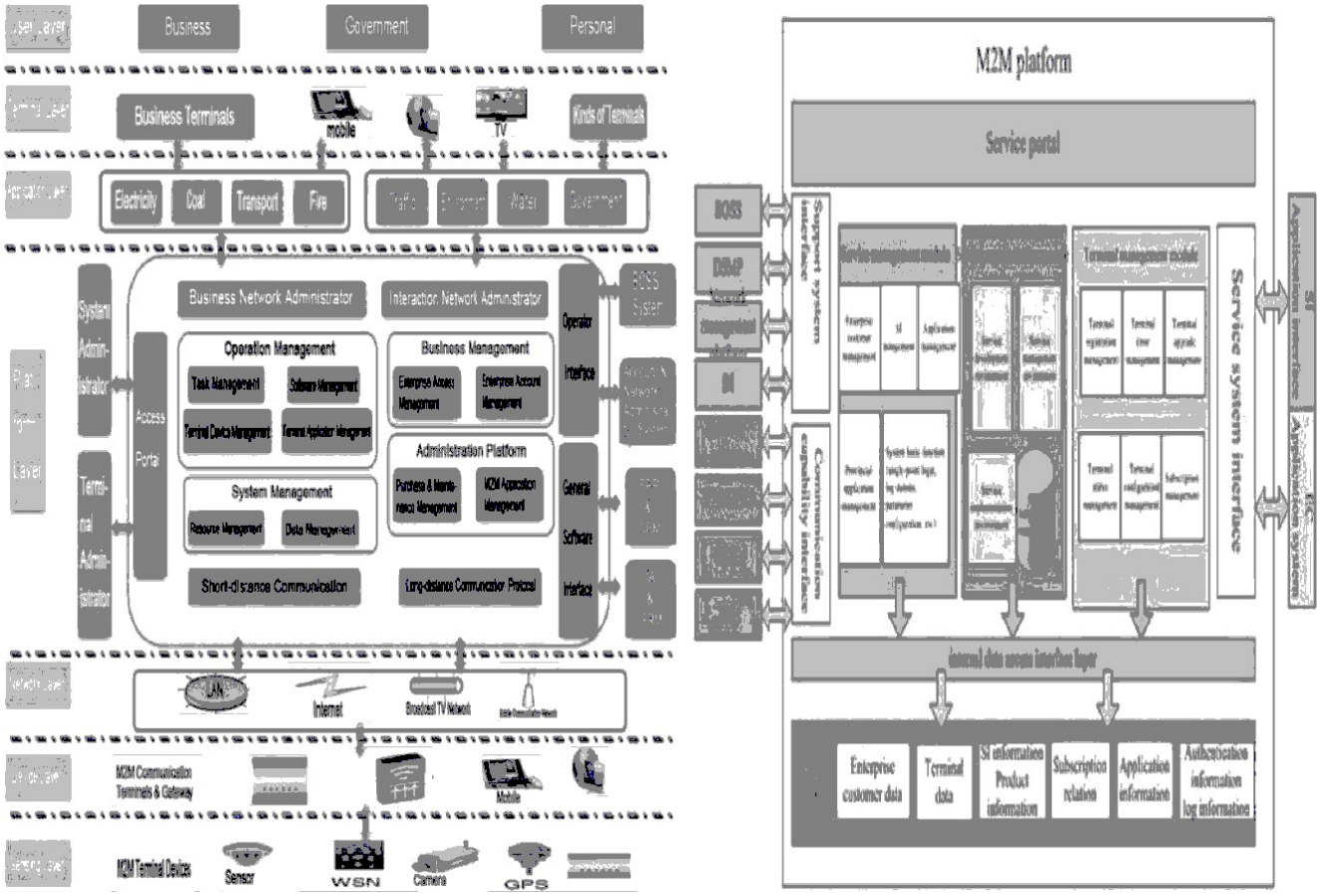


Figure 4.9 Unified iot architecture efforts in China.

- ***More Ingredients: LBS, GNSS, RTLS, and Others***

Other technologies and components are widely used and required in IoT applications; however, those ingredients are not needed for all IoT systems at all times. According to SRI Consulting Business Intelligence, the technologies of the Internet of Things are summarized in Table 4.3. In the Building Blocks column we have discussed almost all of the IoT technologies, which is the goal of this book, except the Location Technology, which must be covered. Positioning capabilities and location-based services (LBS) are required for all mobility IoT applications such as telematics, fleet management, assets tracking in supply chain, and so on.

LBS is a type of context-aware computing, a term first introduced by Schilit in 1994 [71]. In 1996, the Federal Communications Commission issued the order for
enhanced-911 (E-911) to provide the location of wireless callers using 911 emergency services, resulting in significant devel- opment in wireless location technologies and later location- based services. LBSs enable a customer to see the location of its devices in real time and retrieve basic information such as whether the device is registered as well as the history of data sessions. This valuable information enables the customer and the M2M solution providers to determine if the device is functioning as intended and its exact location. Should a ser- vice call, such as a part change, become necessary, they have the means to quickly and accurately locate the device. LBS can enhance the stickiness of any M2M/IoT application, especially for highly mobile solutions; new business lines and incremen- tal revenue streams can be realized using LBS [182] creatively. A group of startups such as FourSquare, Gowalla, Loopt, myTown, BrightKite, Rummble, and others as well as Google's Latitude are providing innovative LBS services.
LBSs work using one or more of a combination of three technology protocols to determine a device's location. If the device has a GPS chip and line of sight to the navigation satel- lites, GPS provides the most accurate location: 15 to 100 feet. Should a pure GPS reckoning not be available due to atmo- spheric conditions or line-of-sight issues, assisted GPS or dif- ferential GPS will be used, providing a hybrid of satellite and cell tower location–based information, resulting in accuracy of 15 to 50 feet. If a device does not have any type of GPS technology, then enhanced cell ID will be used, which will triangulate the location of the device according to the nearest cell towers and the relative signal strength between them.
A global navigation satellite system (GNSS) is a system of satellites that provides autonomous geospatial positioning with global coverage. It allows small electronic receivers to deter- mine their location (longitude, latitude, and altitude) to within a few meters using time signals transmitted along a line of sight by radio from satellites. Such satellites are often medium earth orbit communications satellites (discussed in the last sec- tion) that are also used for M2M communications.
The U.S. NAVSTAR GPS was the only fully operational GNSS before October 2011. The Russian GLONASS (Global Orbiting Navigation Satellite System) achieved full global coverage
in October 2011 after the successful launch of the latest GLONASS satellite. China is in the process of expanding its regional Compass (Beidou) navigation system into a GNSS by 2020. The European Union's Galileo positioning system is a GNSS in initial deployment phase, scheduled to be fully oper-

ational by 2020 at the earliest. All of those GNSS satellites use CDMA for communications. The Indian Regional Navigational Satellite System is an autonomous regional satellite naviga- tion system being developed by Indian Space Research Organization. Other countries such as France and Japan are also developing their own GNSSs.

A local positioning system (LPS) is a navigation system that provides location information in all weather, anywhere within the coverage of the network where there is an unobstructed line of sight to three or more signaling beacons of which
the exact position on Earth is known. Beacons include cel- lular base stations, Wi-Fi access points, RFID readers, radio broadcast towers, and so on. In the past, long-range LPSs have been used for navigation of ships and aircraft. Examples are the Decca Navigator System and LORAN. Nowadays, LPSs are often used as complementary or alternative positioning tech- nology to GPS, especially in areas where GPS does not reach or is weak, for example, inside buildings or urban canyons.

A special type of LPS is the real-time locating system (RTLS), which uses simple, inexpensive badges or tags attached to the objects, and readers receive wireless signals from these tags to determine their locations. According to IDTechEx, the market for RTLS is $380 million in 2011 rising to $1.6 billion in 2021.

A wide variety of wireless systems can be leveraged to provide real-time locating including active RFID, infrared, low-frequency signpost identification, ultrasonic ranging, ultra- wideband (UWB), Wi-Fi, Bluetooth, and so on. The locat- ing methods or algorithms include angle of arrival, line of sight, time of arrival, time difference of arrival, time of flight, received channel power indicator, received signal strength indication, symmetrical double sided–two way ranging, near- field electromagnetic ranging; and so on.

A geographic information system (GIS)—a fusion of car- tography, photogrammtery (the author worked at the Institute of Photogrammetry of ETH Zurich on related research in the late 1980s), statistical analysis, and database technology—is a system designed to capture, store, manipulate, analyze, man- age, and present all types of geographically referenced data. A GIS map labeled with a variety of points of interests is a fun- damental tool for many vertical IoT applications. Traditionally, maps are made up only of the more permanent fixtures of the earth's surface: roads, rivers, mountains, streets, to name a few. Over the past two decades, however, the widespread availabil- ity of GPS and mapping software has changed the landscape. Today, for example, a GPS device fed by sensors can show the state of congestion of the roads in real time on a GIS, such
as the INRIX traffic services, an air-traffic controller is able to see a real-time GIS map of airplane traffic, and so on.

All these possibilities and more are shifting GIS from the relatively leisurely process of analyzing static data to a far more dynamic process of real-time monitoring and decision making. With the advent of the Internet of Things, GIS will involve much more real-time situation monitoring and assessment that treat infor- mation as continually changing.

UNIT – II

Middleware and IOT

An overview of Middleware

There are several historical stories that linguistically unite humanity across the planet: the Tower of Babel, Enmerkar and the Lord of Aratta, Xelhua, and Toltecs. Middleware deals with the babble between distributed systems and has a similar objective in bringing linguistic or communicative unity to disparate technological systems.

The term *middleware* stems from distributed computing and refers to a set of enabling services such as standardized APIs, protocols, and infrastructure services for supporting the rapid and convenient development of distributed services and applications based on the client/ server and later multitier paradigm, which was essential for migrating single- tiered mainframe/ terminal applications to multitier architecture. Middleware is about integration and interoperability of applications and services running on heterogeneous computing and communications devices.
The services it provides, including identification, authentication, authorization, soft- switching, certification, and security, are used in a vast range of global appliances and systems, from smart cards and wireless devices to mobile services and e- commerce. When the first distributed applications became widely used in the early 1990s, application developers were increasingly faced with a multitude of heterogeneous programming languages, hardware platforms, operating systems, and communication protocols, which complicated both the programming and deployment of distributed applications.

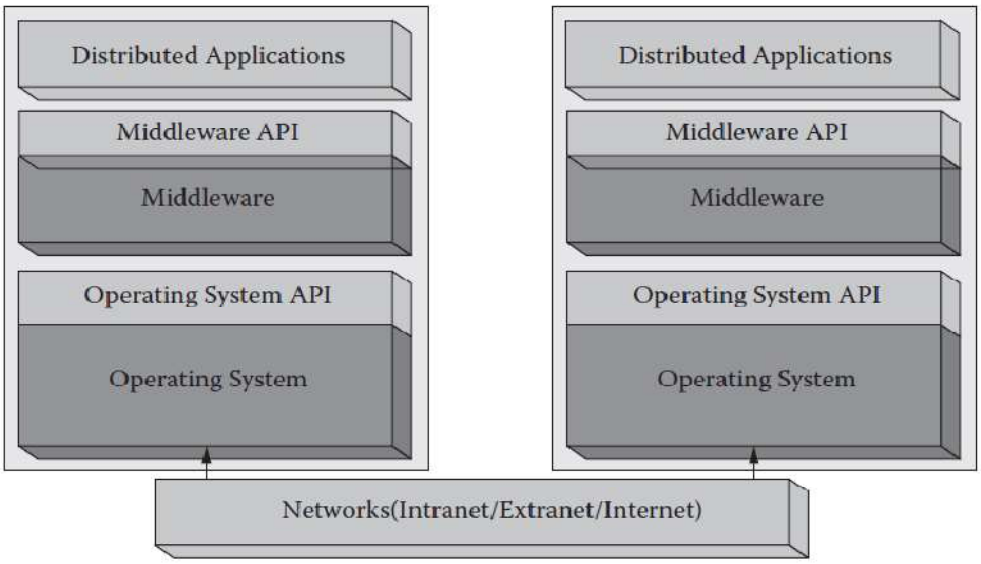


Figure 5.1 Omnipresent middleware.

The term *middleware* refers to a layer that is arranged on top of operating systems and communications stacks and thus hides heterogeneity from the applications through a set of common, well- defined interfaces (Figure 1). In this way, the distributed client and server components of which an application is made up can be programmed in the same manner as if they were executed on the same host.

Middleware brings the following values to the table:

- Enables applications running across multiple platforms to communicate with each other
- Shields the developer from dependencies on network protocols, operating systems, and hardware platforms
- Is a software layer that lies between the operating system and the applications on each site of the system
- Hides heterogeneity and location independence
- Increases software portability
- Provides common functionality needed by many applications
- Aids application interoperability
- Aids scalability
- Helps integrate legacy facilities

Middleware is omnipresent and it exists nearly everywhere in an information and communications technology (ICT) system. Many kinds of middleware are described in related books [160,161]. A list of middleware is compiled below:

- Message- Oriented Middleware (MOM/M Q/ JMS/ ESB)
- CEP (complex event processing) Middleware (Tibco, Sybase)
- Adaptive and Reflective Middleware (TAO/ DynamicTAO/ OpenORB [80])
- Transaction Middleware (TPM/ Tuxedo)
- Peer- to- Peer Middleware (JXTA)
- Grid Middleware (PVM/ MPI/ Schedulers)
- Model- Driven Middleware (CoSMIC)
- Games Middleware (Autodesk)
- Mobile Computing Middleware (OSA/ Parlay/ JAIN/ OMA)
- Radio- frequency Identification (RFID) (Smart Cards) Middleware (Edgeware)
- Three- tiered Application Server Middleware (Weblogic, Websphere)
- Real- time CORBA Middleware (Real- time CORBA)
- High- Availability (Fault Tolerance) Middleware (Fault- Tolerant CORBA)
- Security Middleware (Siteminder)
- CATV/ IPTV Middleware (MHP/ GEM/ OCAP) [181]
- RFID Edge Middleware (OATSystems, Sybase, Oracle,Tibco, SeeBeyond, IBM, SAP, Connectera, GlobeRanger, Manhattan Associates)
- Process- Oriented Middleware (WebMethods, SeeBeyond, Tibco, IBM, SAP, Oracle)
- Business- to- Business (B2B)-Oriented Middleware(SeeBeyond/ Oracle, Tibco, webMethods)
- Middleware for Location- Based Services
- Surveillance Middleware

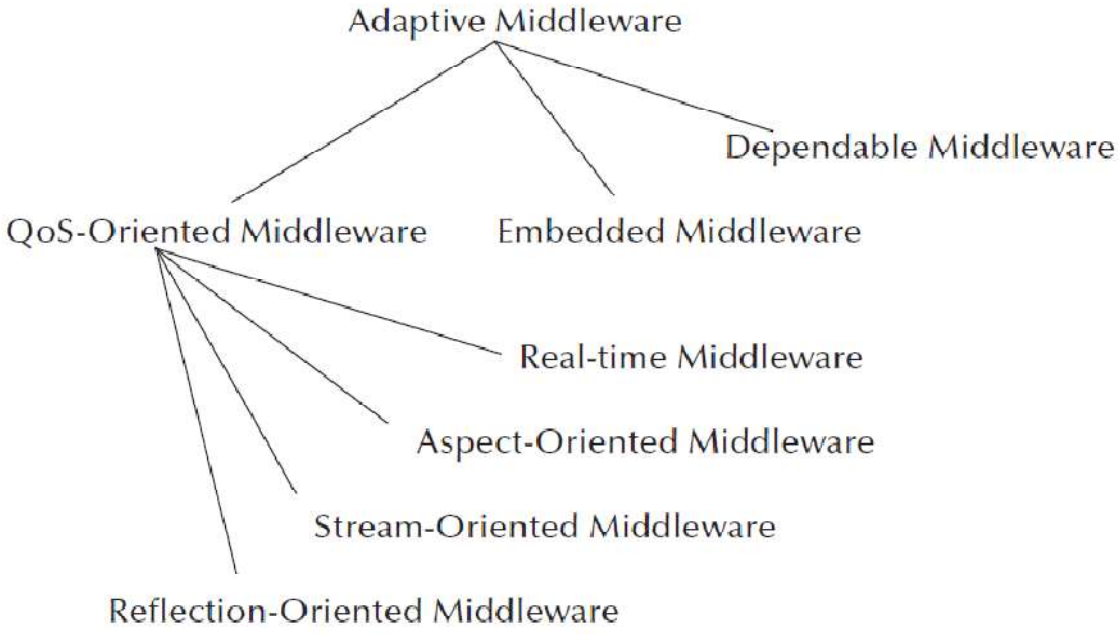


Figure 2

It's argued by some that with middleware proliferation these days, middleware is everywhere (there is also the concept of Everyware, an IoT software platform based on OSGi). It seems like every time that more than two applications need to be integrated, a piece of middleware has been deployed to handle the task. The trouble is that this has led to a lot of middleware sprawl because most of these middleware deployments are tactical, as opposed to being part of information technology (IT) strategy.

Middleware is also the software "glue" that helps programs and databases running on different computers to work together. Gartner formally defines middleware as: "Runtime system software that directly enables application- level interactions among programs in a distributed computing environment"

The basis for nearly all middleware approaches was formalized by the International Organization for Standardization (ISO), which defined the common principles and structures of middleware in a framework known as Reference Model for Open Distributed Processing (RM- ODP). The main objective of ODP is to achieve distribution, interworking, and portability in an environment of heterogeneous IT resources and multiple organizational domains of different participants. ODP groups the functions of middleware into different transparency mechanisms, such as location, failure, persistence, transaction, and scalability. Each of them provides a number of APIs and services to the developer for masking the complexity associated with the respective functions.

The common principles of ODP have been adopted by many of the major middleware platforms, such as OSF DCE (Open Software Foundation's Distributed Computing Environment),

Common Object Request Broker Architecture (CORBA), Java's

Remote Method Invocation (RMI) and Java EE, .NET/ DCOM of Microsoft, LAMP (Linux, Apache, MySQL, PHP/ Perl/ Python), and several approaches for web services. All of these provide several infrastructure services and support different communication patterns, for example, synchronous and asynchronous interactions.

A taxonomy of middleware functionality is outlined by Gartner with three major categories: the integration middleware, the basic middleware, and the development and management tools. More than a dozen different functions that can be performed by middleware have been identified.

The integration middleware covers business- and application- oriented commonalities that include the following:

- Business process management
- Business rule engine/ workflow
- Business event management
- Data routing and adapters

The basic middleware is the foundation, which applies to the Internet of Things (IoT) infrastructure also, and it can be further categorized as follows:

- Data management middleware: helps programs read from and write to remote databases or files. Examples of this kind of middleware include distributed and parallel file systems, such as Google File System, IBM GPFS, Network File System, and Windows, and also include the remote database access middleware, such as Open Database Connectivity or Java Database Connectivity libraries that are bundled into DBMSs such as IBM DB2, Oracle, and Microsoft SQL Server.
- Communication middleware: software that support protocols for transmitting messages or data between two points as well as a system programming interface (SPI) to invoke the communication service. More- advanced communication middleware (such as message- oriented middleware) also support safe (e.g., using strong security) and reliable (e.g., guaranteed once and only once) delivery of messages. Protocols and SPIs used in communication middleware can be proprietary (e.g., IBM WebSphere MQ/ MQ- TT or Microsoft MSMQ) or based on industry standards such as ASN.I, DCE remote procedure call (RPC), CORBA/ IIOP, Java Message Service (JMS), or web services (based on SOAP or REST). Today's communication middleware generally runs on Internet- based protocols such as HTTP (HTTPS), IP, SMTP, and so forth. It may implement higher level protocols, including industry standards (e.g., ebXML messaging and web services), and proprietary protocols (e.g., Oracle AQ), and it may run over the Internet or private networks. Communication middleware also includes *embedded middleware.* Research has been done on middleware and associated standard protocols for home automation and building controls

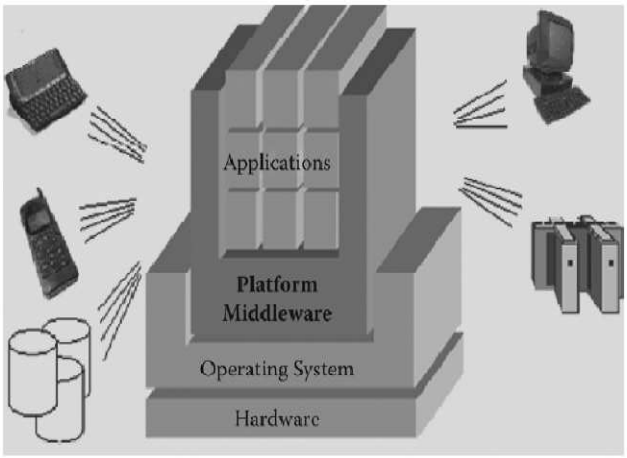


Figure 3 Platform middleware

- Platform middleware: provides the runtime hosting environment (a container) for application components (see Figure 3). It uses embedded or external communication middleware to help programs interact with other programs. It also provides resource management services for hosting application modules at runtime (caching, starting, stopping, and multiplexing programs, load balancing, fault tolerance, access security, monitoring and management, distributed transaction processing, etc.). Platform middleware also provides interfaces to one or several forms of communication middleware (one- way messaging and request/ reply). Platform middleware is well known today as *application servers* (JAVA EE or .NET Framework/ COM+). However, historically, many other product categories have served as then- prevailing platform middleware. Examples include mainframe transaction processing monitors (TPMs such as IBM CICS), Unix- distributed TPMs (such as BEA Tuxedo; the author used to be part of the team), extended RPC implementations, extended object request brokers (ORBs) and object transaction monitors, DBMS stored procedures platforms, proprietary fourth- generation languages, and programmable web servers. Platform middleware has been evolving further in part because of the growing interest in *portal* services such as personalization, multichannel access, and content management. Numerous vendors offer portal services as separate products such as BEA Weblogic Portal, Plumtree, Vignette, and others that are meant to complement web servers and application servers.
- Middleware and the applications software built on top of it are becoming increasingly important in the networked device marketplace. For the nonnetworked device market, the profit is from the device product itself. For the networked device market, additional profits come from consumables, services, and contents. According to Harbor Research, after the "transition point," "the device itself becomes secondary to the value it brings to the customers. Connectivity become the means to cultivate an ongoing relationship." R. Achatz, chairman at Siemens Corporate Research, noted, "We have more software developers than Oracle or SAP, but you don't see this because it is embedded in our trains, machine tools and factory automation" The landscape of CapEx (Capital Expenditure) and OpEx (Operation Expenditure) is changing.
- Device miniaturization, wireless computing, and mobile communication are driving ubiquitous, pervasive, and transparent computing. Supporting these rapidly evolving technologies requires middleware solutions that address connectivity- level, location- dependent, and context- dependent issues. Many companies have developed common application *platform middleware* frameworks for M2M or IoT applications, which will be discussed in more detail in Chapter 7. We will talk more about *communication middleware* in the following sections with regard to its association with M2M or IoT applications

Communication Middleware for IOT

In a runtime environment, the DCM (device, connect, and manage) three- layer model can be further extended into more layers depending upon the geographical scope of the area network (AN) from BAN to interplanetary Internet as listed below:

Body (BAN)
Personal (PAN)
Near- me (NAN)
Machine- to- machine, or M2M (MAN)
Local (LAN)

– Home (HAN)
– Storage (SAN)
Campus (CAN)
Backbone
Metropolitan (MAN)
Wide (WAN)
Internet
Interplanetary Internet

In this section, we will talk about the extensions and enhancements of the existing technologies in the device and connect layers. If the IoT applications are to be extended from the current insolated Intranet or Extranet environments to the wide area as well as global Internet landscape, some fundamental changes in the networking systems have to be considered in a converged next- generation network (NGN) setting.

Some efforts such as the (open- source) Hydra project are under way to build a *unified communication network middleware* for IoT applications. Hydra [133] (networked embedded system middleware for heterogeneous physical devices in a distributed architecture) is a European Union–sponsored IoT open- source project (FP6 IST-2005-034891) that aims to reduce the complexity by developing service- oriented middleware.

MTC/M2M Middleware

The 3GPP (Third Generation Partnership Project) is a collaboration between groups of telecommunications associa tions known as the Organizational Partners. The Organizational
Partners are the European Telecommunications Standards
Institute (ETSI), Association of Radio Industries and Businesses/ Telecommunication Technology Committee (Japan), China Communications Standards Association, Alliance for Telecommunications Industry Solutions (North America), and Telecommunications Technology Association (South Korea). The project was established in December 1998.
The connect layer of DCM can be further divided into three layers based on 3GPP's efforts for GSM/ WCDMA family (3GPP2 for CDMA family) cellular wireless M2M standardization: the M2M area network layer, the access/ core network layer, and the external/ Internet network layer, as depicted in the 3GPP/ ETSI graphic in [230]. The M2M platform in the graphic is an IoT platform middleware at the "M" layer in the DCM value chain.

- M2M area network—provide wired or wireless connectivity between M2M devices and M2M gateways, such as personal area network
- M2M access/ core network—ensure M2M devices interconnection from the gateways to the access/ core communication network, such as GPRS/ GSM (GGSN [Gateway GPRS Support Node], SGSN [Serving GPRS Support Node], etc.; WCDMA, and others
- External/ Internet networks (long distance)—communicate between the 3GPP access/ core network and the M2M middleware platform for applications, such as Internet, corporate WANs, and others

Even though 3GPP introduced the concept of the M2M area network and tries to cover RFID, wireless sensor network (WSN), and supervisory control and data acquisition (SCADA) application scenarios, it is applicable for GSM/ WCDMA

cellular M2M only. 3GPP's coverage/ scope for the entire four- pillar IoT networking possibilities are limited. Other IoT applications, for example, SCADA, may not use cellular networks at all.

The concept of machine- type communication (MTC) was introduced by 3GPP [76]. MTC is the term 3GPP used for cellular M2M communication. It refers to communication without (or with limited) human intervention; data are input or generated by machines instead of humans, which can be significantly faster. Most future big data growth will be in the area of M2M machine- generated data, examples of which include

- Satellite- based telemetry application- generated data
- Location data such as RFID chip readings, global positioning system (GPS) output
- Temperature and other environmental sensor readings
- Sensor readings from factories and pipelines
- Output from many kinds of medical devices, in hospitals and homes alike

In 2009, Gartner estimated that data will grow by 650 percent in the following five years. Most of the growth in data is the by- product of machine- generated data, which could also create M2M data burst to the network systems. New communication middleware will play an important role in alleviating or protecting such overloads.

Current mobile networks are optimized for human- to- human communication, not for MTC. The following are some of the characteristics of MTC summarized by 3GPP

- Time tolerant—data transfer can be delayed
- Packet switched only—network operator shall provide
- PS service with or without a Mobile Station International
- Subscriber Directory Number (MSISDN)
- Online small data transmissions—MTC devices frequently send or receive small amounts of data
- Location- specific trigger—intending to trigger MTC device in a particular area, e.g., wake up the device
- Group- based MTC features—MTC device may be associated with one group
- Extra- low power consumption—improving the ability of the system to efficiently service MTC applications

3GPP started the specification for MTC in early 2010; efforts are proposed as follows [66]:

- ▪ Provide network operators with lower operational costs when offering MTC services
- ▪ Reduce the impact and effort of handling large MTC groups
- ▪ Optimize network operations to minimize impact on device battery power usage
- ▪ Stimulate new MTC applications by enabling operators to offer services tailored to MTC requirements
- ▪ Prepare for number and IP address shortages
- Below are issues with current telco networks for M2M:
- ▪ 3GPP SA1 has required solutions to cater for at least two orders of magnitude more devices compared with human to human.
- ▪ Shortage of telephone numbers.
- Shortage of IPv4 addresses.

- ISMI range seems large enough for most operators.
- Network agnostic middleware approaches for matching application and service requirements with available network capabilities in the telecommunication domain are abundant:
- OSA- Parlay of 3GPP, Parlay- X
- JAIN (Java APIs for integrated networks)
- Open Mobile Alliance (OMA)
- Universal Plug and Play (UPnP)
- Devices Profile for Web Services (DPWS)
- Home Audio- Video Interoperability (HAVi)
- Jini and other middleware alternatives

It seems what the 3GPP's M2M effort lacks is specifying a unified middleware framework for all MTC networks. Middleware for networks is discussed in many works [78,79]. Sahin Albayrak et al. [77] emphasized that "we firmly believe that a new middleware architecture with innovative aspects in terms of: full support along the whole path rather than at the front and backend nodes, highly service aware networks, network aware services, and intelligent coordination and cooperation capabilities is the right answer to the upcoming challenges in next generation networks."
As networks evolve today, middleware based on the aforementioned OSA/ Parlay, JAIN, and others for MTC is an area that requires more investigation and integration in the near future. In addition to the MTC optimization of the cellular wireless network, other optimizations or service enablement middleware (described in Chapter 3) are discussed [226,227] and their standardizations are also needed for M2M applications. Service enablement can be built as middleware that provides reliable and efficient connectivity for adjacent industry applications and to enable operators to

- Act as horizontal service providers across applications and industries
- Expand their role as managed service providers
- Capture maximum value as smart service providers

Nokia is one of the earliest vendors that offered M2M middleware. The Nokia M2M platform [228] is based on open, widely accepted middleware (built on CORBA) and communications architecture, and it supports standard GSM technology with a choice of wireless bearers. Open interfaces facilitate easy development, operation, and maintenance of various M2M applications and services, and provide an easy upgrade path for future technologies. IBM also built an MQ- TT (telemetry transport) middleware (http://mqtt.org/) for M2M applications over IP and non- IP networks.
Other kinds of M2M terminals are the CATV STB (set top box), globally executable MHP (GEM), and MHP (multimedia home platform, based on Java technologies) [128]. These are two of the middleware standards for cable TV, IPTV, Blu- Ray player terminals (embedded middleware), and head- end (platform middleware) applications. GEM, based on MHP, is also a recommended standard by ETSI and ITU.

STB- based home gateway terminal is also an important IoT/ M2M application that has been developed for many years. Other middleware for STB M2M devices and head- end systems include Multimedia and Hypermedia Information Coding Expert Group (MHEG), Open Cable Application Program
(OCAP), OpenTV, Media Highway, Digital Video Broadcasting (DVB)- HTML, etc. All- IP convergence applications based on converged middleware will make the "triple network convergence" of China a reality.

In the digital home (or home automation, domotics) scenario, middleware technology refers to a layer of software that lies on top of a home device's or appliance's operating system. Middleware facilitates rapid development and increases scalability of a system and integration of services in digital homes. It bundles hardware and software into a single solution and provides transparent interaction between home systems and databases, enables unified user interfaces, reduces infrastructure requirements, and makes multiple services easier to manage. A typical digital home could have a number of home devices and appliances, which allows the physical interconnection of multiple systems and services. Home systems and services are inevitably supplied by different manufacturers and use a wide range of different protocols and standards for communication. The home systems and services must be interconnected seamlessly with a consistent middleware platform. An example of the integration architecture of middleware with various digital home services based on standards such as UPnP, DPWS, Jini, HAVi, and so forth is available

SCADA Middleware

The concept of MAN (M2M area network) was introduced in 3GPP/ ETSI's MTC specification. This concept also applies to other pillar segments of IoT. However, not all IoT applications will use a cellular network. In fact, most of the traditional SCADA applications have been using local wireline networks for communications. The remote terminal units (RTUs), programmable logic controllers (PLCs), or even process control systems (PCSs) communicate to the SCADA middleware server via gateways (similar to MAN but all wired) that aggregate data from different wired field buses. The SCADA system is accessed in a LAN environment (sometimes xDSL, cable, WiFi, or WiMax can be used) before it is integrated into the corporate back office system.

Considering that many of the field buses also support IP, such as Modbus TCP/ IP, BacNet IP, and others, it is possible or easier than wireless networks to adopt an all- IP approach to implement SCADA applications. This approach has been used in some of the projects done by the author in building management systems. Figure 5.3 (redrawn based on concepts from [264]) depicts the role of SCADA middleware in such a scenario in more detail.

Companies providing such SCADA middleware products include the following:

- Central Data Control: CDC provides the software platform Integra, which utilizes data agents to translate protocols from different building system components into single management system.
- Elutions: Its Control Maestro product has a SCADA heritage. SCADA may be best known for industrial processes but is also deployed for infrastructure (water treatment plants, gas pipelines, etc.) as well as facility systems. Control Maestro is web- based, uses human–machine,

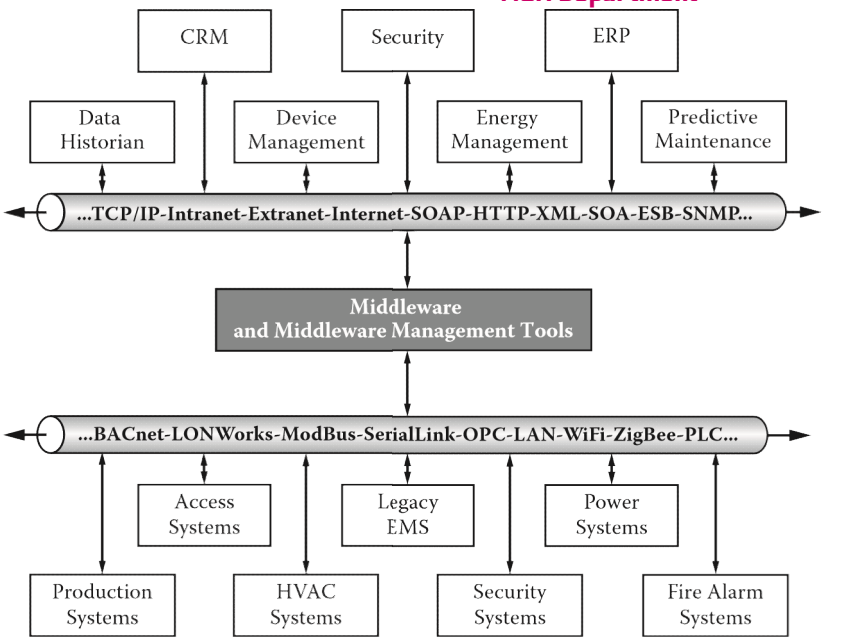


Figure 4 SCADA middleware architecture.

interfaces (HMI), and is able to deliver real- time and historical information.

- Richards Zeta: RZ's middleware solution is a combination of system controllers and software.
- Tridium: It provides the Niagara Java- based middleware framework and JACE hardware controllers. The Niagara platform provides protocol translation for a range of systems and the tools to build applications. Niagara has open APIs to all Niagara services and an extensible component model (XML) that enable development of applications by third parties. It also provides support for web- services data handling and communications with enterprise applications.

With the development of wireless technologies, systems have been developd that blend wireless with wired communication in SCADA applications. SensiLink™ is a middleware and software suite from MeshNetics that links wireless sensor networks with SCADA systems. Sensor data collected from the nodes is channeled through RS232, RS485, USB, Ethernet, or GPRS gateway to the SensiLink server.

OPC middleware products are one of the important communications layer SCADA middleware that are designed to enhance any OPC standards- based applications. Originally, OPC was defined as a standardized solution for the recurring task of connecting PC- based SCADA/H MI applications with automation and process control devices. Today, the OPC standard has evolved into a robust data carrier able to transport entire enterprise resource planning documents and even video signals. OPC is for Windows only (details about the standard is discussed in Chapter 6). Tridium is arguably the first SCADA middle ware based on Java technology. Recent developments have integrated new technologies such as Java and iOS (application store) to build OS platform agnostic middleware for broader IoT applications; adopting new technologies for SCADA is a trend.

RFID Middleware

RFID networking shares a similar three- tiered communication architecture (as shown in Figure 5.4). RFID readers are the gateways similar to MAN. Data from the readers go to the corporate LAN and then are transmitted to the Internet as needed. However, just like the scenarios of M2M and SCADA, most current RFID systems stop at the corporate LAN level and are IoT systems only.

RFID middleware (including the edge middleware or edge ware) is currently no doubt the most well- defined, comprehensive, standardized middleware compared with the other three pillar segments of IoT. Before 2004, an RFID middleware- based system was defined by EPC global, which included:

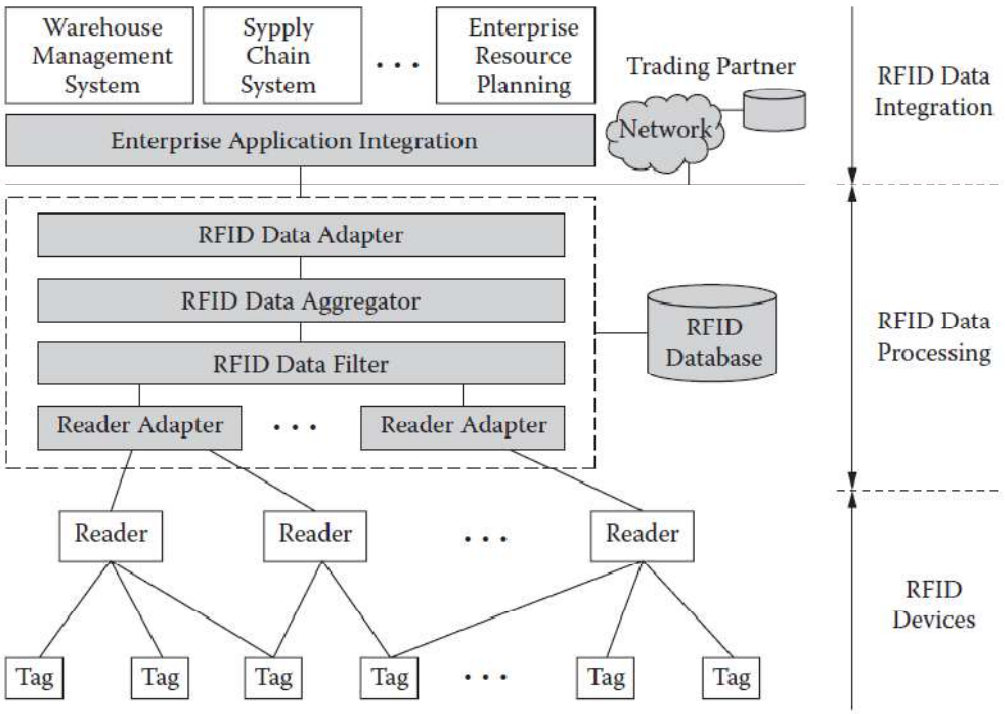


Figure 5.RFiD architecture

- A format for the data called physical mark up language (PML), based on XML (Figure 5 is an example)
- An interface to the servers containing PML records
- A directory service called ONS (object naming service), analogous to the DNS. Given a tag's EPC, the ONS will provide pointers to the
  PML servers containing records related to that tag.

However, since 2004, the unified PML schema has been dropped [51] due to, most likely, practical reasons because most RFID systems are still in the "Intranet of Things" scope. Using the generic PML/ ONS approach would involve overhead and sacrifice efficiency. Instead, the PML- like schema was left to the vertical applications to define their own XML scheme. Consequently, the overall system architecture of RFID has evolved from a dedicated structure to a more generic, open architecture.

However, the PML approach is believed to be a good IoT data representation method that should be used when the day of the full- blown IoT system comes. Other efforts such as M2MXML (from BiTX) and oBIX (an OASIS standard) are under way that are trying to build a generic IoT data schema, which is discussed in the next chapter.

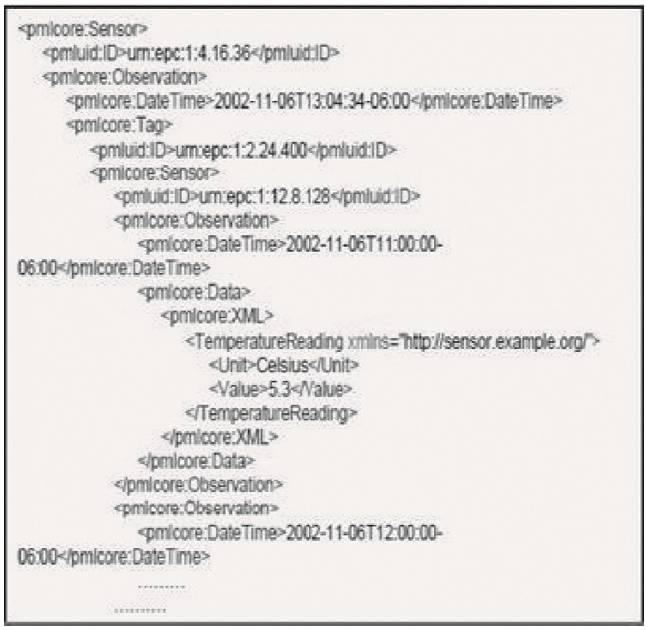


Figure 6 Physical markup language sample.

An example of commercial RFID middleware product is IBM's WebSphere Sensor Events. WebSphere Sensor Events delivers new and enhanced capabilities to create a robust, flexible, and scalable platform for capturing new business value from sensor data. Web Sphere Sensor Events is the platform for integrating new sensor data, identifying the relevant business events from that data using situational event processing, and then integrating and acting upon those events with SOA business processes.

The blending or convergence of different pillar IoT applications to build cross- segment IoT systems is a trend that has been demonstrated [228], in which unified data representation and associated communication middleware became more and more important.

WSN Middleware

Middleware also can refer to software and tools that can help hide the complexity and heterogeneity of the underlying hardware and network platforms, ease the management of system resources, and increase the stableness of application executions. WSN middleware is a kind of middleware providing the desired services for sensor- based pervasive computing applications that make use of a WSN and the related embedded operating system or firmware of the sensor nodes [57]. In most cases, WSN middleware is implemented as embedded middleware on the node [82].

It should be noted that while most existing distributed system middleware techniques aim at providing transparency abstractions by hiding the context information, WSN- based applications are usually required to be context aware, as mentioned in Chapter 1 [18].

A complete WSN middleware solution should include four major components: programming abstractions, system services, runtime support, and quality of service (QoS) mechanisms. Programming abstractions define the interface of the middleware to the application programmer. System services provide implementations to achieve the abstractions. Runtime support serves as an extension of the embedded operating system to support the middleware services. QoS mechanisms define the QoS constraints of the system. The system architecture of WSN middleware is shown in Figure 6.

Middleware for WSN should also facilitate development, maintenance, deployment, and execution of sensing- based applications. Many challenges arise in designing middleware for WSN due to the following reasons and more:

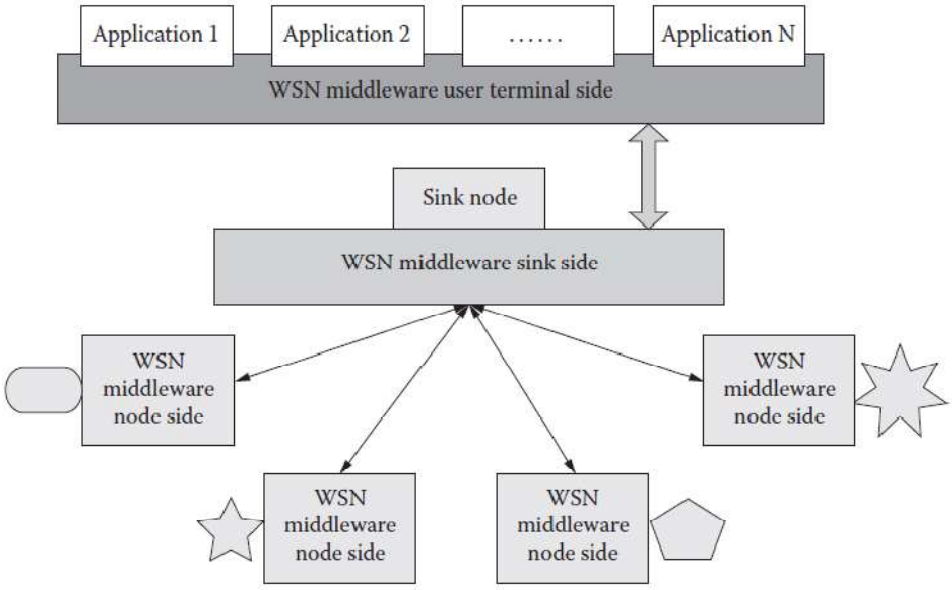


Figure 6 WSn middleware architecture.

- Limited power and resources, e.g., battery issues
- Mobile and dynamic network topology
- Heterogeneity, various kinds of hardware and network protocols
- Dynamic network organization, ad- hoc capability

WSN middleware is designed using a number of approaches such as virtual machine, mobile agents, database based, message- oriented, and more. Example middleware are as follows [83]:

- Magnet OS (Cornell University): power- aware, adaptive; the whole network appears as a single JVM, standard Java programs are rewritten by MAGNET as network components, and components may then be "injected" into the network using a power- optimized scheme.

- IMPALA: modular; efficiency of updates and support dynamic applications; application adaption with different profiles possible; energy efficient; used in the Zebra Net project for wildlife monitoring.
- Cougar: represents all sensors and sensor data in a relational database; control of sensors and extracting data occurs through special SQL- like queries; decentralized implementation; message passing based on controlled flooding.
- SINA (system information networking architecture): based on a spreadsheet database wherein the network is a collection of data sheets and cells are attributes; attribute- based naming; queries performed in an SQL- like language; decentralized implementation based on clustering.
- MIRES: publish/ subscribe; multi hop routing; additional service (e.g., data aggregation); sense—advertise over P/ S and route to sink.
- MQTT- S (Message Queue Telemetry Transport for Sensors, IBM): a publish/ subscribe messaging protocol for WSN, with the aim of extending the MQTT protocol beyond the reach of TCP/ IP infrastructures (non- TCP/ IP networks, such as Zigbee) for sensor and actuator solutions; a commercial product.
- MiLAN: provides a mechanism that allows for the adaptation of different routing protocols; sits on top of multiple physical networks; acts as a layer that allows network- specific plug- ins to convert MiLAN commands to protocol-specific ones that are passed through the usual network protocol stack; can continuously adapt to the specific features of whichever network is being used in the communication.
- The WSN middleware is considered to be "proactive" middleware in the middleware family. A more comprehensive list of existing WSN middleware platforms, software/ OS, and programming languages is shown in Table 5.3. A comparison of some of the WSN middleware is available [84].

As an example, the Agilla middleware is examined here in more detail (Figure 5.7). The Agilla [229] runs on top of TinyOS and allows multiple agents to execute on each node. The number of agents is variable and is determined primarily by the amount of memory available. Each agent is autonomous but shares middleware resources with other agents in the system.

Table Sample WSN Middleware and WSN Languages

WSN Middleware			
Agilla	eCos	MagnetOS	SINA
AutoSec	EMW	MANTIS	SOS
Bertha	Enviro- Track	Mate	TinyDB
BTnut Nut/ OS	EYESOS	MiLAN	TinyGALS
COMiS	FACTS	Mire	TinyOS
Contiki	Global Sensor Networks (GSN)	Netwiser	t- Kernel
CORMOS	Impala	OCTAVEX	VIP Bridge

COUGAR	jWebDust	SenOS	
DSWare	LiteOS	SensorWare	
WSN Languages			
c@t	DCL (Distributed Compositional Language)	galsC	nesC
Protothreads	SNACK	SQTL	

Agilla provides two fundamental resources on each node: a neighbor list and a tuple space. The neighbour list contains the addresses of neighbouring nodes. This is necessary for agents to decide where they want to move or clone to next. The tuple space provides an elegant decoupled- style of communication between agents. It is a shared memory architecture that is addressed by field- matching rather than memory addresses. A tuple is a sequence of typed data objects that is inserted into the tuple space. The tuple remains in the tuple space even if the agent that inserted it dies or moves away. Later, another agent may retrieve the tuple by issuing a query for a tuple with the same sequence of fields. Note that tuple spaces decouple the sending agent from the receiving agent: they do not have to be co- located, or even aware of each other's existence, for them to communicate. This is basically a fault- tolerant distributed computing technology.

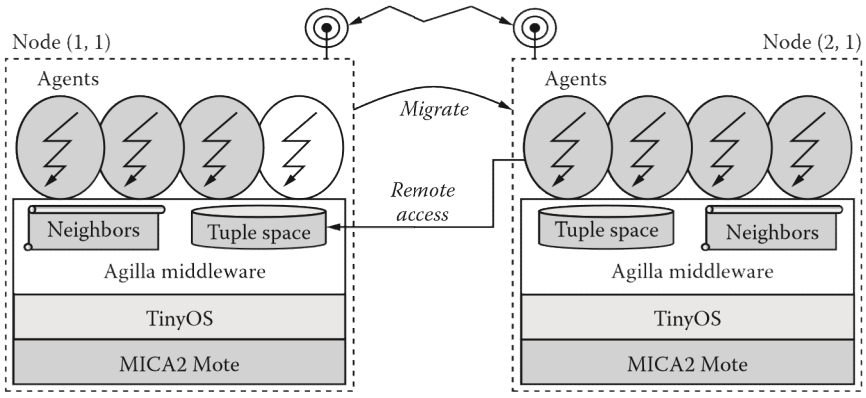


Figure 7 The Agilla middleware model.

All of the above WSN middleware are at the device level up to the gateways (equivalent to the MAN of MTC). Most of them are research projects conducted at universities and research institutions with a few experimental uses and of limited commercial value. This situation is very much like the research on parallel computing architecture one or two decades ago. There was a proliferation of parallel architectures [85] such as hypercube, wave front arrays, pyramids, systolic arrays, and others, which the author has gone through [86–95]. Many research papers have been produced, but none of these architectures exist in the real world now. Nowadays, 99 percent of the world's fastest high performance computing (HPC) supercomputers use the simple massive parallel processing (MPP) architecture [96]. David Culler, the inventor of TinyOS and Mote, professor at University of California–Berkeley, was one of the prominent researchers on parallel architecture at

that time. He has been doing research on WSN since the wane of parallel architecture research. In fact, some of the WSN architecture and middleware ideas are inherited from parallel computer architectures, which will most likely diminish the same way as time passes by, especially the ad hoc wireless networks (they may have greater value in military uses).

Nevertheless, once the data from the ad hoc mesh WSN reaches the gateways, or if the wireless sensors are directly connected to the higher- tier networks, the remaining process and route to reach the Internet of Things will be the same as the other pillar segments of IoT. The WSN middleware at the system level may be the same as SCADA or M2M or RFID systems, which share the same three- tiered architecture

LBS and Surveillance Middleware

Other than the communication middleware and the platform middleware (which will be covered in Chapter 7) for IoT applications, other middleware are related IoT or are part of IoT. Location- based service (LBS) and surveillance middleware are two of the examples we choose to cover in this chapter.

LBS is a service that integrates a mobile device's location or position with other information so as to provide added value to a user [97]. There are several uses of LBS, and some of them are direct IoT applications:

- News: information dissemination based on the location of a user, such as weather information
- Point of interest (POI): shows points of interest near the user or vehicles
- Directions: shows directions from the current location of a user
- Yellow pages: finds services near the user
- Fleet management: tracks positions of a transportation fleet
- Local advertisement: user receives advertisements according to his or her position
- Emergency: tracks current position of a user in an emergency
- Location- based games: player interacts with another player according to his or her position
- LBS scenarios involve collecting, analyzing, and matching different types of information including user profiles (e.g., personal information and interests) and information dissemination profiles. For each piece of information, LBS systems have to handle different aspects:
- Spatial: LBS middleware must be able to collect information about mobile position and fixed elements, associate them with physical/ logical maps, and efficiently match locations and regions.
- Temporality: Location information has a temporal dimension that must be included in query capability.
- Inaccuracy, imprecision, and uncertainty: LBS must deal with inaccuracy and imprecision associated with location positioning technologies.
- Large volumes: In real scenarios, LBS must handle large volumes of data; scalability is a very important issue.
- Continuous queries: In an LBS scenario, query executions are continuous, so the query engine of an LBS middleware must be efficient.
- An example middleware architecture for LBS systems can be found at locationet.com (http://www.locationet.com/ LBSmiddleware.php). Most LBS middleware can be categorized as event based (publish/ subscribe), tuple space based, context aware, and data sharing based:
- Publish/ subscribe: one of the most prominent middleware models, in which communication is defined in terms of exchanging asynchronous messages based on subscription.

- Tuple space: originally proposed to coordinate concurrent activities in parallel programming systems such as Linda, in which a process communicates with another process in a global collection of tuples. A tuple is a data element that contains values of a specified data type.
- DBMS- based: comprises the use of database interaction to implement a communication and coordination; many geographic information systems (GISs) operate according to this scheme. LBS architecture naturally fits the DBMS- model, such as user management systems and accounting information systems.

As an example, LocatioNet middleware is a product that meets mobile operators' needs for in- house location- privacy management, location billing functionality, provisioning interfaces, and links to various content databases. LocatioNet comprises a set of modules offered in any required
combination:

- Comprehensive location privacy management: allows users to decide who can see their location, when, and how precisely, application by application
- Billing for location: gives operators a flexible set of billing options for their location and GPS services
- Provisioning: enables operators to provision user-to- location and GPS applications
- Content interfaces: enables operators to take advantage of content properties they have access to (such as local news, the weather, points of interest, traffic) by linking them to the location and GPS infrastructure

A Location API for J2ME has been specified as JSR-179 that enables mobile location- based applications for resource-limited devices. Java middleware and applications can be developed based on the Location API standard. The Open GIS Consortium (OGC) also produced a specification about location services called OpenLS™ in 2003.

Automated video surveillance networks are a class of sensor networks (people argued that a video surveillance network without automatic image recognition and event detection or alert generation is not a sensor network but instead simply a video or image capture and transmission system) with the potential to enhance the protection of facilities such as airports and power stations from a wide range of threats. However, current systems are limited to networks of tens of cameras, not the thousands required to protect major facilities. Realizing thousand- camera automated surveillance networks demands sophisticated middleware and architectural support as well as replacing the ad hoc approaches used in current systems with robust and scalable methods.

The IBM Smart Surveillance Solution [100] is based on the MILS (middleware for large- scale surveillance) surveillance middle ware and designed to work with a number of video management systems from partner companies. The MILS provides the data management services needed to build a large- scale smart surveillance application. While MILS builds on the extensive capabilities of IBM's Content Manager and DB2 systems, it is essentially independent of these products and can be implemented on top of third- party relational databases. The MILS take the automatically detected events from the SSE (smart surveillance engine) as inputs. An SSE is a class of surveillance algorithms such as the HMM (Hidden Markov Model) [99].

The IBM SSS system provides two distinct functionalities:

- Real- time user- defined alerts: The user defines the criteria for alerting with reference to a specific camera view, for example, parked car detection, tripwire, and so forth.
- Indexed event search: The system automatically generates descriptions of events that occur in the scene and stores them in an indexed database to allow the user to perform a rapid search.
- Another middleware approach for video surveillance networks is proposed [98]. This surveillance middleware approach partitions systems based on an activity topology— a graph describing activity observed by the surveillance camera network. Processing within topological partitions uses well- known architectural styles such as blackboards, and pipes and filters. Communication between partitions uses a service- oriented architecture. This middleware enables building intelligent video surveillance systems at a far larger scale than was previously possible. Communication on the surveillance network follows the service- oriented model with publish/ subscribe messaging, providing scalability, availability, and the ability to integrate separately developed surveillance services.

IOT Protocol Standardization efforts

We have touched on the issues of IoT standardization sporadically in the previous chapters of the book. Now we are going to give a summarized description of the four pillars as well as the generic IoT standardization efforts focusing on data representations and APIs (i.e., protocols). The standards on platform architecture and middleware framework will be discussed in the next chapter. However, because in most cases, the data representation and APIs are intertwined with architecture and framework, it is hard to separate; so there may be some overlaps. Some of the IoT projects such as the Internet of Things Strategic Research Roadmap by CERP- IoT [8] are still at the grand concept level with limited materialized results. The IoT- A (Internet of Things architecture [113]) is one of the few efforts targeting a holistic architecture for all IoT sectors. This consortium consists of 17 European organizations from nine countries. They summarized the current status of IoT standardization as follows:

- Fragmented architectures, no coherent unifying concepts, solutions exist only for application silos.
- No holistic approach to implement the IoT has yet been proposed.
- Many island solutions do exist (RFID, sensor nets, etc.).
- Little cross- sector reuse of technology and exchange of knowledge.

The author had the same observation (also one of the first who introduced the Intranet/ Extranet of Things concept independently [74]) before 2010 based on the four- pillar classification of IoT. Even though the IoT- A consortium doesn't categorize the IoT as four pillars, they do believe solutions for radio- frequency identification (RFID), sensor nets, and so forth are island solutions. In fact, IoT- A doesn't have a systematic, clean- cut, and comprehensive classification of IoT sectors as the foundation. Their "holistic" view of IoT is based on the following scenarios, which is actually not complete and holistic currently.

The key objectives of the IoT- A consortium [103] are as follows:

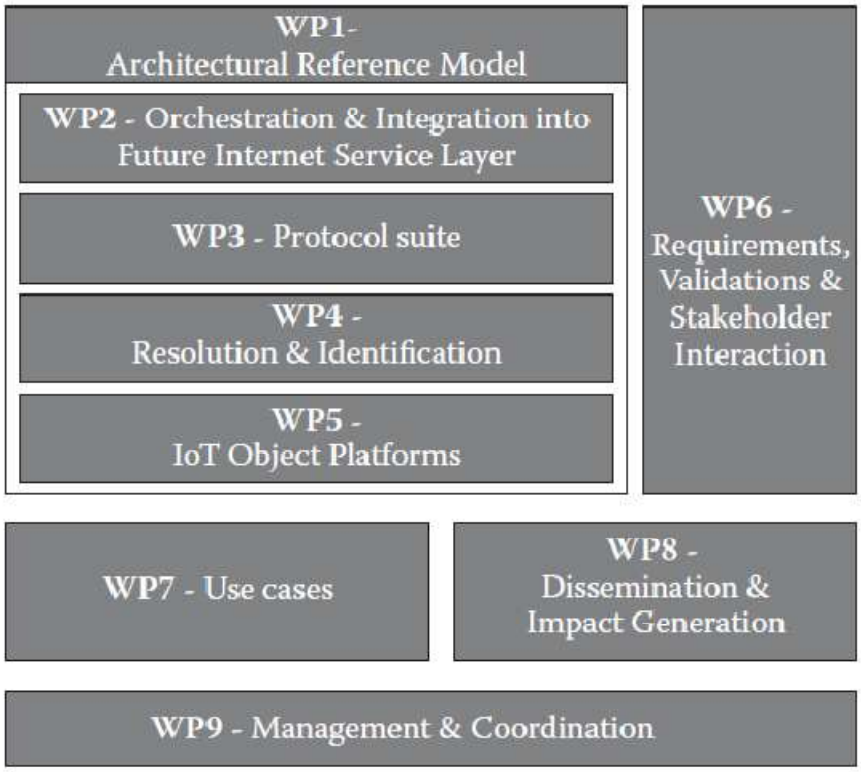
- Create the architectural foundations of an interoperable Internet of Things as a key dimension of the larger future Internet
- Architectural reference model together with an initial set of key building blocks:

– Not reinventing the wheel but federating already existing technologies
– Demonstrating the applicability in a set of use cases
– Removing the barriers of deployment and wide- scale acceptance of the IoT by establishing a strongly involved stakeholder group

- Federating heterogeneous IoT technologies into an interoperable IoT fabric

A WP (work package) framework of ongoing works has been proposed [103]. Also, the ITU- T has a few study groups (SGs 2, 3, 5, 9, 11, 12, 13, 15, 16, and 17, http://www.itu.int/ en/ ITU- T/ tech watch/ Pages/i nternetofthings.aspx) doing IoT- related works (Figure 6.6).

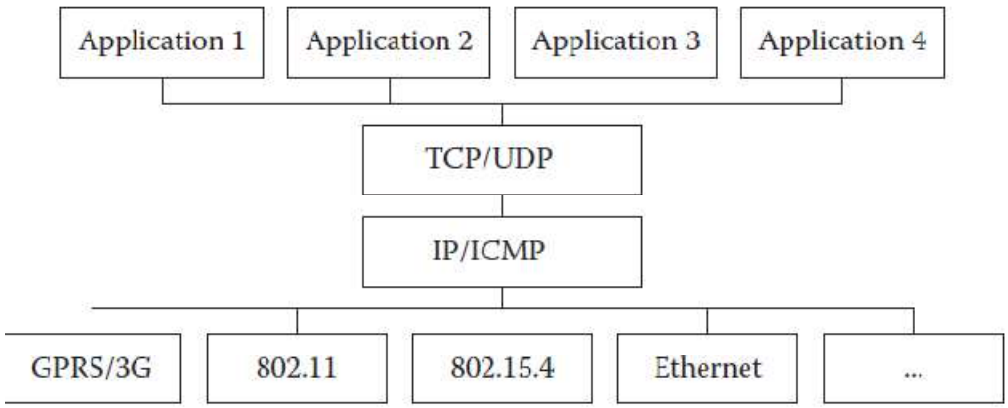
IPSO (Internet Protocol for Smart Objects, http://www.ipsoalliance.org/) Alliance, formed in 2008, is another effort aiming to form an open group of companies to market and educate about how to use IP for IoT smart objects based on an all- IP holistic approach [81] (Figure 6.7).



ITU-T Study Group	Study Group Name	Activities Related to IoT
SG 2	Operational aspects of service provision and telecommunication management	Numbering, naming and addressing
SG 3	Tariff and accounting principles including related telecommunication economic and policy issues	
SG 5	Environment and climate change	
SG 9	Television and sound transmission and integrated broadband cable networks	
SG 11	Signalling requirements, protocols and test specifications	Testing architecture for tag-based identification systems and functions
SG 12	Performance,QoS and QoE	
SG 13	Future networks including mobile and NGN	NGN requirements and architecture for applications and services using tag-based ID
SG 15	Optical transport networks and access network infrastructures	
SG 16	Multimedia coding, systems and applications	Requirements and architecture for multimedia information access triggered by tag-based ID
SG 17	Security	Security and privacy of tag-based applications
Focus Groups	Smart Grid	Smart metering, M2M
	Cloud Computing	Cloud network requirements, e.g., for IoT
	Future Networks	Describe future networks underlying the IoT
	Car Communication	

Working groups of iot standards.

The emerging application space for smart objects requires scalable and interoperable communication mechanisms that support future innovation as the application space grows. IP has proven itself a long- lived, stable, and highly scalable communication technology that supports a wide range of applications, devices, and underlying communication technologies. The IP stack is open, lightweight, versatile, ubiquitous, scalable, manageable, stable, and end- to- end. It can run on tiny, battery- operated embedded devices. IP therefore has all the qualities to make the Internet of Things a reality, connecting billions of communicating devices.



A smart object is defined by IPSO as

- An intelligent (RFID) tag
- A sensor: device that measures a physical quantity and converts it to an analog or digital signal, such as power consumption and quality, vibration of an engine, pollution, motion detection, temperature
- An actuator: device that controls a set of equipment, such as controls and/ or modulates the flow of a gas or liquid, controls electricity distribution, performs a mechanical operation
- An embedded device: a purpose-built connected device that performs a specific function, such as a factory robotic arm, vending machine, smart grid analyzer
- Any combination of the above features to form a more complex entity

The IPSO Alliance works closely with Internet Engineering

Task Force (IETF), the Institute of Electrical and Electronics Engineers (IEEE), the European Telecommunication Standard Institute (ETSI), the International Society of Automation (ISA), and others, and relies on the standards developed by them. IPv4, IPv6, and 6LoWPAN were all developed by engineers within IETF, and the role of the alliance is to ensure how they are used, deployed and provided to all potential users.

The Mobile IP protocol is a related IETF- proposed standard that provides a network layer solution to node mobility across IPv4 (Mobile IPv4) and IPv6 (Mobile IPv6) networks. Mobile IP allows a node to change its point of attachment to the Internet without having to change its IP address.
Another solution to the problem is network mobility (NEMO). NEMO is an extension of Mobile IP that enables an entire network to change its attachment point to the Internet. NEMO works by moving the mobility functionality from Mobile IP mobile nodes to a moving network's router. The router is able to change its attachment point to the Internet in a manner that is transparent to attached nodes.

SHIM6 [114], a serverless Mobile IPv6 protocol, allows two communicating nodes to overcome connection loss problems that may arise if one node changes its IP address (locator) during an established communication.

Sensinode [115], as an example, provides embedded networking software and hardware products based on IP- based 6LoWPAN technology for demanding enterprise applications. NanoStack™ 2.0 is an advanced 6LoWPAN protocol stack software product for 2.4 GHz radios. The NanoRouter™ 2.0 platform includes software and hardware solutions for 6LoWPAN- Internet routing infrastructure.
Also, since its creation in 2003, ETSI TISPAN

(Telecommunications and Internet converged Services and Protocols for Advanced Networking) has been the key standardization body in creating the next- generation networks (NGN) specifications, which is a synonym of IoT.

M2M and WSN Protocols

Most M2M applications are developed today in a highly customized fashion, and vertical- specific industry bodies are busy crafting standards for markets ranging from the auto industry to the smart grid. A broad horizontal standard is a key requirement for the M2M industry to move from its current state of applications existing in isolated silos based on vertical market or underlying technology to a truly interconnected Internet of Things. Such a horizontal standard is expected to be the major impetus to growth in the future.

Efforts to develop broad, horizontal standards for the M2M market are gaining momentum [49,105]. The most important activity is occurring within the context of the International Telecommunication Union's (ITU) and ETSI's (M2M Technical Committee) Global Standards Collaboration (GSC), which has established the M2M Standardization Task Force (MSTF, created during the GSC-15 meeting in Beijing, China, in September 2010) to coordinate the efforts of individual standards development organizations (SDOs), including China Communications Standards Association, Telecommunications Industry Association TR-50 Smart Device, etc.

The end result of these efforts is to define a conceptual framework for M2M applications that is vertical industry and communication technology agnostic, and to specify a service layer that will enable application developers to create applications that operate transparently across different vertical domains and communication technologies without the developers having to write their own complex custom service layer [105]. The high- level M2M architecture from MSTF does include fixed and other noncellular wireless networks, which means it's a generic, holistic IoT architecture even though it is called M2M architecture (M2M and IoT sometimes are used interchangeably in the United States and in the telco- related sectors). Despite all of the positives, it seems the voices from the SCADA (supervisory control and data acquisition) and RFID communities are relatively weak; efforts to incorporate existing SCADA standards such as OPC, ISA-95, and RFID EPCIS, ONS, and others are not seen yet. It remains to be seen whether all of the stakeholders from the four pillars of IoT will be equally included in the loop.

This is a more comprehensive approach than the 3GPP's MTC effort described in the previous chapter. Considering 3GPP is only one of the SDOs in the MSTF, this makes sense and good results are much anticipated from MSTF. Some vertical applications on top of the unified horizontal M2M architecture are already under way . Companies such as Telenor Objects, Numerex, and others are building MSTF standards compliant products already.

Other M2M standards activities include the following:

- Data transport protocol standards: M2MXML, JavaScript Object Notation (JSON) (originally not for IoT applications, used by the Mango open source M2M project), BiTXML [117], WMMP (shown in Figure 6.8), MDMP, open
- Building Information Exchange (oBIX), EEML, open M2M Information exchange (oMIX)
- Extend OMA DM to support M2M devices protocol management objects
- M2M device management, standardize M2M gateway
- M2M security and fraud detection
- Network API's M2M service capabilities
- Charging standards
- MULTI IMSI, M2M sevices that do not have MSISDN
- IP addressing issues for devices IPV6

- Remote diagnostics and monitoring, remote provisioning and discovery
- Remote management of devices behind a gateway or firewall
- Open REST- based API for M2M applications

One of the benefits of using sensor data is that the data typically can be repurposed many times, thereby reducing cost and maximizing benefit. For example, weather observations (temperature, wind speed and direction, humidity, and so on) can be used in climate modeling, weather forecasting, plume modeling, insurance risk analysis, ski area location decisions, and dozens of other applications. However, the ability to access and use the same sensors in multiple application domains, to share sensor data, and to maximize the full value of sensor networks and data is severely hindered by a lack of interoperability. Hundreds of sensor manufacturers build sensors for specific purposes, often using their own "language" or encodings, different metadata, and so forth. Standard data representation (together with WSN middleware) is the key to materialize data integration and increase interoperability

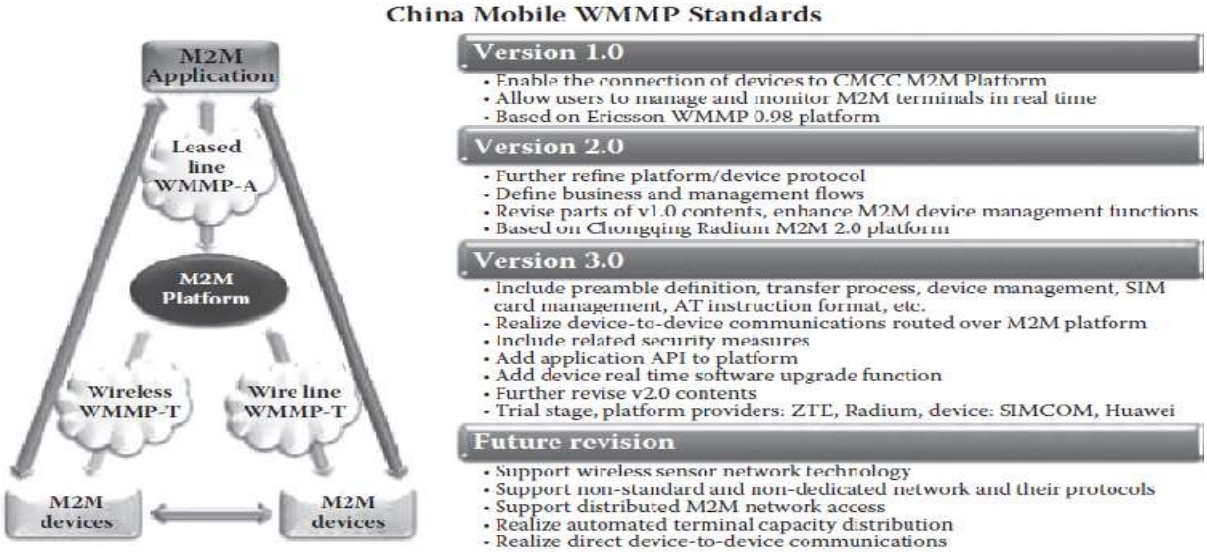


Figure China Mobile's WMMP standard

There are a number of standardization bodies in the field of WSNs. The IEEE focuses on the physical and MAC layers; the IETF works on layers 3 and above. IEEE 1451 is a set of smart transducer interface standards developed by the IEEE Instrumentation and Measurement Society's Sensor Technology Technical Committee that describe a set of open, common, network- independent communication interfaces for connecting transducers (sensors or actuators) to microprocessors, instrumentation systems, and control/ field networks. One of the key elements of these standards is the definition of transducer electronic data sheets (TEDS) for each transducer. The TEDS is a memory device attached to the transducer, which stores transducer identification, calibration, correction data, and manufacturer- related information. The IEEE 1451 family of standards includes the following:

- 1451.0-2007 Common Functions, Communication Protocols, and TEDS Formats

- 1451.1-1999 Network Capable Application Processor Information Model
- 1451.2-1997 Transducer to Microprocessor Communication Protocols & TEDS Formats
- 1451.3-2003 Digital Communication & TEDS Formats for Distributed Multi- drop Systems
- 1451.4-2004 Mixed- mode Communication Protocols & TEDS Formats
- 1451.5-2007 Wireless Communication Protocols & TEDS Formats
- 1451.7-2010 Transducers to Radio Frequency Identification(RFID) Systems Communication Protocols and TEDS Formats

The goal of the IEEE 1451 family of standards is to allow the access of transducer data through a common set of interfaces whether the transducers are connected to systems or networks via a wired or wireless means. IEEE p1451.3 is XML based and allows the manufacturer to change the contents.

Cross- network (e.g., between Bluetooth and ZigBee) standards are not as proliferate in the WSN community compared to other computing systems, which make most WSN systems incapable of direct communication with each other. The contents on WSN described in the previous chapters are more devices or network focused. OGC (Open Geospatial Consortium) and W3C has been doing research and standardization work following a data- focused approach [233].

The Semantic Sensor Web (SSW) is an approach to annotating sensor data with spatial, temporal, and thematic semantic metadata based on OGC SWE (Sensor Web Enablement). The following data- encoding specifications have been produced by OGC SWE Working Group (in addition to the web service specifications that will be described in Chapter 7):

- SWE Common—common data models and schema
- Sensor ML—models and schema for sensor systems and processes surrounding measurements
- Observations & Measurements (O&M)—models and schema for packaging observation values
- Transducer Mark up Language (TML)—models and schema for multiplexed data from sensor systems

The European Union SENSEI [109] project creates an open, business driven architecture that fundamentally addresses the scalability problems for a large number of globally distributed wireless sensor and actuator networks (WSAN) devices. It provides necessary network and information management services to enable reliable and accurate context information retrieval and interaction with the physical environment. By adding mechanisms for accounting, security, privacy, and trust, it enables an open and secure market space for context awareness and real- world interaction. An ambient ERP system supported the SENSEI.

Tangible results of the SENSEI project are as follows:

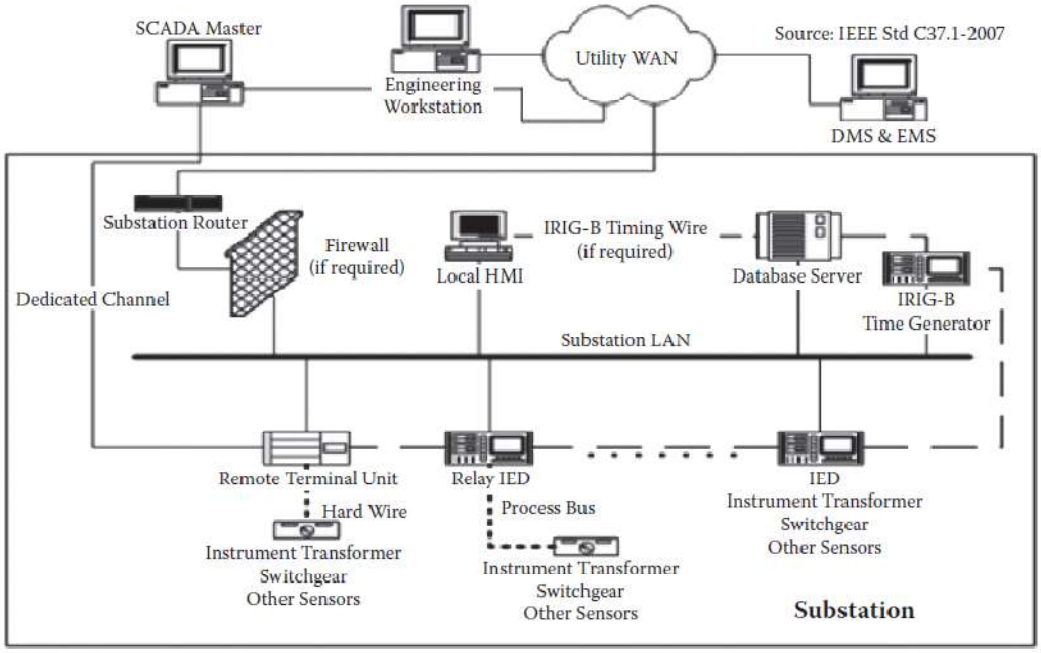
- A highly scalable architectural framework with corresponding protocol solutions that enable easy plug- and- play integration of a large number of globally distributed WSAN into a global system, providing support for network and information management, security, privacy and trust, and accounting
- An open service interface and corresponding semantic specification to unify the access to context information and actuation services offered by the system for services and applications
- Efficient WSAN island solutions consisting of a set of cross- optimized and energy- aware protocol stacks including an ultra- low- power multi- mode transceiver

- Pan European test platform, enabling large- scale experimental evaluation of the SENSEI results and execution
- of field trials, providing a tool for long- term evaluation of WSAN integration into the NGN
- ISO/ IEC JTC1 WG7 (Working Group on Sensor Networks), established in 2009, preceded by JTC 1 SGSN SC6, created the ISO/ IEC 29182 Reference Architecture for sensor networks application and services focusing on telecommunication and information exchange between systems. The architecture is defined through the following set of documents:
- ISO/ IEC 29182 Part 1: General overview and requirements
- ISO/ IEC 29182 Part 2: Vocabulary/ terminology
- ISO/ IEC 29182 Part 3: Reference architecture views
- ISO/ IEC 29182 Part 4: Entity models
- ISO/ IEC 29182 Part 5: Interface definitions
- ISO/ IEC 29182 Part 6: Application profiles
- ISO/ IEC 29182 Part 7: Interoperability guidelines

SCADA and RFID Protocols

As described before, we use the SCADA term as one of the IoT pillars to represent the whole industrial automation arena in this book. Industrial automation has a variety of vertical markets and there are also many types of SCADAs.

IEEE created a standard specification, called Std C37.1™, for SCADA and automation systems in 2007, targeting mostly power SCADA applications. It's recognized in the specification that in recent years, network- based industrial automation has greatly evolved with the use of intelligent electronic devices (IEDs), or IoT devices in our terms, in substations and power stations. The processing is now distributed, and functions that used to be done at the control center can now be done by the IED, that is, M2M between devices. Despite the fact that many functions can be moved to the IED, utilities still need a master station, the IoT platform, for the operation of the power system. Due to the restructuring of the electric industry, traditional vertically integrated electric utilities are replaced by many entities such as GENCO (Generation Company), TRANSCO (Transmission Company), DISCO (Distribution Company), ISO (independent system operator), RTO (regional transmission organization), and so forth. To fulfill their role, each of these entities needs a control center, that is, a substation, to receive and process data and take appropriate control actions.



IEEE Std. C37.1 SCADA Architecture.

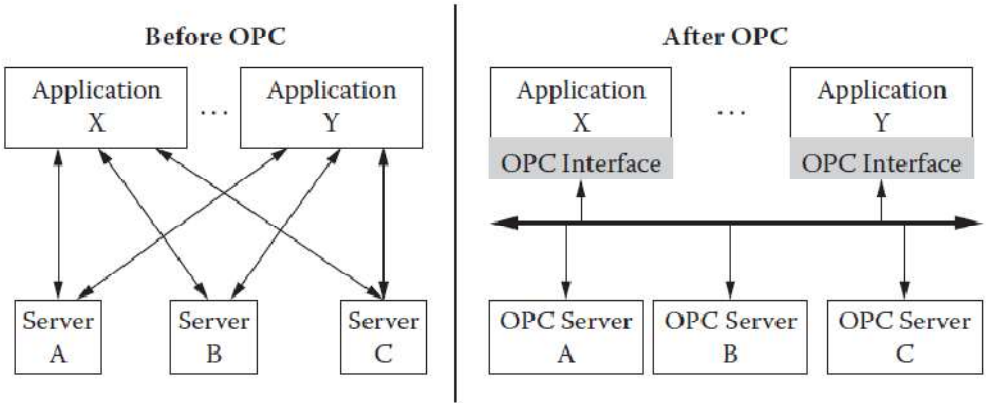
This specification addressed all levels of SCADA systems and covered the technologies used and, most importantly, the architecture of how those technologies interact and work together. However, no XML data formats and componentized architecture details are specified, which is perhaps why SCADA has long been regarded as a traditional control system market. People working in that area are often not aware of Internet- based IT innovations and cannot relate their work to a new concept such as IoT.

Wireless sensor systems have the potential to help industry use energy and materials more efficiently, lower production costs, and increase productivity. Although wireless technology has taken a major leap forward with the boom in wireless personal communications, applications for industrial field device systems must meet distinctly different challenges. That's where the ISA100, Wireless Systems for Industrial Automation, comes in. The ISA100 was developed by the standards committee of the Industrial Society for Automation, which was formed in 2005 to establish standards and related information that will define procedures for implementing wireless systems in the automation and control environment with a focus on the field level. The committee is made up of more than 400 automation professionals from nearly 250 companies around the world, lending their expertise from a variety of industrial backgrounds.

The ISA100 family of standards is designed with coexistence in mind, bringing peace of mind for the end user. We know that customers have other wireless solutions installed today and have the need for any future system to coexist with these installed systems. Therefore, the standards will feature technology to ensure the best performance possible in the presence of other wireless networks. For example, the ISA100 has created a new subcommittee to address options for convergence of the ISA100.11a and Wireless HART standards. This initiative is a key step in the mission of the ISA100

committee to develop a family of universal industrial wireless standards designed to satisfy the needs of end users across a variety of applications.

OPC, which stands for Object Linking and Embedding (OLE) for Process Control, is the original name for a standard specification developed in 1996 by an industrial automation industry task force. The standard specifies the communication of real- time plant data between control devices from different manufacturers (Figure 6.10). OPC is managed by the OPC Foundation [120] with more than 220 members worldwide including major firms in industrial automation, instruments manufacturers, building automation, and others.



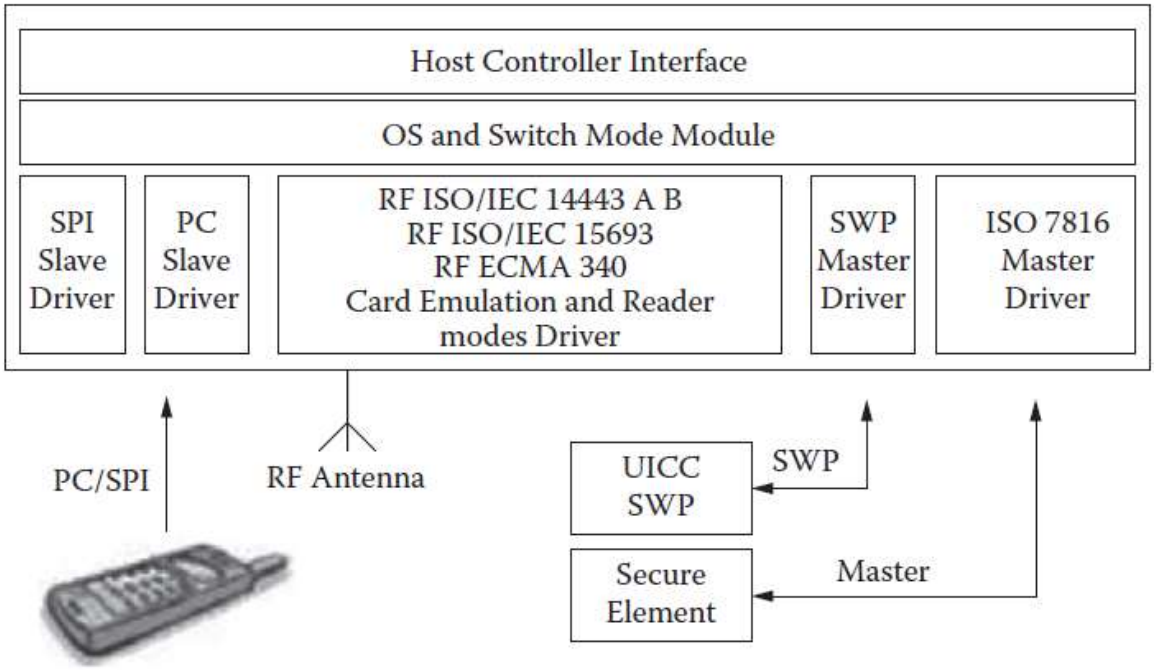
OPC originated from the DDE (dynamic data exchange) technologies based on DOS for PCs. The introduction of Windows 3.0 in 1990 made Windows an inexpensive, mainstream computing platform, providing the ability for a PC to run multiple applications simultaneously and a standard mechanism for those applications to exchange data at runtime. Wonder ware's InTouch™ SCADA software had the greatest impact for the transition from DDE to OPC. It introduced a means of networking DDE traffic (NetDDE™, which was later taken up by Microsoft) and also greatly increased the effective bandwidth of DDE by packing multiple data items into each packet or message. OLE (based on COM, common object model) and OCX (now ActiveX based on .NET) were launched in 1992. A number of SCADA vendors saw the chance to standardize the interface between the SCADA core and the device drivers that were actually responsible for acquiring the data, and the first- draft version of the OPC specification was released in December 1995 by the OPC Foundation sponsored by Microsoft.

OPC was designed to provide a common bridge for Windows- based software applications and process control hardware. Standards define consistent methods of accessing field data from plant floor devices. This method remains the same regardless of the type and source of data. An OPC server for one hardware device provides the same methods for an OPC client to access its data as each and every other OPC server for that same or another hardware device. The aim was to reduce the amount of duplicated effort required from hardware manufacturers and their software partners, and from the SCADA and other HMI producers, in order to interface the two. When a hardware manufacturer had developed their OPC server for the new hardware device, their work was done to allow anyone to access their device; and when the SCADA producer had developed their OPC client, their work was done to allow access to any hardware, existing or yet to be created, with an OPC- compliant server.

OPC has achieved great success in many application areas, most of them closely related to or part of IoT applications. However, OPC's success story is accompanied by some caveats. For example, standard OPC DA (data access) is based on Microsoft's COM and DCOM technology and is consequently restricted to the Windows operating system. In addition, DCOM communication is easily blocked by firewalls that prevent OPC clients from accessing data over a wide- area network and the World Wide Web. New approaches, such as XML- DA and United Architecture (UA) [234], have been developed to make OPC technology available on other platforms or accessible by other systems.

The RFID protocols and data formats are relatively well defined, mostly by EPC global, and unified compared with protocols and formats of the other three pillars of IoT. The RFID protocols (such as PML, Object Naming Service [ONS], Edgeware, EPC Information Service [EPCIS], Application Level Event [ALE], etc.) have been described in the previous chapters, so we will talk only about protocols for the related contactless smart cards here.
The smart cards with contactless interfaces (RFID is a subset) are becoming increasingly popular for payment and ticketing applications such as mass transit and stadiums. Visa and MasterCard have agreed to an easy- to- implement version deployed in the United States. Smart cards are also being introduced in personal identification and entitlement schemes at regional, national, and international levels. Citizen cards, drivers' licenses, and patient card schemes are becoming more prevalent. Some examples of widely used contactless smart cards are Taiwan's Easy Card, Hong Kong's Octopus card, Shanghai's Public Transportation Card, and Beijing's Municipal Administration and Communications Card.

The standard for contactless smart card communications is ISO/ IEC 14443. It defines two types of contactless cards (A and B) and allows for communications at distances up to 10 cm. An alternative standard for contactless smart cards is ISO/ IEC 15693, which allows communications at distances up to 50 cm (Figure 6.11).



iSO / IEC 14443/15693 smart card standards