12 The Internet of Things (IOT)

12.1 Moving toward a Smarter Internet

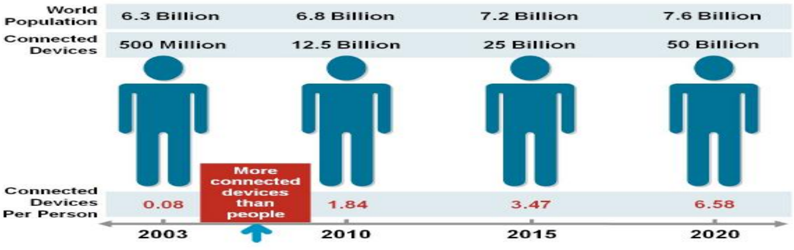
Imagine a world where billions of objects can sense, communicate and share information, all interconnected over public or private Internet Protocol (IP) networks. These interconnected objects have data regularly collected, analysed and used to initiate action, providing a wealth of intelligence for planning, management and decision making. This is the world of the Internet of Things (IOT).

The IOT concept was coined by a member of the Radio Frequency Identification (RFID) development community in 1999, and it has recently become more relevant to the practical world largely because of the growth of mobile devices, embedded and ubiquitous communication, cloud computing and data analytics.

Since then, many visionaries have seized on the phrase “Internet of Things” to refer to the general idea of things, especially everyday objects, that are readable, recognisable, locatable, addressable, and/or controllable via the Internet, irrespective of the communication means (whether via RFID, wireless LAN, wide- area networks, or other means). Everyday objects include not only the electronic devices we encounter or the products of higher technological development such as vehicles and equipment but things that we do not ordinarily think of as electronic at all - such as food and clothing. Examples of “things” include:

- People;
- Location (of objects);
- Time Information (of objects);
- Condition (of objects).

These “things” of the real world shall seamlessly integrate into the virtual world, enabling anytime, anywhere connectivity. In 2010, the number of everyday physical objects and devices connected to the Internet was around 12.5 billion. Cisco forecasts that this figure is expected to double to 25 billion in 2015 as the number of more smart devices per person increases, and to a further 50 billion by 2020 (see Figure 1).



Source: Cisco IBSG, April 2011

Figure 1: Number of connected devices by 2020[1]

---

[1] Dave Evans. The Internet of Things: How the Next Evolution of the Internet Is Changing Everything. [Online] Available from: http://postscapes.com/cisco-internet-of-things-white-paper-how-the-next-evolution-of-the-internet-is-changing-everything%20 [Accessed 9th July 2012].

With more physical objects and smart devices connected in the IOT landscape, the impact and value that IOT brings to our daily lives become more prevalent. People make better decisions such as taking the best routes to work or choosing their favourite restaurant. New services can emerge to address society challenges such as remote health monitoring for elderly patients and pay-as-you-use services. For government, the convergence of data sources on shared networks improves nationwide planning, promotes better coordination between agencies and facilitates quicker responsiveness to emergencies and disasters. For enterprises, IOT brings about tangible business benefits from improved management and tracking of assets and products, new business models and cost savings achieved through the optimisation of equipment and resource usage.

12.2 Market Trends

In today’s IT industry, companies are staying competitive by adopting new technologies, streamlining business processes and innovating new services to increase productivity and save costs.

In the logistics and supply chain, the traditional supply of goods is based on established agreements between manufacturers and suppliers. Orders are made in advance and tracking is done by various stakeholders in the supply chain, i.e., assembly lines, manufacturers and logistics managers. With the use of smart technologies such as active RFID (executable codes in tag), it is possible to envision that goods may be transported without human intervention from manufacturers to suppliers. Warehouses will become completely automatic with goods moving in and out; forwarding of the goods will be made, using intelligent decisions based on information received via readers and positioning systems to optimise transiting routes. Suppliers will have the flexibility to purchase parts from various manufacturers (possibly from competing manufacturers) and buy them in a sequence of individual orders. Such automation creates a dynamic production and transportation network and provides better asset management to improve the overall efficiency in the supply chain.

In healthcare, hospitals are shifting from providing healthcare on premise, i.e., in hospitals and clinics, to remote self-monitoring for patients. Self-monitoring benefits patients by giving them greater freedom and independence in monitoring their health and frees up hospital equipment for the treatment of emergencies. In the USA, electronic health monitoring has been given the go-ahead by the Federal Communications Commission (FCC). FCC allows the use of allotted frequencies for sensors to control devices wirelessly in the monitoring of health at hospitals and homes. Such monitoring allows doctors to inform their patients of critical conditions before they happen and subsequently improves the quality of healthcare by untethering patients from tubes and wires. FCC has also forecast savings of an average of US$12,000 per patient[2] by decreasing hospital-acquired infections. Moving into the future, there are newer trends of developing biodegradable materials for sensors and “lab-on-chip” equipment that can be implanted on or in patients. The sensor chips can detect internal organ responses to new medication and guide the application of drugs to infected areas for better treatment.

In smart grid and metering, smart grid systems allow the monitoring and managing of the entire life cycle of power generation, transmission, distribution and consumption. Consumers traditionally do not have control over their exact consumption of power but are now empowered to manage and track their own consumption. This shift potentially creates huge saving for consumers and also for power companies as they are able to provision power at peak periods of the day. Frost and Sullivan has forecast 5% cost savings[3] from changes in consumption patterns resulting from the ability to monitor consumption habits

---

[2] Tam Harbert. FCC Gives Medical Body Area Networks Clean Bill of Health. [Online] Available from: http://spectrum.ieee.org/tech-talk/biomedical/devices/fcc-gives-medical-body-area-networks-clean-bill-of-health [Accessed 9th July 2012].

[3] Frost and Sullivan IOT cost saving data provided to IDA

for consumers, and 10% cost savings[4] on passive energy efficiencies related to smart grid implementation, e.g., diagnostic capability, conservation via voltage reduction and control, measurement and verification for efficiency. For example, in South Korea, a smart grid test-bedding project[5] is currently being trialled on Jeju Island where it will become the world's largest smart grid community to conduct testing of the most advanced smart grid sensor technologies and R&D results. The target is to achieve a 30% reduction of CO2 by 2020, and achieve a low carbon economy and society capable of monitoring power consumption and distribution.

In automotive transportation, the traffic conditions today are monitored by cameras and motion sensors placed along major road junctions and highways. However, with road traffic growing and land space for road development restricted, these sensing technologies are reaching their limits in providing real-time traffic updates to ease road congestions and help prevent accidents. There are shifting trends in the automotive industry to equip vehicles with dedicated short-range communication (DSRC) to provide vehicle-to-vehicle (V2V) communications to improve vehicle safety and provide better road visibility for traffic management. For instance, when there is a traffic jam, the first car may tell the cars behind if there is an accident, and this will eventually inform the intelligent navigation systems to re-route the path to another less crowded road. These cars can make breakdown calls when appropriate, collecting data about the surrounding infrastructures such as traffic lights and buildings, and about itself (such as the faulty parts in the vehicle and type of loads it is carrying) in the event of an emergency. Vehicles gradually become smart "things" which can react, based on real-time situations on the roads, and contribute to a safer traffic system.

In retail, businesses have problems identifying the right customer at the right time to sell them their products. Various techniques of marketing products involve using short messaging system (SMS) broadcast, digital signages and recently the use of Quick Response (QR) codes to bundle promotions. These methods often fail to deliver the right customer to the right product and vice-versa. New trends of marketing have evolved with businesses shifting from mass market advertising to context-aware[6] systems to anticipate customer needs and proactively serve the most appropriate products or services. For example, a male shopper, looking to buy business suits for a job interview, will be informed of exact store locations selling suits that match his body size, style and budget. Behind the scene, the context-aware system tries to understand the profile and sentiments of the male shopper, and combines data from the mall to "intelligently" make recommendations to suit the shopper. Gartner has forecast that context-aware technologies will affect US$96 billion[7] of annual consumer spending by 2015, with 15% of all payment card transactions being made on the back of contextual information.

---

[4] Frost and Sullivan IOT cost saving data provided to IDA

[5] Korea Smart Grid Institute. Korea's Jeju Smart Grid Test-bed Overview. [Online] Available from: http://www.smartgrid.or.kr/10eng3-1.php [Accessed 9th July 2012].

[6] Rsasirekha. Gartner's IT Top Predictions for 2011. [Online] Available from: http://itknowledgeexchange.techtarget.com/enterprise-IT-tech-trends/gartners-it-top-predictions-for-2011/ [Accessed 9th July 2012].

[7] Retail Pro. Gartner: Context-aware technology to change consumer spending. [Online] Available from: http://www.retailpro.com/community/blog/index.php/2011/10/31/gartner-context-aware-technology-to-change-consumer-spending/ [Accessed 9th July 2012].

The use of RFID and near field communications (NCF) tags on packages, shelves and payment counters is also being gradually adopted by businesses to enhance retail experiences. It is estimated that an estimated 2 billion phones will be sold by 2012. Almost every phone will have RFID and NFC readers, meaning that eventually shoppers will no longer need to consult salespersons or floor readers to know the history of a product. They can simply scan the product tags using their mobile phones (or the shelves if the products are sold out). Virtual shopping carts can be created and orders placed automatically with warehouses for goods to be delivered to their homes.

12.3 Technology Trends

Several technology trends will help shape IOT. Here are seven identified macro trends: the miniaturisation of devices, advances in RFID technologies, Internet Protocol version Six (IPv6), improvements in communication throughput and latency, real-time analytics, adoption of cloud technologies and security.

12.3.1 Miniaturisation of devices

IOT uses technologies to connect physical objects to the Internet. The size (and cost) of electronic components that are needed to support capabilities such as sensing, tracking and control mechanisms, play a critical role in the widespread adoption of IOT for various industry applications. The progress in the semiconductor industry has been no less than spectacular, as the industry has kept true to Moore's Law of doubling transistor density every two years.

In 2000, the state of the art was 1,000 nanometers (nm) but from 2010 to 2011, the industry shifted to commercially available System-on-Chip (SoC) chip solutions that utilise 28 nm - 45 nm lithography to achieve a 2-3 chipset package that can integrate an entire radio transceiver complete with digital signal processing, baseband microprocessors or graphic accelerators. Many applications such as remote healthcare and environmental monitoring require these integrated chipsets to be not only small but also concealable and to act as “tiny” computers to sense the physical subjects. Fortunately, the miniaturisation of devices has been taking place at lightning speed and the number of transistors manufactured per die has increased exponentially over the years (Figure 2).

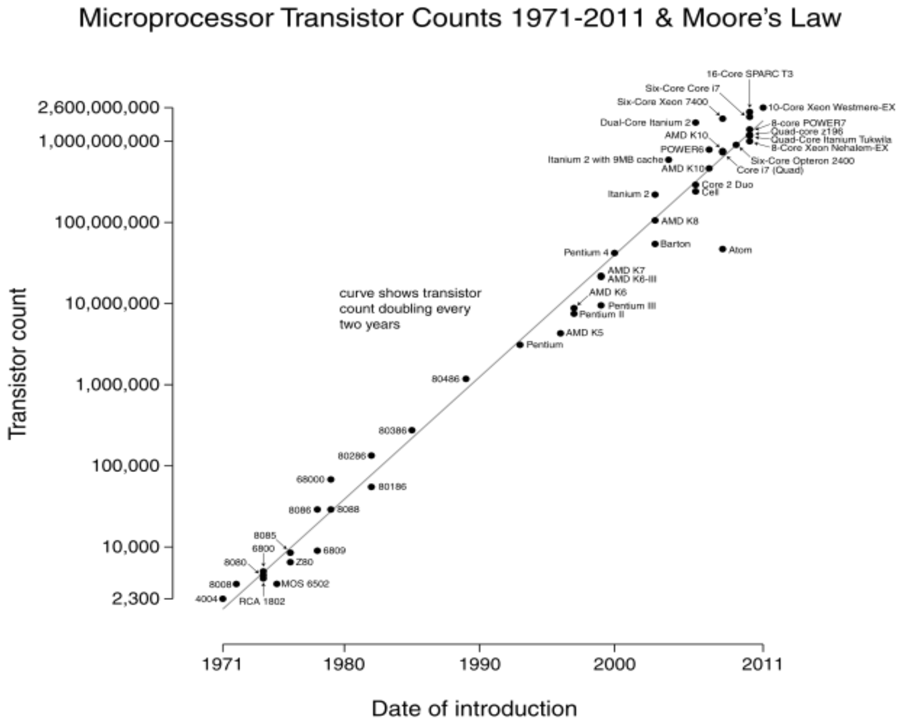


Figure 2: Moore's Law[8]

Today, wafer chip manufacturing technology is predominantly driven by planar metal oxide semiconductor field-effect transistor (MOSFET) technology. Advances in the areas of chip design and architecture have allowed semiconductor industries to reduce the size, density and cost of production of transistors. Technologies such as lithography, metrology and nanotechnology are used (and explored) to dramatically increase the number of transistors to be fabricated on a single chip. For example, semiconductor manufacturing processes have also improved from the current 32 nanometer (nm) node in 2010 to the 22nm node in late 2011, moving to 16nm by 2013 and 11 nm by 2015 (see Figure 3). Intel[9], in April 2012, officially launched the world's first commercial microprocessor - "Ivy Bridge" on the 22nm wafer, using 3-D Tri-Gate technology. This 22nm wafer is capable of fitting more than 2.9 billion transistors with 37% improvements in performance and more than 50% power reduction, compared to its predecessor transistors.

---

[8] Wikipedia. Moore's law. [Online] Available from: http://en.wikipedia.org/wiki/Moore%27s_law [Accessed 9th July 2012].

[9] Intel. Intel 22nm Technology. [Online] Available from: http://www.intel.com/content/www/us/en/silicon-innovations/intel-22nm-technology.html [Accessed 9th July 2012].

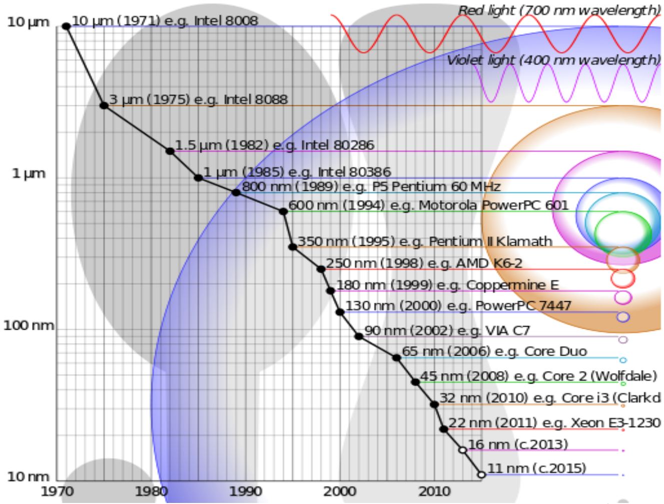
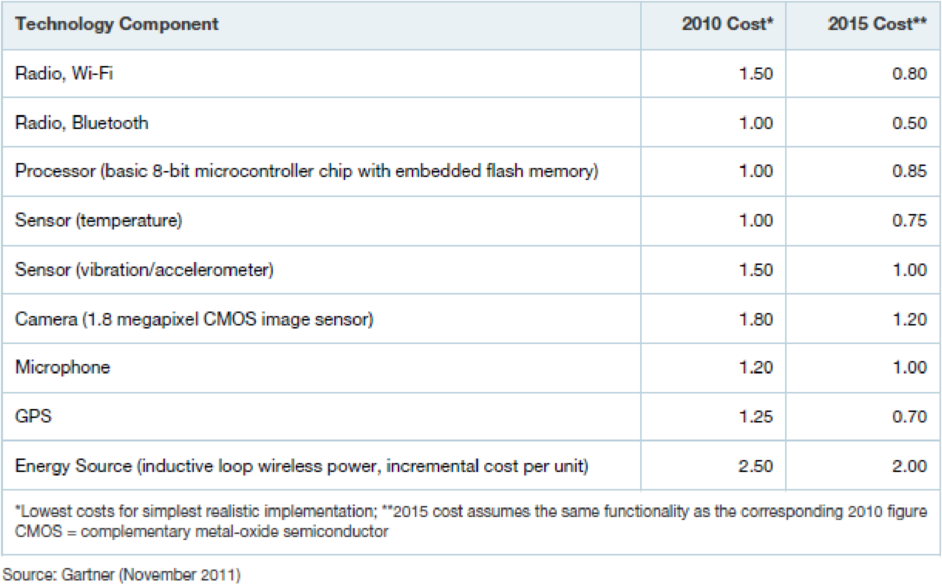


Figure 3: Semiconductor device fabrication trends[10]

Just as the size of the chips is getting smaller, the costs of sensing components are also dropping to become more affordable. Gartner has forecast that most technology components such as radio, WiFi, sensors and global positioning systems (GPS), could see a drop in cost of 15% to 45% from 2010 to 2015 (see Figure 4). To illustrate, with cheaper temperature sensors, cold chain retailers would consider deploying more temperature sensors to monitor their perishable produce as it traverses the supply chain.



Technology Component	2010 Cost*	2015 Cost**
Radio, Wi-Fi	1.50	0.80
Radio, Bluetooth	1.00	0.50
Processor (basic 8-bit microcontroller chip with embedded flash memory)	1.00	0.85
Sensor (temperature)	1.00	0.75
Sensor (vibration/accelerometer)	1.50	1.00
Camera (1.8 megapixel CMOS image sensor)	1.80	1.20
Microphone	1.20	1.00
GPS	1.25	0.70
Energy Source (inductive loop wireless power, incremental cost per unit)	2.50	2.00

*Lowest costs for simplest realistic implementation; **2015 cost assumes the same functionality as the corresponding 2010 figure
CMOS = complementary metal-oxide semiconductor

Source: Gartner (November 2011)

Figure 4: Falling cost of technology components[11]

With the decreasing size and falling cost of technology components, organisations will see greater savings and opportunities in pursuing IOT in the next one to three years.

---

[10] Wikipedia. Semiconductor device fabrication. [Online] Available from: http://en.wikipedia.org/wiki/Semiconductor_device_fabrication [Accessed 9th July 2012].

[11] Jim Tully, Hung LeHon g. Falling Cost of Components Will Help Drive The Pervasiveness of the Internet of Things. [Online] Available from: http://www.gartner.com/id=1839420 [Accessed 9th July 2012].

12.3.2 Radio Frequency Identification (RFID)

Radio Frequency Identification (RFID) technology is of particular importance to IOT as one of the first industrial realisations of IOT is in the use of RFID technology to track and monitor goods in the logistics and supply chain sector. RFID frequency bands range from 125 kHz (low frequency/LF) up to 5.8 Ghz/super high frequency (SHF) and the tags have at least three basic components:

- The chip holds information about the object to which it is attached and transfers the data to reader wirelessly via an air interface.
- The antenna allows transmission of the information to/from a reader.
- The packaging encases chip and antenna, and allows the attaching of the tag to an object for identification.

Today, the one dimension bar (ID) code has made a significant contribution to the supply chain and other businesses such as asset management. Two dimension (2D) bar codes have provided a richer source of data but, once printed, are not up-datable. RFID, with its ability to permanently collect and process data in its environment, is proving to be the next technology for the identification of goods. Many industry verticals, especially in the logistics and supply chain, have been using RFID as tagging solutions to improve their tracking and monitoring processes.

Moving into the future, RFID has the potential to provide streams of data that will provide information systems with real-time, item-specific data and be flexible enough to be placed in extremely small spaces and locations, i.e., coil-on-chip technology. With technology developments in areas such as chip design, energy usage and preservation, RF technologies and manufacturing, new ways of RFID usage will emerge for applications such as automatic meter reading, remote home automation and real-time vehicle tracking.

12.3.3 Internet Protocol version 6 (IPv6)

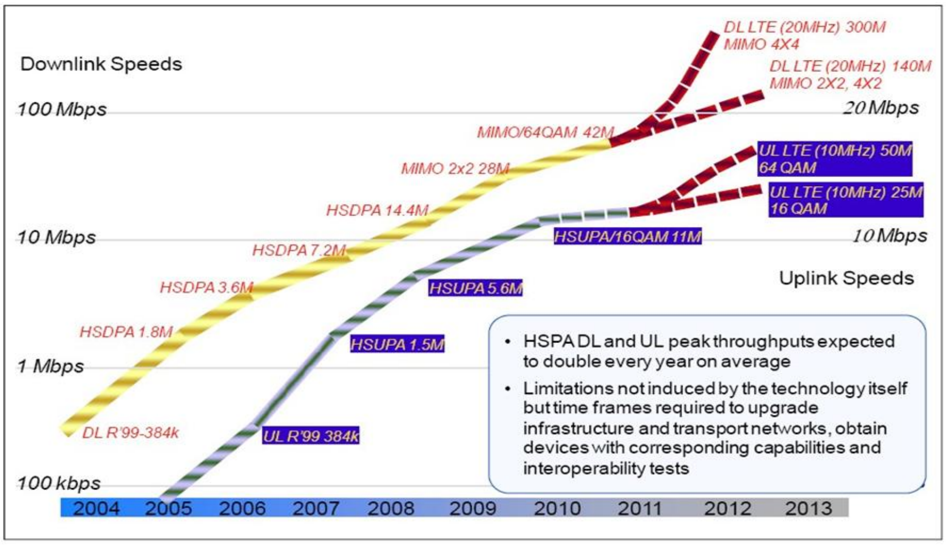
The IPv4 address pool is effectively exhausted, according to industry accepted indicators. The final allocations under the existing framework have now been made, triggering the processes for the Internet Assigned Numbers Authority (IANA) to assign the final five IPv4/8 blocks, one to each of the five regional registries. With the exhaustion of the IANA pool of IPv4 addresses, no further IPv4 addresses can be issued to the regional registries that provide addresses to organisations.

IPv6 is the next Internet addressing protocol that is used to replace IPv4. With IPv6, there are approximately 3.4×10^{38} (340 trillion trillion trillion) unique IPv6 addresses, allowing the Internet to continue to grow and innovate. Given the huge number of connected devices (50 billion), IPv6 can potentially be used to address all these devices (and systems), eliminating the need of network address translation (NAT) and promoting end-to-end connectivity and control. These features provide seamless integration of physical objects into the Internet world.

12.3.4 Increasing communication throughput and lower latency

IOT relies on a pervasive communication network to allow “everything and everywhere” connectivity to occur. Over the years, network operators have been enhancing their infrastructure to support data capability and improving network throughput for their existing cell sites, transceivers, and interconnection facilities. With the addition of General Packet Radio Service (GPRS) infrastructure, Global System for Mobile (GSM) operators have largely upgraded their data services to Enhanced Data rates for GSM Evolution (EDGE). Today, most operators worldwide are deploying Universal Mobile Telecommunications System (UMTS) with High Speed Packet Access (HSPA) technology for higher throughput and low latency. HSPA, also commonly known as “3G”, has also shown us the power and potential of always-on, everyplace network connectivity that has ignited a massive wave of industry innovation that spans devices and applications.

As the technology trend shifts towards providing faster data rates and lower latency connectivity (see Figures 5 and 6), the Third Generation Partnership Project (3GPP) standards body has developed a series of enhancements to create the “HSPA Evolution”, also referred to as “HSPA+”. HSPA Evolution represents a logical development of the Wideband Code Division Multiple Access (WCDMA) approach, and is the stepping stone to an entirely new 3GPP radio platform called 3GPP Long Term Evolution (LTE). LTE offers a number of distinct advantages such as increased performance attributes, high peak data rates, low latency and greater efficiencies in using the wireless spectrum.



- HSPA DL and UL peak throughputs expected to double every year on average
- Limitations not induced by the technology itself but time frames required to upgrade infrastructure and transport networks, obtain devices with corresponding capabilities and interoperability tests

Figure 5: Higher throughput in communication technologies[12]

---

[12] Rysavy Research. Mobile Broadband Explosion. [Online] Available from: http://www.rysavy.com/Articles/2011_09_08_Mobile_Broadband_Explosion.pdf [Accessed 9th July 2012].

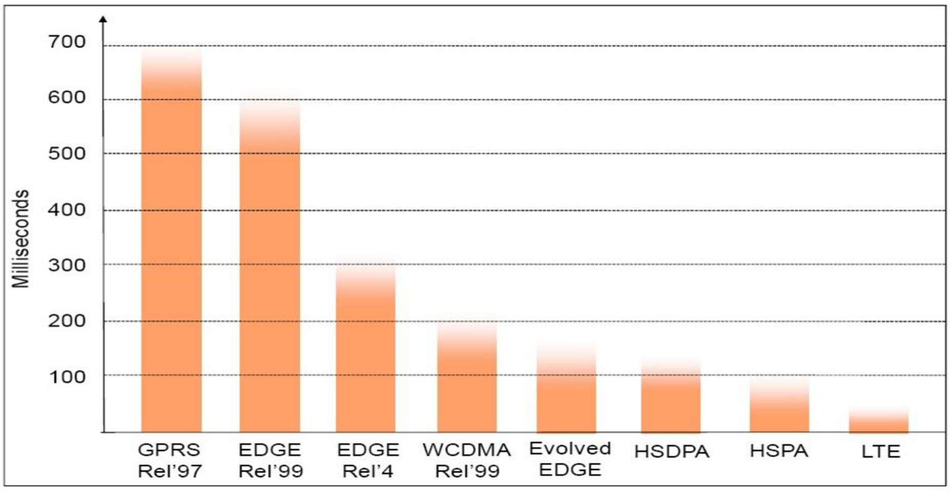


Figure 6: Lower latency in communication technologies[13]

Low latency makes it possible for IOT applications to query or receive quicker updates from sensor devices. LTE networks have latencies on the order of 50-75 ms which will open up new types of programming possibilities for application developers. For example, wearable computers which require interactive and real-time feedback will require moving large chunks of data to be analysed in the cloud or back-end systems to create a seamless user experience. Higher peak data rates can support applications such as Voice over IP (VoIP) and digital video that require better quality of service (QoS). With further advancements in communication technologies such as Software Define Radio (SDR) and Long Term Evolution-Advanced (LTE-A), devices will be able to communicate with better QoS and support better access to new services with more efficient use of the radio frequency spectrum.

12.3.5 Real-time Analytics

In today's decision making process, the availability of real-time, accurate information is crucial. With the growing volume of data from connected devices, social media platforms, etc, good decision making relies heavily on advances in analytic capability technologies, to bring out the intelligence in data. Traditional analytics is performed as a back-end resource and is done with the pre-creation of metadata before the actual analytics process takes place. The modelling of the metadata depends on the analytics requirements; with new project requirements, the metadata has to be re-modelled.

New forms of analytics have emerged to remove the need to pre-model metadata, resulting in faster query and more dynamic data processing. In-memory processing is a form of analytics where detailed data is loaded into the system memory from a variety of data sources. New data is analysed and stored in the system memory to improve the relevance of the analytics content to augment the speed in decision making. Companies such as SAP[14],

---

[13] Rysavy Research. Mobile Broadband Explosion. [Online] Available from: http://www.rysavy.com/Articles/2011_09_08_Mobile_Broadband_Explosion.pdf [Accessed 9th July 2012].

[14] SAP. SAP HANA redefines In-Memory. [Online] Available from: http://www.sap.com/solutions/technology/in-memory-computing-platform/index.epx [Accessed 9th July 2012].

Microsoft[15], IBM[16] and Teradata[17] are building in-memory database solutions that can perform high analytical and transactional processing. Another form of real-time analytics such as streaming analytics uses complex algorithms to instantaneously process streams of event data it receives from one or more sources. Some examples for IOT applications that require streaming analytics could be road traffic data and telephone conversations.

IOT creates opportunities for analytics to be performed in real time and also allows large volumes of data to be stored for analysis at a later time.

12.3.6 Cloud Computing

IOT connects billions of devices and sensors to create new and innovative applications. In order to support these applications, a reliable, elastic and agile platform is essential. Cloud computing is one of the enabling platforms to support IOT.

Cloud computing is an architecture that orchestrates various technology capabilities such as multi-tenancy, automated provisioning and usage accounting while relying on the Internet and other connectivity technologies like richer Web browsers to realise the vision of computing delivered as a utility. Cloud computing is seeing growing adoption (Figure 7) and there are three commonly deployed cloud service models namely Cloud Software as a Service (SaaS), Cloud Platform as a Service (PaaS) and Cloud Infrastructure as a Service (IaaS). For example, in IaaS, the use of hardware such as sensors and actuators can be made available to consumers as cloud resources. Consumers can set up arbitrary services and manage the hardware via cloud resource access control. PaaS can provide a platform from which to access IOT data and on which custom IOT applications (or host-acquired IOT applications) can be developed. SaaS can be provided on top of the PaaS solutions to offer the provider's own SaaS platform for specific IOT domains. Companies such as Axeda[18], ThingWorx[19], DeviceWise[20] are already providing software development platform to build innovative M2M and IOT applications.

---

[15] Wikipedia. Microsoft SQL Server. [Online] Available from: http://en.wikipedia.org/wiki/Microsoft_SQL_Server [Accessed 9th July 2012].

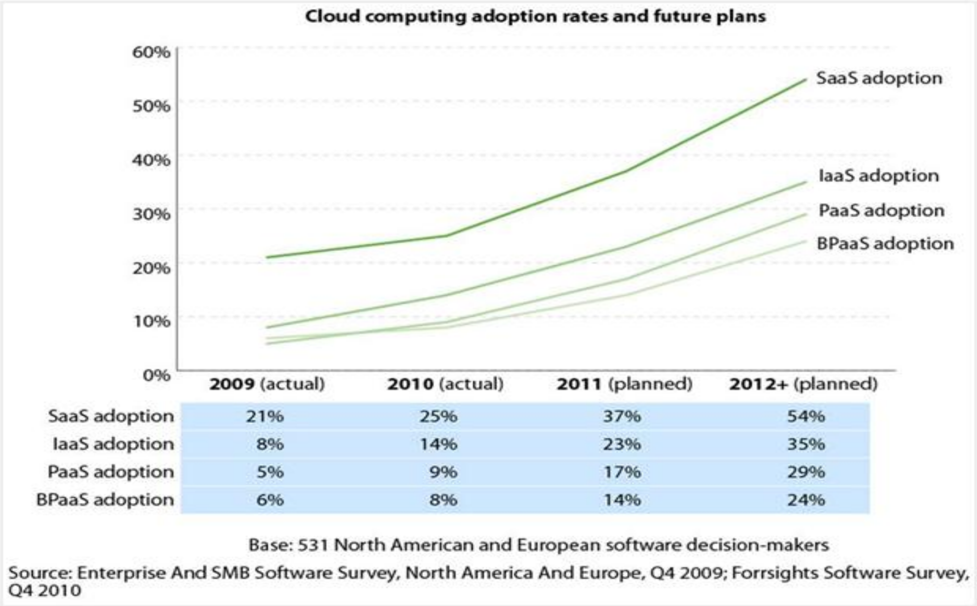
[16] IBM. IBM solidDB – Fastest Data Delivery. [Online] Available from: http://www-01.ibm.com/software/data/soliddb/ [Accessed 9th July 2012].

[17] Teradata. Teradata Database. [Online] Available from: http://www.teradata.com/products-and-services/database/teradata-14/ [Accessed 9th July 2012].

[18] Axeda. Axeda Application and Data Integration Platform. [Online] Available from: http://www.axeda.com/products/platform [Accessed 9th July 2012].

[19] ThingWorx. Platform Overview. [Online] Available from: http://www.thingworx.com/platform/ [Accessed 9th July 2012].

[20] DeviceWise. Platform - Overview. [Online] Available from: http://www.devicewise.com/platform/overview/ [Accessed 9th July 2012].


Cloud computing adoption rates and future plans

	2009 (actual)	2010 (actual)	2011 (planned)	2012+ (planned)
SaaS adoption	21%	25%	37%	54%
IaaS adoption	8%	14%	23%	35%
PaaS adoption	5%	9%	17%	29%
BPaaS adoption	6%	8%	14%	24%

Base: 531 North American and European software decision-makers

Source: Enterprise And SMB Software Survey, North America And Europe, Q4 2009; Forrsights Software Survey, Q4 2010

Figure 7: Cloud computing adoption rates[21]

12.3.6 Security and Privacy

Today, various encryption and authentication technologies such as Rivest Shamir Adleman (RSA) and message authentication code (MAC) protect the confidentiality and authenticity of transaction data as it “transits” between networks. Encryptions such as full disk encryption (FDE) is also performed for user data “at rest” to prevent unauthorised access and data tampering.

In future, new standards and technologies should address security and privacy features for users, network, data and applications. In areas of network protocol security, Internet Protocol Version 6 (IPv6) is the next generation protocol for the Internet; it contains addressing and security control information, i.e., IPSec to route packets through the Internet. In IPv4, IPSec is optional and connecting computers (peers) do not necessarily support IPsec. With IPv6, IPSec support is integrated into the protocol design and connections can be secured when communicating with other IPv6 devices. IPSec provides data confidentiality, data integrity and data authentication at the network layer, and offers various security services at the IP layer and above. These security services are, for example, access control, connectionless integrity, data origin authentication, protection against replays (a form of partial sequence integrity), confidentiality (encryption), and limited traffic flow confidentiality. Other IP-based security solutions such as Internet Key Exchange (IKEv2) and Host Identity Protocol (HIP) are also used to perform authenticated key exchanges over IPSec protocol for secure payload delivery.

---

[21] Louis Columbus. Predicting Cloud Computing Adoption Rates. [Online] Available from: http://softwarestrategiesblog.com/2011/07/24/predicting-cloud-computing-adoption-rates/ [Accessed 9th July 2012].

At the data link layer, Extensible Authentication Protocol (EAP) is an authentication framework used to support multiple authentication methods. It runs directly on the data link layer, and supports duplicate detection and re-transmission error. In order to enable network access authentication between clients and the network infrastructure, a Protocol for carrying Authentication for Network Access (PANA) forms the network-layer transport for EAP. In EAP terms, PANA is a User Datagram Protocol (UDP)-based EAP lower layer that runs between the EAP peer and the EAP authenticator.

For data privacy, policy approaches and technical implementations exist to ensure that sensitive data is removed or replaced with realistic data (not real data). Using policy approaches, Data Protection Acts are passed by various countries such as the USA and the European Union to safeguard an individual's personal data against misuse. For technical implementations, there are Privacy Enhancing Techniques (PETs) such as anonymisation and obfuscation to de-sensitize personal data. PETs use a variety of techniques such as data substitution, data hashing and truncation to break the sensitive association of data, so that the data is no longer personally identifiable and safe to use. For example, European Network and Information Security Agency (ENISA) has proposed to approach data privacy by design[22], using a “data masking” platform which uses PETs to ensure data privacy.

With the IOT-distributed nature of embedded devices in public areas, threats coming from networks trying to spoof data access, collection and privacy controls to allow the sharing of real-time information, IOT security has to be implemented on a strong foundation built on a holistic view of security for all IOT elements at various interacting layers.

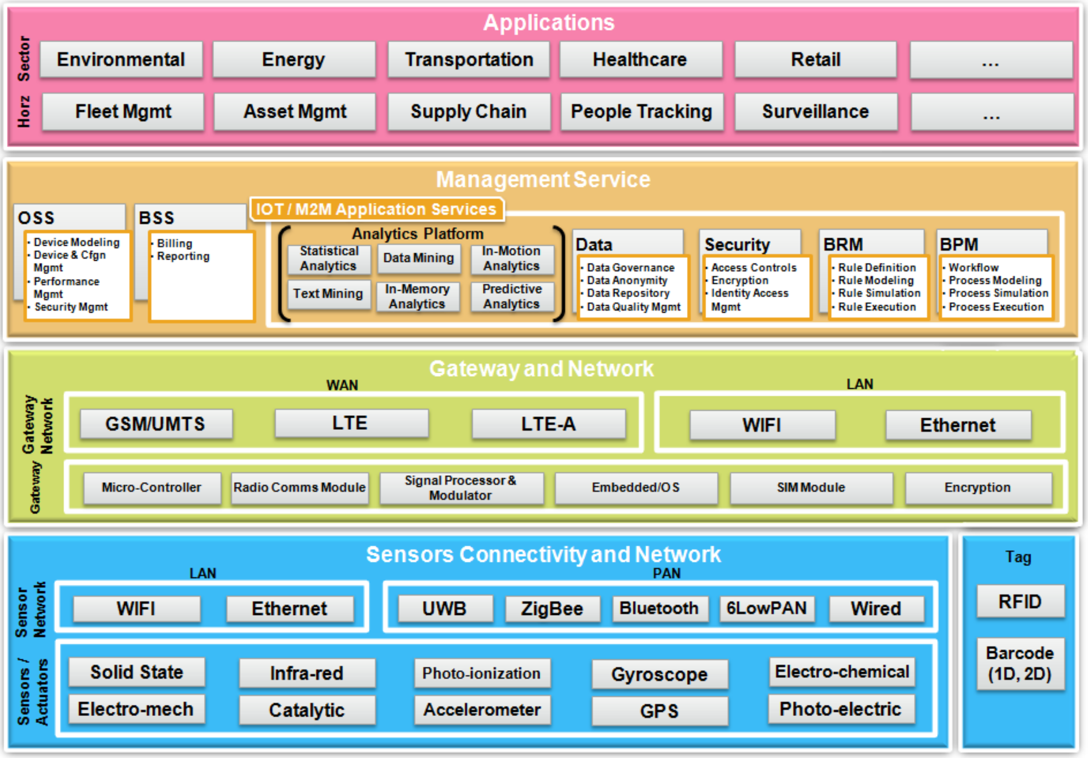
---

[22] European Network and Information Security Agency. Privacy, Accountability and Trust – Challenges and Opportunities. [Online] Available from http://www.enisa.europa.eu/activities/identity-and-trust/library/deliverables/pat-study [Accessed 9th July 2012].

12.4 IOT Architecture

IOT architecture consists of different suite of technologies supporting IOT. It serves to illustrate how various technologies relate to each other and to communicate the scalability, modularity and configuration of IOT deployments in different scenarios.

The functionality of each layer is described below:



Copyright Reserved by IDA

Figure 8: IOT Architecture

12.4.1 Sensor Layer

The lowest layer is made up of smart objects integrated with sensors. The sensors enable the interconnection of the physical and digital worlds allowing real-time information to be collected and processed. The miniaturisation of hardware has enabled powerful sensors to be produced in much smaller forms which are integrated into objects in the physical world.

There are various types of sensors for different purposes. The sensors have the capacity to take measurements such as temperature, air quality, movement and electricity. In some cases, they may also have a degree of memory, enabling them to record a certain number of measurements. A sensor can measure the physical property and convert it into signal that can be understood by an instrument. Sensors are grouped according to their unique purpose such as environmental sensors, body sensors, home appliance sensors and vehicle telematics sensors, etc.

Most sensors require connectivity to the sensor aggregators (gateways). This can be in the form of a Local Area Network (LAN) such as Ethernet and WiFi connections or Personal Area Network (PAN) such as ZigBee, Bluetooth and Ultra- Wideband (UWB). For sensors that do not require connectivity to sensor aggregators, their connectivity to backend servers/applications can be provided using Wide Area Network (WAN) such as GSM, GPRS and LTE. Sensors that use low power and low data rate connectivity, they typically form networks commonly known as wireless sensor networks (WSNs). WSNs are gaining popularity as they can accommodate far more sensor nodes while retaining adequate battery life and covering large areas.

12.4.2 Gateways and Networks

Massive volume of data will be produced by these tiny sensors and this requires a robust and high performance wired or wireless network infrastructure as a transport medium. Current networks, often tied with very different protocols, have been used to support machine-to-machine (M2M) networks and their applications.

With demand needed to serve a wider range of IOT services and applications such as high speed transactional services, context-aware applications, etc, multiple networks with various technologies and access protocols are needed to work with each other in a heterogeneous configuration. These networks can be in the form of a private, public or hybrid models and are built to support the communication requirements for latency, bandwidth or security.

A possible deployment could consist of a converged network infrastructure that resolves the fragmentation by integrating disparate networks into a single network platform. Converged network layer abstraction allows multiple organisations to share and use the same network independently for their information to be routed without compromising their privacy, security and performance requirements. Each organisation thus utilises the network as if it is a private network resource to them.

12.4.3 Management Service Layer

The management service renders the processing of information possible through analytics, security controls, process modelling and management of devices.

One of the important features of the management service layer is the business and process rule engines. IOT brings connection and interaction of objects and systems together providing information in the form of events or contextual data such as temperature of goods, current location and traffic data. Some of these events require filtering or routing to post-processing systems such as capturing of periodic sensory data, while others require response to the immediate situations such as reacting to emergencies on patient’s health conditions. The rule engines support the formulation of decision logics and trigger interactive and automated processes to enable a more responsive IOT system.

In the area of analytics, various analytics tools are used to extract relevant information from massive amount of raw data and to be processed at a much faster rate. Analytics such as in-memory analytics allows large volumes of data to be cached in random access memory (RAM) rather than stored in physical disks. In-memory analytics reduces data query time and augments the speed of decision making. Streaming analytics is another form of analytics where analysis of data, considered as data-in-motion, is required to be carried out in real time so that decisions can be made in a matter of seconds. For example, this requirement is typical in the transportation sector where real-time traffic information enables drivers to optimise their routes and travelling times.

Analytics can be carried out at other layers within the IOT architecture. For example, analytics may be carried out in the smart object layer, i.e., local hub or edge device, so that subsets of the information can be carried through the network for further processing. At this layer, analytics helps to reduce the stress placed on the network layer, reduce power needs of sensors by less frequent communication backend and allow faster responses to data received by the sensors.

Data management is the ability to manage data information flow. With data management in the management service layer, information can be accessed, integrated and controlled. Higher layer applications can be shielded from the need to process unnecessary data and reduce the risk of privacy disclosure of the data source. Data filtering techniques such as data anonymisation, data integration and data synchronisation, are used to hide the details of the information while providing only essential information that is usable for the relevant applications. With the use of data abstraction, information can be extracted to provide a common business view of data to gain greater agility and reuse across domains.

Lastly, security must be enforced across the whole dimension of the IOT architecture right from the smart object layer all the way to the application layer. Security is of the utmost importance as the integrity of the data must be protected as data travels across the entire system. The integrity of data enables reliable and authentic decisions to be made. Moreover, security of the system prevents system hacking and compromises by unauthorised personnel, thus reducing the possibility of risks.

12.4.4 Application Layer

There are various applications from industry sectors that can leverage on IOT. Applications can be verticalised ones that are specific to a particular industry sector, and other applications such as Fleet Management, Asset Tracking, and Surveillance can cut across multiple industry sectors. For a list of IOT applications, it is covered under "IOT Applications" in Chapter 12.6.2.

12.5 Technology Outlook

This section covers the various technologies that support IOT. While IOT is architected into layers, the technologies have been categorised into three groups.

The first group of technologies impacts the devices, microprocessor chips:

- Low power sensors for power and energy sustainability;
- Intelligence of sensors in the field;
- Miniaturisation of chipsets;
- Wireless sensor network for sensor connectivity.

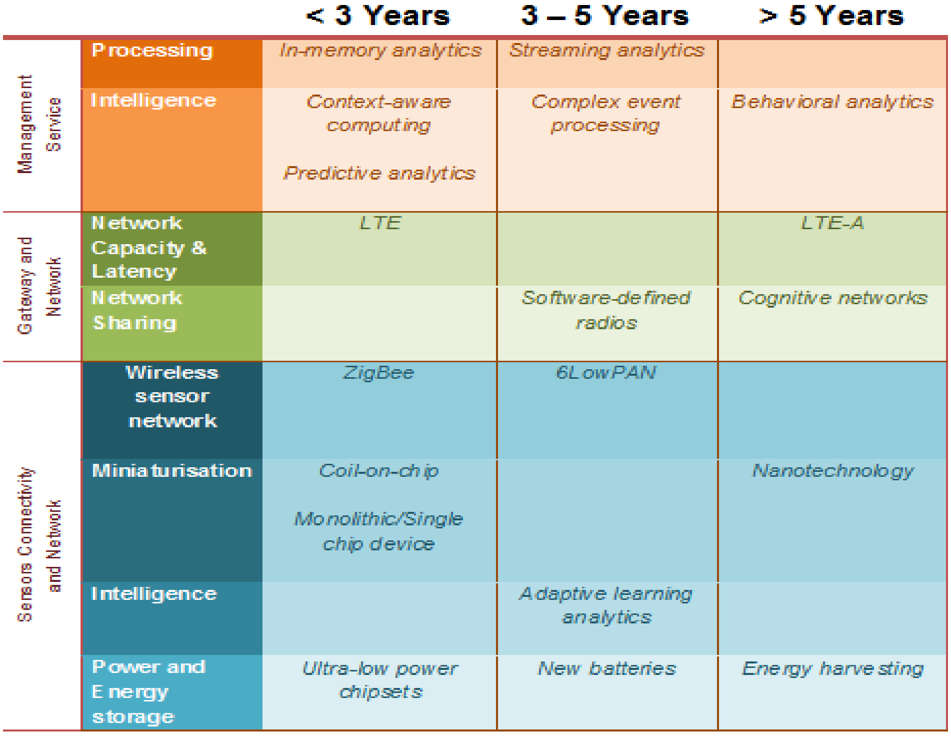
The second group comprises technologies that support network sharing and address capacity and latency issues:

- Network sharing technologies such as software-defined radios and cognitive networks;
- Network technologies that address capacity and latency issues such as LTE and LTE-A.

The third group impacts the management services that support the IOT applications:

- Intelligent decision-making technologies such as context-aware computing service, predictive analytics, complex event processing and behavioural analytics;
- Speed of data processing technologies such as in-memory and streaming analytics.

Technology Radar



		< 3 Years	3 – 5 Years	> 5 Years
Management Service	Processing	In-memory analytics	Streaming analytics	
	Intelligence	Context-aware computing Predictive analytics	Complex event processing	Behavioral analytics
Gateway and Network	Network Capacity & Latency	LTE		LTE-A
	Network Sharing		Software-defined radios	Cognitive networks
Sensors Connectivity and Network	Wireless sensor network	ZigBee	6LowPAN	
	Miniaturisation	Coil-on-chip Monolithic/Single chip device		Nanotechnology
	Intelligence		Adaptive learning analytics	
	Power and Energy storage	Ultra-low power chipsets	New batteries	Energy harvesting

12.5.1 Sensors Connectivity and Network

12.5.1.1 Less than three years

Coil-on-chip

Coil-on-chip technology allows the fabrication of antenna coil onto the surface of the silicon wafer to create chips that can sense and interact with readers via radio frequency. Such chips typically measure an approximate 2.5mm by 2.5mm and can be used on tiny objects or in areas that have extremely small spaces. The chip fabrication technology used is photolithography which creates very small structures with a high degree of precision and reproducibility.

Today, coil-on-a-chip technology has been implemented in certain RFID tags and specialised applications such as magnetic resonance micro-imaging. When compared to conventional ones with external antenna coils, the coil-on-chip RFID tag achieves a smaller footprint and rarely malfunctions because of the deterioration of contacts on the lack of external soldering connections between the antenna coils and the IC chips. RFID coil-on-chip has basic storage capacities ranging from 128 bytes to 4 kilobytes of data and no moving parts so it can withstand the harshest environments, including wet and dry conditions.

For IOT, coil-on-chip technology is particular useful as it allows small physical objects to be tagged with tiny coil-on-chip sensors to be monitored by applications. Some of these physical objects could be living things such as birds and insects that require monitoring on their immigration patterns due to weather. Maxell – a world leader in memory and storage technologies has developed coil-on-chip RFID tags that have read/write capabilities with a memory capacity of 1 kbit (Figure 9). The tiny 2.5mm by 2.5mm RFID chip allows data to be recorded, erased and re-recorded, and new data can be added until the memory capacity is filled. This makes the coil-on-chip a good choice for long-term data management over a project lifecycle. The data on the chip can either be erased and re-used, or saved for archiving after processing.

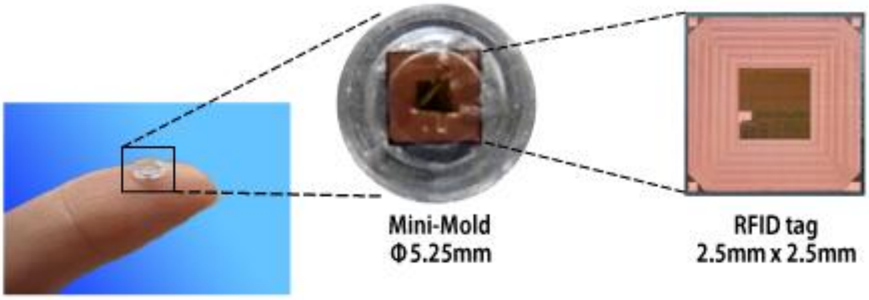


Figure 9: Hitachi Maxell's coil-on-chip RFID[23]

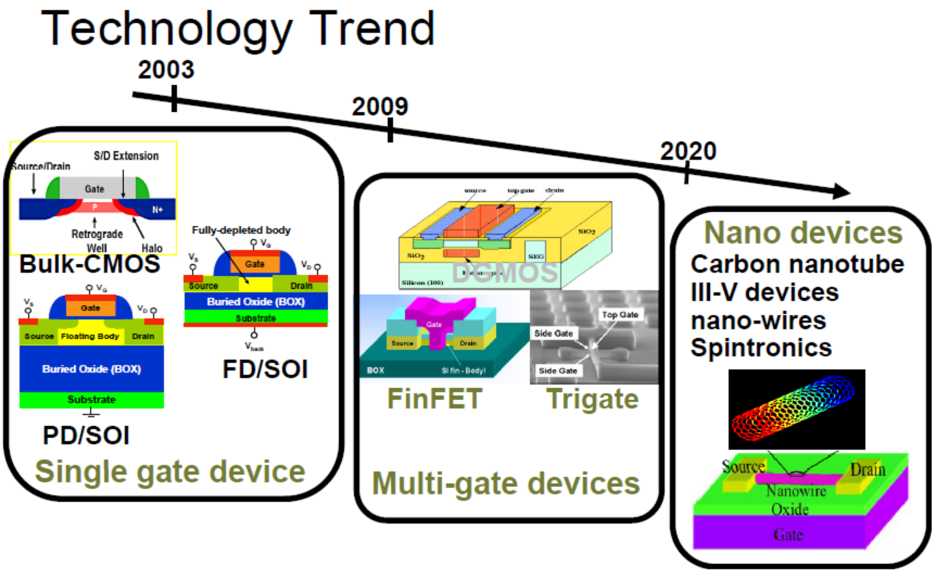
---

[23] Maxell. Coil-on-Chip RFID. [Online] Available from: http://biz.maxell.com/en/product_security/?pci=7&pn=sp0006 [Accessed 9th July 2012].

Low-power devices and batteries

Power consumption has the greatest challenge for sensors. Today sensors need to be able to sustain longer battery lifespan, especially in cases such as outdoor deployments, to shorten hardware maintenance and prevent breakdown of communication. In many deployment cases, in order to prolong the usability of the sensors in the field, large battery sources have to be attached to the sensors, making the sensor setup bulky and cumbersome.

IOT supports the pervasive connectivity of sensors and the need for them to interact with each other i.e., act as both tags and interrogators. In order to support such connectivity and communications, the design and use of low-power chipsets will create a significant impact and consideration on power consumption for future sensors. Ultra-low power designs for chipset circuits have been an ongoing research area, with techniques moving from single gate to multi-gate transistors and to carbon nanotube designs (Figure 10).



Source: Kaushik Roy, Purdue University

Figure 10: Low-power chipset designs[24]

Power scavenging (energy harvesting) technologies that convert energy out of physical energy sources such as temperature differences and applied pressure, have been researched to explore their capability to replace conventional batteries. Two examples of power scavenging technologies are photovoltaic technology which generates electric power by using solar energy and piezoelectrics technology that creates charges on stress or shape change on voltage applied. Newer forms of battery technologies, e.g., polymer battery, fuel-cell and paper batteries will support increasing functionality and longer battery lifetime. Paper and smart label batteries have shown promising use cases in warehousing usage as they allow containers to perform two-way communications with the reader.

---

[24] Mike Keating. The Future of Low Power. [Online] Available from: http://www.synopsys.com/Solutions/EndSolutions/EclypseSolutions/CapsuleModule/future_lp08_mike_keating.pdf [Accessed 9th July 2012].

Monolithic/Single Chip Device

In many sensing applications today (e.g., motion detectors and ambient intelligence systems), there is an increasing need for sensors and tags to exchange data with other devices such as tags, sensors, network nodes and routers. In order for these sensors and tags to communicate heterogeneously across various communication protocols, chip design is allowing additional RF components (e.g., for Bluetooth, ZigBee, Wireless LAN and FM functionality) to be part of the monolithic/single chip device.

The definition of a monolithic chip is a type of “integrated circuit” electronics that contains active and passive devices (transistors, microcontrollers and capacitors) that are made in and on the surface of a single piece of a silicon wafer. The “planar technology” used in a single block (monolith) allows the block to interconnect with the insulating layer over the same body of the semiconductor to produce a solid integral monolithic chip. If the devices are interconnected by bonding wires dangling above the chip, it is not a monolithic chip but a hybrid chip.

With monolithic chips, applications can not only communicate with the reader but are also able to exchange data with other devices (tags, sensors, network nodes, routers, etc). Such designs will lead to a cost-effective solution for industries looking to integrate sensors with communication devices such as Personal Digital Assistants (PDAs), mobile phones, notebooks and navigation systems.

ZigBee (used in the context of Wireless Sensor Networks)

Wireless Sensor Network (WSN) or Ubiquitous Sensor Network (USN) is defined by the ability of sensors (often called nodes) to communicate directly with each other to form a mesh network. The sensors in the network can act as interrogators and are often mobile. If they are inadvertently moved, they can compensate electronically, without human intervention, i.e., they are “self-calibrating”. The nodes can be constructed from a variety of electronic hardware - a sensor, an actuator, a microprocessor, a radio and a power source - and some may be different from others in a given system to form a “heterogeneous network”.

IEEE 802.15.4 wireless technology is a short-range communication system intended to provide applications with throughput and latency requirements in WSN. The key features of 802.15.4 wireless technology are short distance transmission, low power consumption and low cost characteristics that can be supported by devices. Most wireless sensor networks use wireless mesh technology based on IEEE 802.15.4, sometimes referred to as ZigBee. ZigBee is a specification for a suite of high level communication protocols using small, low-power digital radios based on an IEEE 802 standard for personal area networks (PAN). Using ZigBee protocol, sensors are able to communicate with each other on low-power, reliable bit-rate transfer of 250kbps at 2.4GHz band and secure data transfer, i.e., 128 AES plus security. The radio design used by ZigBee has been optimised for low-cost production and has a transmission range of less than 100 m.

WSNs are playing a key role in supporting several IOT application scenarios. With many smart objects having different communication, information and processing capabilities, a reliable network to provide seamless interaction among them becomes imperative. Scalability is another issue for the IOT due to the large scope of communications needed to seamlessly interconnect objects and people. Finally, when dealing with battery-operated smart objects, low-power communication becomes a crucial aspect to ensure continuous connectivity of these objects. The characteristics of WSN support these network requirements. Some examples of WSN deployments are in healthcare, environmental monitoring and smart buildings (Figure 11).

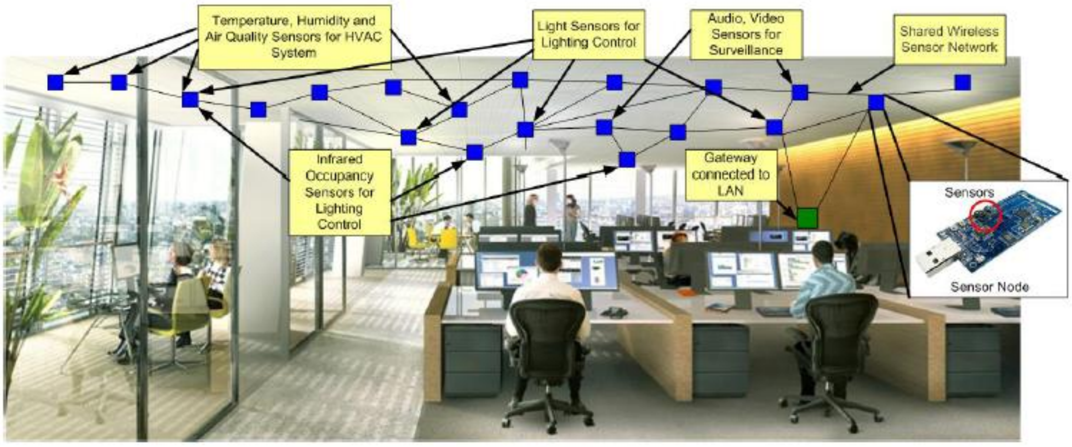


Figure 11: Wireless sensor network (WSN) applications[25]

12.5.1.2 Three to five years

Adaptive Learning Analytics

Adaptive learning analytics is a range of analytics algorithms that is performed by sensors and mobile devices to make intelligent analyses of real-time data. The algorithms tune the data stream processing parameters according to the occurring situations and available computational resources such as battery charge and available memory to make the most optimum decision or recommendations. For example, in a smart room scenario, rather than monitoring sensed context from light, noise and motion sensors individually, this information can be used to reason about situations, such as “meeting”, “presentation” or “study”, providing a better understanding of the environment. It also provides a more abstract view of the environment, rather than focusing on individual pieces of context. Such information output is also formulated, based on optimisation of the computational resources of the sensing devices, and reduces the volume of data needed to be sent to back-end systems for analytics processing.

IPv6 and 6LowPAN

---

[25] Peter Harrop, Raghu Das. Wireless Sensor Networks 2010 – 2020. [Online] Available from: http://media.idtechex.com/pdfs/en/R7036G0761.pdf [Accessed 9th July 2012].

IPv6 is 128-bit Internet address scheme that is used to replace IPv4 addresses which were officially exhausted in February 2011 (in the Asia Pacific Network Information Centre or APNIC). With IPv6, there are approximately 3.4×10^{38} unique IPv6 addresses, more than enough address space to accommodate the universe of cloud-ready devices for the foreseeable future. IPv6 allows network auto-configuration, making devices easy to be managed and configured automatically once they are in the network. IPv6 can bring about efficient allocation and management of addresses by distributing addressing space in a hierarchical manner. For example, IPv6 addresses allocated to an organisation can be further distributed according to the topology of network infrastructure. Alternatively, IPv6 addresses can be assigned according to geographical regions within a country i.e. North, South, East and West and further allocated based on street levels and building levels. By having an efficient allocation scheme for IPv6 addresses, it allows better manageability, usability and ease of devices identification.

6LoWPAN is an acronym for "IPv6 over Low power Wireless Personal Area Networks". It is a communication standard that allows the low-power devices to communicate and exchange data via IPv6. There are many benefits of using IP-based connectivity to form the sensor access network:

- IP connects easily to other IP networks without the need for translation gateways or proxies.
- IP networks allow the use of existing network infrastructure.
- IP is proven to work and scale. Socket API is well-known and widely used
- IP is open and free, with standards, process and documents available to anyone. It encourages innovation and is well understood.

6LowPAN works on the IPv6 protocol suite based on IEEE 802.15.4 standard. Hence it has the characteristics of low-cost, low-rate and low-power deployment. The bottom layer adopts the physical (PHY) and media access control (MAC) layer standards of IEEE802.15.4, and uses IPv6 as the networking technology. The reference model of the protocol stacks as shown below (Figure 12):

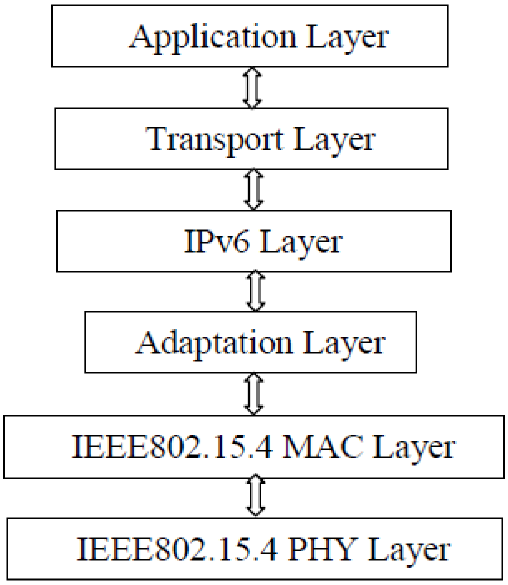


Figure 12: Reference model of 6LowPAN protocol stack

IOT applications that require a two-way, low-power communication network would benefit from the use of 6LoWPAN. For example, in smart grid systems, a power plant can transmit electric power and messages to clients. On the other hand, clients can also send information to the power plant. Through two-way communications, the power utilisation can be adjusted in a more efficient way. With its affordability and practicality, 6LowPAN presents significant opportunities for the market.

12.5.1.3 More than five years

Nanotechnology

Nanotechnology is enabling the development of devices ranging from one to a few hundred nanometers. At this scale, a nanomachine is defined as the most basic functional unit and is integrated with nano-components to perform simple tasks such as sensing or actuation. Coordination and information sharing among several nanomachines will expand the potential applications of individual devices both in terms of complexity and range of operation.

The U.S. National Nanotechnology Initiative has described four generations of nanotechnology development[26] with the first generation being nanostructures, materials designed to perform one task. The second phase sees the introduction of active nanostructures for multitasking, for example, actuators and sensors. The third generation, occurring around 2010, featured nanosystems with thousands of interacting components. From now till 2015, integrated nanosystems, functioning with hierarchical systems within systems to perform complexity tasks, will emerge (Figure 13).

---

[26] Center for Responsible Nanotechnology. What is Nanotechnology. [Online] Available from: http://www.crnano.org/whatis.htm [Accessed 9th July 2012].

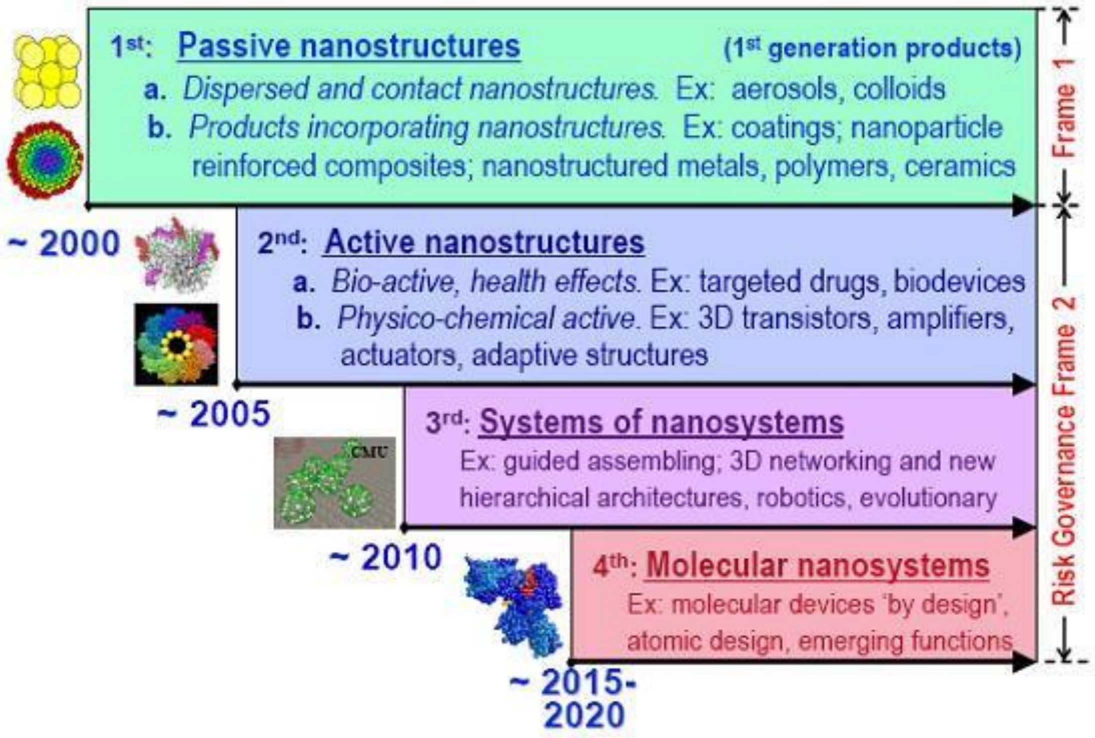


Figure 13: Nanotechnology evolution[27]

With nanotechnology, future sensors designs can get much smaller, less power hungry, and more sensitive than current sensors. Sensing applications will thus enjoy benefits far beyond those offered by existing technologies such as MEMS. In the areas of power generation and storage, nanotechnology uses nanomaterials that can dramatically reduce the size of energy storage devices and increase the energy density of the same devices.

12.5.2 Gateways and Networks

12.5.2.1 Three to five years

Software Defined Radios

Software Defined Radio (SDR) is a software radio algorithm that puts most of the radio frequency (RF) into the digital domain, allowing for great flexibility in the modes of radio operation. Traditional radios are hard-wired to communicate using one specific protocol, e.g., mobile phones will need multiple radios to handle a variety of communication modes with cell phone towers, WiFi base stations and GPS signals. SDR works with raw electromagnetic signals and enables multiple radio functionalities with the use of software. This makes it possible for SDR devices to tune into many different frequencies simultaneously and make multiple communications with other devices which use specific radio protocols.

---

[27] Center for Responsible Nanotechnology. What is Nanotechnology. [Online] Available from: http://www.crnano.org/whatis.htm [Accessed 9th July 2012].

With billions of connected devices on current communication technologies, new communication needs will require faster connectivity and the frequency spectrum will have to be adapted to the new bandwidth requirements. Using SDR, the need to implement hardware upgrades when new protocols emerge can be removed. SDR will eventually find its way into mobile devices to enable them to operate across many radio bands and using many radio protocols and modulation techniques. The technologies could become commercial in the next three to five years and will have a broad impact on IOT, especially M2M devices, so as to enable human-oriented devices to search for the best frequency, depending on a predetermined set of parameters.

Currently one of the earliest examples of a commercial SDR is the range of ZTE 3G Dual-carrier High Speed Packet Access (DC-HSPA)/LTE universal radio-base stations that is deployed at CSL Hong Kong for the operator's new Frequency Division Duplex (FDD)-LTE service, launched in Q1 2012.

12.5.2.2 Less than three years and more than five years

LTE & LTE-A

LTE is a 4G wireless broadband technology developed by the Third Generation Partnership Project (3GPP), an industry trade group. 3GPP engineers named the technology "Long Term Evolution" because it represents the next step in a progression from GSM, a 2G standard, to UMTS, the 3G technologies based upon GSM. LTE provides significantly increased peak data rates, with the potential for 100 Mbps downstream and 50 Mbps upstream, reduced latency and scalable bandwidth capacity. Future developments could yield peak throughput to the order of 300 Mbps.

LTE-A is a major enhancement of the LTE standard developed by 3GPP and has been approved by the International Telecommunication Union (ITU) as the 4th generation (4G) radio technologies system. LTE-A is backward compatible with LTE and uses the same frequency bands while LTE is not backward-compatible with 3G systems. LTE-A has the potential for even faster peak data rates, with the potential for 1 Gbps downstream and 500 Mbps upstream.

With the increasing number of IOT sensors, M2M devices and various applications such as context-aware computing services generating huge amounts of real-time data, mobile data networks could experience millions of transactions or interactions between cloud servers, peer-to-peer communications and back-end systems. Leveraging on LTE and LTE-A networks which have been designed to increase network capacity, speed of mobile communications and low latency, IOT applications would enable substantial improvements in end user throughputs, application response times and user experience.

12.5.2.3 More than five years

Cognitive Networks

The definition of a cognitive network is a network that can perceive current network conditions to plan, decide, and respond, based on those conditions. The network can learn from these adaptations to make future decisions. Today's networking technology limits a network's ability to adapt, often resulting in sub-optimal performance. It is not designed for IOT as network usage tends to be more downlink than uplink intensive. Network traffic is more IP based than hybrid e.g. Zigee and UWB and has a relative fixed access pattern compared to dynamic access pattern which includes multiple sensors, actuators, sensor gateways. Often being limited in scope and response mechanisms, the network elements are unable to make intelligent adaptations to suit networking requirements.

Cognitive networks use observations of network performance as inputs to provide outputs in the form of a set of actions that can be implemented in the modifiable elements of the networks. For example, in addressing network QoS, cognitive networks can utilise the feedback about observed network conditions to identify bottlenecks, change in resource prioritisation and optimising behaviour, to provide the desired end-to-end QoS. In another area such as network security, cognitive networks can react to security threats by analysing feedback from the various layers of the network to find patterns and risks, and to dynamically manipulate access control, trust management or intrusion detection to eliminate security threats.

Some of the technologies needed in cognitive networks are software adaptable networks (SAN) and cognitive radio. SAN consists of the API, modifiable network elements and network status sensors. It implements the actual network functionality and allows the cognitive process to adapt to the network. Cognitive radio is a transceiver that automatically changes its transmission or reception parameters. Its cognitive process has the capability to learn from past decisions and rely on observations, paired with knowledge of node capabilities, to influence future transmission behaviour.

With various network requirements by IOT applications, the cognitive network will allow IOT devices and M2M machines to form a pervasive communication environment, designed to be extensible and flexible to ensure a certain level of QoS and user experience.

12.5.3 Management Services

12.5.3.1 Less than three years

In-Memory analytics

In-memory analytics is an approach to querying data when it resides in a computer's random access memory (RAM), as opposed to querying data that is stored on physical disks. This approach vastly shortens query response times, allowing business intelligence (BI) and analytics applications to support faster decision making.

On the traditional disk-based analytics platform, metadata has to be created before the actual analytics process takes place. The way which the metadata is modelled is dependent on the analytics requirements. Changing the way to model the metadata to fulfill new requirements requires a good level of technical knowledge. In-memory analytics reduces or eliminates the need for data indexing and storing pre-aggregated data in OLAP cubes or aggregate tables. It allows the developer to consider every possible avenue of analysis and improves the relevance of the analytics content.

It is anticipated that as future applications require faster query response times, in-memory analytics can be embedded at the chipset level and traditional data warehouses may eventually be used only for data that is not queried frequently.

In-memory analytics is enabled by a series of in-memory technologies:

- In-memory data management
    I. In-memory database management system (IMDBMS): An IMDBMS stores the entire database in the computer RAM, negating the need for disk I/O instructions. This allows applications to run completely in memory;
    II. In-memory data grid (IMDG):The IMDG provides a distributed in-memory store in which multiple, distributed applications can place and retrieve large volumes of data objects.

- In-memory low-latency messaging
    I. This platform provides a mechanism for applications to exchange messages as rapidly as possible through direct memory communications.

Context-aware computing services

The definition of “context” is the set of environmental states and setting that either determines an application's behaviour or in which an application event occurs and is useful to the user. Thus, context is any information that can be used to characterise the activity or situation of an entity where the entity is a person, place or object that is considered relevant by the application for the user. Context-aware computing is a new computing paradigm that describes software/hardware that utilises contextual information to enable a system to predict and act on behalf of, or in accordance with, a user's profile and predetermined requirements.

Context-aware computing is an exciting and challenging area of human-computer interaction as it gives computers perceptual qualities such as eyes and ears to make them recognise the situations that users interact with. In IOT, the use of sensors allows situations and activities to be captured and classified as contexts. Once the system has recognised the context in which interaction takes place, this information can be used to change, trigger and adapt the behaviour of applications and systems.

Today, there are already many forms of context-aware systems such as mobile device displays that switch the orientation of the screen according to the user’s current orientation and switch on the backlight of the phone in dark surroundings. In transportation, context-aware systems enable vehicles to have anti-lock braking systems and an Electronic Stability Program (ESP) that manoeuvres the cars to safety in extreme situations. More forms of context-aware systems will appear in the next few years as the technologies become more mature and integrate into our daily lives.

Predictive analytics

Predictive analytics is a set of statistical and analytical techniques that are used to uncover relationships and patterns within large volumes of data so that they can be used to predict behaviour or events. Predictive analytics may mine information and patterns in structured and unstructured data sets as well as data streams to anticipate the future.

In IOT, millions of connected devices and sensors will be sending data on the Internet and strategic decisions can be made by determining the patterns and outliers among the huge volumes of data. For example, predictive analytics can identify the buying behaviour trends of consumers from data captured by retail store cashiers to come up with future marketing trends. Predictive analytics can also be used in traffic management to identify and prevent traffic gridlocks.

There are three methods of predictive analytics approaches[28]:

Pattern-based approach: This approach compares real-time system performance and configuration data with unstructured data sources that may include known failure profiles, historic failure records and configuration data. The objective is to extract statistical patterns within the huge multifaceted data repositories, using powerful correlation engines, to determine if the current configuration and performance data indicate a likely failure.

Rule-based approach: A series of rules is defined, based on statistical analyses of historic performance data, previously identified failure modes and the results of system testing. Each rule may be compared to multiple data streams and other external factors such as the time of day, environmental conditions and concurrent external activities against defined thresholds. Breaches of these rules may then be collated and escalation routines used to determine the likely severity and impact of the resultant issue or outage.

Statistical process control-based approach: Control charts[29] have proven to be an invaluable aid in managing complex, process-driven systems. The advent of retrofit-capable real-time telemetry and improvements in data acquisition solutions and network capacity to support large data volumes mean the statistical techniques that underpinned the quality revolution within the manufacturing space can now be used to industrialise IT services delivery. Statistical anomalies can be readily identified and used to initiate preventive action appropriately to ensure that service performance is unaffected.

12.5.3.2 Three to five years

Streaming Analytics

---

[28] Rob Addy. Emerging Technology Analysis: Predictive Support Services. [Online] Available from: http://www.gartner.com/id=1875816 [Accessed 9th July 2012].

[29] Control chart is a time series chart showing performance against upper and lower control limits (also known as tramline charts) that is generally associated with the practice of statistical quality control (SQC) or statistical process control (SPC)

Streaming analytics is a new paradigm of analysing data that does not require the storing of data. It processes the data on the fly, as soon as it arrives in a high speed stream, and then discards the data in order to make room for subsequent data. IOT data captured by sensors and devices is constantly changing but may not represent changes that are meaningful, e.g., periodic updates of temperature information. Streaming analytics has to be applied to draw meaningful changes in data and detect complex patterns over time to come up with the actions that better fit the new environment.

Some examples of real-time applications that require streaming analytics include traffic data networks, telephone conversations, ATM transactions and sensor data.

Complex event processing

Complex event processing (CEP) encompasses ways to process events while they occur and derive patterns in newly arrived event data. It is a style of computing that is implemented by event-driven, continuous intelligence systems. A CEP system uses algorithms and rules to process streams of event data that it receives from one or more sources such as ERP applications, financial applications, Web and operational analytics to generate insights. It generates new summary-level facts or complex events, and puts them in context to identify threat and opportunity situations. For example, sales managers who receive an alert message containing a complex event that says, "Today's sales volume is 30% above average" grasp the situation more quickly than if they were shown the hundreds of individual sales transactions (base events) that contributed to that complex event. This information is then used to guide the response in sense-and-respond business activities. Computation of CEP is triggered by the receipt of event data. CEP systems store large numbers of events within the memory spaces, aggregating unrelated events from multiple sources and executing highly complex analyses when the event data arrives (Figure 14).

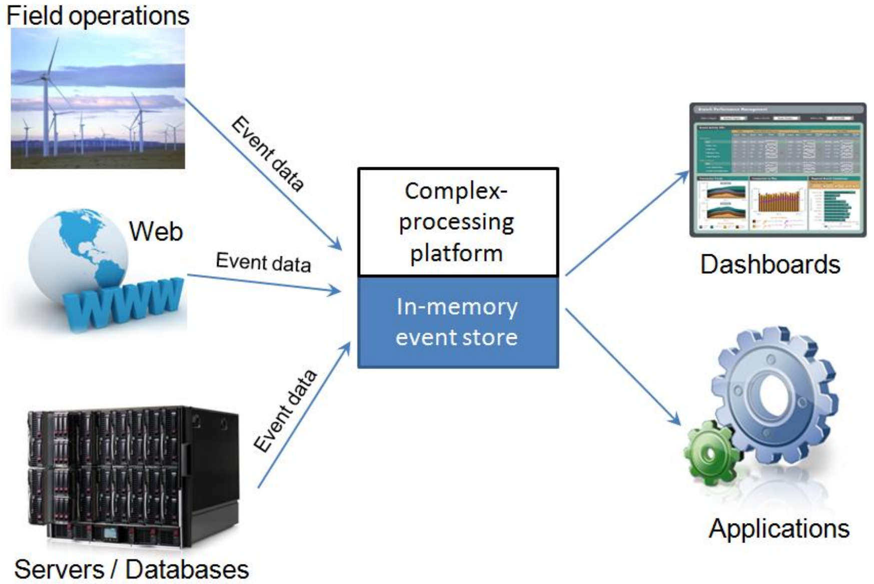


Figure 14: Complex event processing

CEP is particularly useful for IOT as numerous events are produced daily. Traditionally, events that are generated by RFID readers or sensors are considered primitive events. The information inside primitive events is quite limited. Moving ahead with real-time IOT applications, information is becoming more complicated and includes business logic and rules; to derive useful insights, the combination of primitive events into a complex event becomes important. In manufacturing process monitoring and control, CEP systems are particularly helpful in low-latency data collection, and analysis of plant-floor devices and sensors to ensure that these processes are running optimally.

12.5.3.3 More than five years

Behavioural Analytics

Behavioural analytics is the combination of strategies and tools that allows the application to identify individual consumers and their consumption preferences and behaviour. It can be used to identify a selected user or a user across multiple platforms over time and is most efficient when it is connected with different data sources, e.g., mobile networks or subscription databases.

In IOT, one example of behavioural analytics application is in the consumer-oriented video systems designed to track customer movements in the retail stores to determine buying intentions. Retailers can actively explore such systems to measure shoppers’ behaviour and suggest the appropriate recommendations to target the right customer. Other areas of application can be in automating the analysis of surveillance systems at high-risk, restricted areas. The analytics capability can be deployed to track behavioural patterns for the likelihood of criminal and terrorist activities.

12.6 Market Opportunities

IOT presents many opportunities for industry verticals and brings about new innovations to businesses. With real-time data and potentially cross-domain data sharing, new business models can be created. Opportunities such as IOT as a service, new markets and value chains can be formed to enhance competitive edge. This section, though not comprehensive in coverage, considers the business drivers for IOT and potential applications that can be developed to support various industry sectors and cross-domain businesses.

12.6.1 IOT Business Drivers

12.6.1.1 Supply Chains

One of the most well-regarded applications in the supply chain involves using sensors to track RFID tags placed on products moving through supply chains, to improve inventory management while reducing working capital and logistics costs. Back in January 2005, Wal-Mart required its top 100 suppliers to attach RFID tags on their shipping palates and extended the same compliance to more than 15,000 suppliers in January 2008. With such an implementation, Wal-Mart had its manual orders decreased by 10 to 15% which contributed to inventory reduction. Furthermore, Wal-Mart's suppliers such as Kimberly-Clark and Procter & Gamble were also able to be informed of the movement of their pallets, cases and promotional displays within 30 minutes[30] of the tags being read, enabling them to measure the impact and execution of promotions on sales. To date, Wal-Mart has moved toward item-level RFID tagging of apparel, for instance, in order to improve inventory accuracy and on-shelf availability.

With IOT, a rich variety of other information such as temperature, geographic location and transport routes, can be collected as the products traverse the supply chain. With the real-time visibility provided by IOT, customers are able to query rich information on their parcels or containers at any time, compared with the current intermittent updates on parcel information offered by the logistics service providers.

12.6.1.2 Government

IOT presents opportunities for government to improve public services and enhance homeland security. Agencies have data feeds from monitoring sources such as cameras, environmental sensors and building emissions. The sharing of this data can lead to better inter-agency coordination and the development of new public services for citizens. Some examples of new public services are shared video surveillance service for citizens to monitor their unlit streets and water levels in nearby reservoirs/canals, as well as the availability of preferred public housing units.

---

[30] Wal-Mart, Suppliers Affirm RFID Benefits. [Online] Available from: http://www.rfidjournal.com/article/view/3059/2 [Accessed 9th July 2012].

In areas of homeland security, the Memphis Police Department (MPD) in New York[31] has enhanced its crime fighting techniques by using statistical and predictive analytics software to monitor crime activities. Using predictive analytics software, MPD is able to forecast precincts that will see the most crime activities and detect crime patterns in real time to prevent criminal activities from happening. With the implementation of this software, MPD saw a reduction of more than 30% in serious crime, including a 15% reduction in violent crime.

12.6.1.3 Retail

Implementing IOT in retail allows retailers to track their inventory and sales in real time. With products RFID-tagged and smart shelves equipped with sensors such as weight and motion detectors, retailers can optimise many applications to detect the movement of the products and track out-of-stock items on shelves. The savings potential in a retail store is large. For example, sales losses that occur when shelves go empty are estimated to be 3.9% of sales worldwide[32]. Retailers are relying on sensor readings and video footages to gather and process data from thousands of shoppers as they journey through stores. This data can provide insights as to how long the shoppers linger at individual displays and the types of product they ultimately buy. Applying context-aware computing services on this data will help profile the behaviour of shoppers and guide retailers on how to optimise retail layouts to better attract shoppers.

12.6.1.4 Healthcare

The medical body area network (MBAN) is an area in healthcare that has great potential for IOT. MBAN refers to an application of wearable computing devices whereby the devices or sensors actively monitor the human body's vital signs (e.g., heartbeat, temperature and blood pressure) and communicate wirelessly with a single body central unit (BCU) using, for example, ZigBee protocol. With sensors monitoring health conditions remotely and continuously, doctors can be given early warning of conditions that would otherwise lead to unplanned hospitalisation and expensive emergency care. Other benefits of implementing MBAN are greater flexibility for the elderly to monitor their health at home, less hospital-related infections, greater comfort for families supporting independent living for family members. Remote health monitoring is shift towards preventive healthcare that allows hospitals to better plan for emergencies for their patients and change the way insurance companies charge for insurance policies and premiums.

In the USA, the Federal Communications Commission (FCC)[33] has forecast potential cost savings using MBAN:

- With MBAN technology, doctors can intervene before a patient's condition seriously deteriorates, resulting in less time spent in the intensive care unit, and cost savings

---

[31] IBM. Memphis Police Department Reduces Crime Rates with IBM Predictive Analytics Software. [Online] Available from: http://www-03.ibm.com/press/us/en/pressrelease/32169.wss [Accessed 9th July 2012].

[32] T. W. Gruen, D. S. Corsten, and S. Bharadwaj, "Retail Out of Stocks." Technical report, 2002.

[33] "FCC Proposal to Spur Innovation in Medical Body Area Networks": http://www.fcc.gov/document/chairman-proposal-spur-innovation-medical-body-area-networks [Accessed 9th July 2012].

from fewer follow-up visits. One healthcare company estimates US$1.5 million per month could be saved if emergency transfers could be reduced by early detection and treatment of a patient's health condition.

- Disposable wireless sensors can help decrease hospital-acquired infections. The industry estimates that disposable sensors could help to save an estimated US$2,000 to US$12,000 per patient and more than US$11 billion nationwide.
- Remote monitoring of patients with congestive heart failure alone could result in annual savings of over US$10 billion.

12.6.1.5 Transportation

In transportation, IOT applications can help to improve transport utilisation and road safety with the convergence of real-time road traffic data. For example, in the city of Madrid[34], a control centre has been built to monitor the underground traffic of a busy highway, M30, through which an average of 200,000 vehicles pass daily. The intention is to address traffic congestion and reduce accidents in the tunnel, using sensors to monitor the real-time traffic conditions 24/7. Data is collated from various sensors such as fibre optic cables and CCTV cameras, and the analytics of the real-time data is performed by back-end systems, taking into consideration past data trends. There are plans to publish relevant real-time tunnel situation details to the public. With such an implementation, M30 has registered a drop of 60% in accident rates and an overall increase of traffic load by 5%.

In Stockholm, Sweden[35], the city cut traffic by 20% through the use of smart tolling traffic systems; spin-off perks included 15% lowered carbon emissions and 40,000 additional daily users of public transportation.

12.6.2 IOT Applications

Below are some of the IOT applications that can be developed in the various industry sectors (these applications are not exhaustive).

12.6.2.1 Supply Chains

Dynamic Ordering Management Tool

Traditionally, the order picking management in the warehouse picks up multiple types of commodities to satisfy independent customer demands. The order picker (done manually) tries to minimise the travelling distance for time and energy savings via route optimisation and order consolidation. Using the dynamic ordering tool, the network of smart objects will identify the types of commodities and decompose the order picking process to distributed sub-tasks based on area divisions. The application will plan the delivery routes centrally

---

[34] Madrid Calle 30 Control Centre data on traffic monitoring provided to IDA

[35] "IBM helps City of Stockholm Significantly Reduce Inner City Road Traffic": http://www-03.ibm.com/press/us/en/pressrelease/24414.wss [Accessed 9th July 2012].

before activating order pickers for the delivery. Using executable algorithms in active tags, the tags can choose the best paths for the order pickers to take, as well as paths that are within their responsible areas. This results in a more optimised order processing, time savings and lower cost of delivery.

12.6.2.2 Government

Crowd Control during Emergencies and Events

The crowd control application will allow relevant government authorities to estimate the number of people gathering at event sites and determine if necessary actions need to be taken during an emergency. The application would be installed on mobile devices and users would need to agree to share their location data for the application to be effective. Using location-based technologies such as cellular, WiFi and GPS, the application will generate virtual “heat maps” of crowds. These maps can be combined with sensor information obtained from street cameras, motion sensors and officers on patrol to evaluate the impact of the crowded areas. Emergency vehicles can also be informed of the best possible routes to take, using information from real-time traffic sensor data, to avoid being stuck among the crowds.

Intelligent Lampposts

The intelligent streetlamp is a network of streetlamps that are tied together in a WAN that can be controlled and monitored from a central point, by the city or a third party. It captures data such as ambient temperature, visibility, rain, GPS location and traffic density which can be fed into applications to manage road maintenance operations, traffic management and vicinity mapping. With the availability of such real-time data, government can respond quicker to changing environments to address citizen needs.

12.6.2.3 Retail

Shopping Assistants

In the retail sector, shopping assistant applications can be used to locate appropriate items for shoppers and provide recommendations of products based on consumer preferences. Currently, most of shopping malls provide loyalty cards and bonus points for purchases made in their stores but the nature of these programmes are more passive, i.e., they do not interact with, and often do not make any recommendations for, the buyers.

The application can reside in the shopper’s personal mobile devices such as tablets and phones, and provide shopping recommendations based on the profile and current mood of the shopper. Using context-aware computing services, the application captures data feeds such as promotions, locations of products and types of stores, either from the malls’ websites or open API if the mall allows it. Next the application attempts to match the user’s shopping requirements or prompts the user for any preferences, e.g., “What would you like to buy today?” If the user wants to locate and search for a particular product in the mall, the application guides the user from the current location to the destination, using local-based technology such as WiFi embedded on the user's mobile phone.

12.6.2.4 Healthcare

Elderly Family Member Monitoring

This application creates the freedom for the elderly to move around safely outdoors, with family members being able to monitor their whereabouts. The elderly sometimes lose their way or are unable to identify familiar surroundings to recall their way back home. Family members who do not know their relatives' whereabouts may be at a loss to know where and how to start searching. The application can be a tiny piece of wearable device such as a coil-on-chip tag attached to the elderly. This tag will be equipped with location-based sensors to report the paths that the wearer has travelled. It can emit signals to inform family members if the wearer ventures away from predetermined paths. It can also detect deviations in their daily routines. Family members can also track the location of their elderly online via the user interface (UI) application.

Continuous Patient Monitoring

Continuous patient monitoring can be an extension to the "Elderly Family Member Monitoring" application; this application, however, requires the medical services companies to support it. Continuous patient monitoring requires the use of medical body sensors to monitor vital body conditions such as heartbeat, temperature and sugar levels. The application examines the current state of the patient's health for any abnormalities and can predict if the patient is going to encounter any health problems. Analytics such as predictive analytics and CEP can be used to extrapolate information to compare against existing patterns and statistics to make a judgment. Energy harvesting sensors can be used to convert physical energy to electrical energy to help power these sensors to prevent the patient from having to carry bulky batteries or to perform frequent re-charging.

Smart Pills

Smart pills are essentially ingestible sensors that are swallowed and can record various physiological measures. They can also be used to confirm that a patient has taken his or her prescribed medication, and can measure the effects of the medication.

12.6.2.5 Transportation

Special Needs and Elderly Transportation Assistant

The transportation assistant application serves to address the group of commuters with special needs and who require assistance as they commute using public transportation. When these commuters travel, e.g., using the public train service, the transport assistant will inform the nearest transport staff so they can provide special assistance such as audio and visual services, and physical assistance for the passengers. When commuters are outdoors, the transport assistant will alert oncoming public vehicles to slow down as the passengers require special assistance to board the vehicle. The transportation assistant application can embedded into watches, bracelets and panic button devices with built-in intelligent capabilities such as context-aware computing services and predictive analytics. Depending on the wearer's (user's) profile, the application recommends the most suitable assistance required by the wearer, gathering inputs from the current surroundings to make the decision. Using sensors on these wearable devices, the application communicates with other sensors or receivers, e.g., staff badges using radio frequency or Zigbee, to establish connectivity and make the request.

Accident Avoidance Detection

Vehicles can play a part in providing better road safety by monitoring and sensing each other on the roads. The accident avoidance detection application can be programmed into vehicles' on-board equipment (OBE). With the use of sensors placed within the vehicles, the application can warn the drivers of accidents or dangers that may lie ahead on the road. For example, the application is capable of interpreting a series of complex events such as poor visibility conditions resulting from heavy rain, slippery roads and strong wind to the possibility of vehicles suddenly stopping. Consequently, it can alert and advise the driver on how to drive in such conditions. Sensors using infrared (IR) can help to detect the distance between each vehicle or the conditions of the road (e.g., rain levels and fallen debris), feeding the application with the information to alert drivers to avoid and steer clear of a potential accident site.

Distributed Urban Traffic Control systems

Distributed Urban Traffic Control systems enable the tracking of car locations in real time and provide an appropriate traffic management response to handle road conditions. The control system can be used in times of emergency such as setting up of fast lane corridors for emergency services, i.e., ambulances, police cars and fire brigades, to pass through during heavy traffic conditions.

12.6.2.6 Energy Management

Facilities Energy Management

Facilities energy management involves the use of a combination of advanced metering and IT and operational technology (OT) that is capable of tracking, reporting and alerting operational staff in real time or near real time. Systems are capable of allowing highly dynamic visibility and operator influence over building and facility performance. They also provide dashboard views of energy consumption levels, with varying degrees of granulation, and allow data feeds from a wide range of building equipment and subsystems.

Home Energy Management/Consumer Energy Management

Home energy management (HEM) optimises residential energy consumption and production. Solutions include software tools that analyse energy usage, and home-area network (HAN) energy management sensors that respond to variable power prices. A combination of these solutions contributes towards reducing overall carbon emissions for homes.

12.6.2.7 Cross Domain Examples

Automotive Industry and Financial Institutions

Automotive vehicles can be embedded with location sensors to track the movement and interactions of them with other vehicles on the road. This information can be used by insurance companies to base the price of insurance policies on how well the vehicle has been driven and the places and distances it has travelled. The pricing can be customized to the actual risks of operating the vehicle rather than based on requirements such as a driver’s age, years of driving experiences, gender, etc.

Retail, Manufacturing and Supply Chain Industries

For retail owners, smart store manager applications can be used to facilitate the tracking of inventory in the store warehouse for inventory management and control. The application can be provided by an augmented reality service that uses cameras to scan RFID tags to determine the availability of inventory on each level. When the application detects that an inventory level is low or below a certain threshold, it can automatically inform Manufacturer whom will determine the production of goods. The information will also be shared with the Supply Chain to optimise the next delivery of goods to the store warehouse.

Logistics Industry

Logistics companies are tapping on traffic patterns, road congestions information from road cameras and sensors and early knowledge of weather conditions to make constant routing adjustments for their delivery trips. These cross-domain information help them increase their delivery efficiencies and reduce overall congestion costs.

12.6.3 Challenges to achieve full IOT potential

There are key challenges and implications today that need to be addressed before mass adoption of IOT can occur.

12.6.3.1 Silo IOT solutions

The commonly observed trend in today’s IOT domain is the emergence of a variety of solutions targeted at specific application domains. These application developments have limited interoperability between systems and technologies, and they do not adopt a common standardisation and understanding of the IOT domain.

An example is in greenfield applications such as home healthcare where there have been new families of products that have shown economic and social benefits. However, these domain-specific solutions are developed for a single application and with only one scenario in mind. Moving into the future of IOT, system interfaces should be standardised, and solutions made interoperable at various levels (e.g., communication and service levels) and across various platforms in order to promote integration and scalability.

Today, industry organisations are trying to form non-profit and open industry alliances to bring about interoperable solutions for industry verticals. Continua Healthcare Alliance[36] is one of such alliance that is working with standards organisations to allow information that has been collected from the healthcare devices and gateways to be made available for the electronic medical records (EMR) and other clinical applications. Continua has adopted a set of standards from the Integrating the Healthcare Enterprise (IHE) initiative and Health Level 7 (HL7) to be used for health information exchanges in various countries around the world.

12.6.3.2 Cost versus Usability

IOT uses technology to connect physical objects to the Internet. For IOT adoption to grow, the cost of components that are needed to support capabilities such as sensing, tracking and control mechanisms need to be relatively inexpensive in the coming years.

Gartner has forecast that most technology components such as radio, WiFi, sensor and GPS, could see a drop in cost of 15% to 45% from 2010 to 2015[37]. For organisations planning to adopt IOT, the reduction in costs of these components needs to be less than the increase in revenue margins that can be gained from a better product and service. The trend forecast by Gartner could incentivise organisations to pursue opportunities in IOT in the next one to three years.

12.6.3.3 Privacy and Security

As the adoption of IOT becomes pervasive, data that is captured and stored becomes huge.

One of the main concerns that the IOT has to address is privacy. The most important challenge in convincing users to adopt emerging technologies is the protection of data and privacy. Concerns over privacy and data protection are widespread, particularly as sensors and smart tags can track user movements, habits and ongoing preferences. Invisible and constant data exchange between things and people, and between things and other things, will take place, unknown to the owners and originators of such data. IOT implementations would need to decide who controls the data and for how long. The fact that in the IOT, a lot of data flows autonomously and without human knowledge makes it very important to have authorisation protocols in place to avoid the misuse of data. Moreover, protecting privacy must not be limited to technical solutions, but must encompass regulatory, market-based

---

[36] Continua Healthcare Alliance website [Online] Available from http://www.continuaalliance.org/index.html [Accessed 9th July 2012].

[37] Jim Tully, HungLeHong. Falling Cost of Components Will Help Drive The Pervasiveness of the Internet of Things. [Online] Available from: http://www.gartner.com/id=1839420 [Accessed 9th July 2012].

and socio-ethical considerations. Another area of protecting data privacy is the rising phenomenon of the “Quantified Self”[38] where people exercise access control to their own personal data e.g. food consumed, distance travelled, personal preferences. These groups of people gather data from their daily lives and grant trusted third-party applications to access their data in exchange for benefits such as free data storage and analysis. The third-party applications or providers do not have access to the raw data or usually have commercial relationships with these consumers and hence cannot use their personal data for other purposes.

In the retail/consumer example, data collected from users can range from location data, user preferences, payment information to security parameters. This data gives insight into the lives of the users and hence, appropriate privacy and security mechanisms have to be in place to protect the use and dissemination of the data.

With new IOT applications being developed from evolving data that has been processed and filtered, IOT systems must be able to resolve the privacy settings from this evolving data and also for corresponding applications.

12.6.3.4 Interoperability

Different industries today use different standards to support their applications. With numerous sources of data and heterogeneous devices, the use of standard interfaces between these diverse entities becomes important. This is especially so for applications that supports cross organizational and various system boundaries. For example in the Logistics sector, the supply chains involve multiple stakeholders like retailers, manufacturers, logistics, the IOT systems need to handle high degree of interoperability in order for information to be processed down the value chain.

12.6.3.5 Network Capacity Constraints

With convergences brought about by connected machines and smart mobile devices, there is an increasing demand for network infrastructure to support these data “hungry” [39] devices with a certain level of expected QoS. New mobile applications that perform contextual-aware services may require frequent bursts of small blocks of data for updating and synchronizing. These sessions will typically occur in “rapid-fire” bursts for a few seconds or minutes with size of the payload in few kilobytes. The rapidity of these sessions will have an impact on the latency and bandwidth of the network.

---

[38] UK, MyHealthTrainer, [Online] Available from https://connect.innovateuk.org/c/document_library/get_file?folderId=9447195&name=DLFE-102773.pdf [Accessed 9th July 2012].

[39] On 6 January 2012, Telecomsasis.net reported that iPhone 4S users downloaded 2.76 times as much data as users of the iPhone 3G. HTC Desire S users typically uploaded 3.23 times as much data as iPhone 3G users, with iPhone 4S users uploading almost as much (3.20x).

The issue of limited network capacity has prompted many global operators to develop initiatives that leverage technologies in unlicensed spectrum such as whitespace[40] and increase the use of WiFi to offload mobile data traffic for wireless usage. A pervasive high quality network infrastructure will be needed to support the rapid development and deployment of IOT applications for both domestic and international markets.

---

[40] Whitespace refers to the emerging wireless technology where devices operate in vacated license-exempt TV spectrum bands because of the upcoming analogue-to-digital switchover after 2015.