Challenges for Database Management in the Internet of Things

Article *in* IETE Technical Review · September 2009

DOI: 10.4103/0256-4602.55275 · Source: DOAJ

2 authors:

Joshua Cooper
Hildebrand Technology Limited

Anne E. James
Coventry University

ISSN : 0256-4602

The Institution of Electronics and Telecommunication Engineers

IETE Technical Review

Volume 26 • No. 5 • Sep-Oct 2009

www.ietejournals.org

Challenges for Database Management in the Internet of Things

Joshua Cooper and Anne James[1]

Hildebrand, London, UK,

[1]Distributed Systems and Modelling Applied Research Group, Department of Computing and the Digital Environment, Coventry University, UK

Abstract

This article discusses the challenges for Database Management in the Internet of Things. We provide scenarios to illustrate the new world that will be produced by the Internet of Things, where physical objects are fully integrated into the information highway. We discuss the different types of data that will be part of the Internet of Things. These include identification, positional, environmental, historical, and descriptive data. We consider the challenges brought by the need to manage vast quantities of data across heterogeneous systems. In particular, we consider the areas of querying, indexing, process modeling, transaction handling, and integration of heterogeneous systems. We refer to the earlier work that might provide solutions for these challenges. Finally we discuss a road map for the Internet of Things and respective technical priorities.

Keywords

Database management challenges, Internet of Things, Road map, Technical priorities.

1. Introduction

The Internet of Things (IoT) is a self-styled term to describe objects that are able to communicate via the Internet. Objects range from sensor inputs to actuators that control physical objects with new interactions requiring advances in machine and human interfaces. It is widely forecast that these objects will number in the trillions over the next five years of Internet development. Haller *et al.* [1] have provided the following definition.

"A world where physical objects are seamlessly integrated into the information network, and where they, the physical objects, can become active participants in business processes. Services are available to interact with these 'smart objects' over the Internet, query their state and any information associated with them, taking into account security and privacy issues."

Historically, the IoT referred mainly to Radio-frequency identification (RFID) tagged objects that used the Internet to communicate. Its origins lie in the manufacturing area, for example, the Auto-id project [2]. The Cambridge Auto-id laboratory produced a number of white articles, journals, and conference articles on the project.

RFID is not the only means of connection to the IoT. Wireless sensor networks will provide continuous streams of data on various environmental characteristics, which may be fed into the IoT. Other more sophisticated bridges to the IoT include identification of objects via sensing devices, for example, object recognition via digital imaging. Evermore capable display, personal computing devices, and smart materials are increasing the interactivity with human users and the physical environment. Biometric identification will be used for security and personalization of the IoT systems.

Taken to its extreme, any everyday object might become part of the IoT and be made intelligent: each book we read, every device in our homes, our pets, every food product, every item of clothing, and even ourselves. Of course one can imagine good and bad scenarios in this vision. It might be convenient to arrive home after a period of absence to a reception of our most comfortable home environment in terms of heating, lighting, and digital entertainment. This environment is set up and maintained economically and energy-efficiently through sensors that capture every environmental detail, making decisions based on their input, knowledge of commodity tariff factors, and our preferences. Let us consider another case. We cannot find the book we are reading so we press a button on the screen embedded into our living room wall and in an instance the location of the book is revealed. Or even more futuristically the book detects our arrival home, knows we want to read it, and therefore bleeps to indicate its presence.

A scenario presented in a related publication [3] describes an international car journey. The car's RFID sensor system detects possible tire failure triggering the driver to call into a garage where a diagnostic tool using sensors and radio technology conducts a comprehensive check of the car. The check results in the rear tires being replaced with new ones, incorporating RFID tags and sensors for monitoring pressure, temperature, and deformation. The driver is able to define the required privacy-related options regarding the new RFID tags. The scenario describes a beverage machine recognizing a customer and preparing her favorite drink, which is paid for by a wave of an internet watch. The information stored in the car's control system is intended for maintenance purposes, but can be read at different points of the car journey where RFID readers are available, unless the data owner's privacy option is switched on. There is no need to stop at border control as the car contains the driver's license and passport, which can be electronically read as the car passes through. During the journey the driver receives a video call on her internet-enabled sun glasses.

The IoT offers a wealth of possibilities, but perhaps there is a downside. Our possessions, movements, and activities might be detected by rogue receivers operated by people with intentions to put such information to illegal use. Even overzealous legitimate data collection can seem an invasion of privacy. Security systems are not fail-proof. Would we wish surveillance agents to know everything about us, for instance our reading habits, the music we listen to, the food we eat, with whom we associate, and where we are? Such matters need to be carefully considered.

There are predictions that the IoT makes, which will only happen if there is an economic imperative and investments are made by businesses [4]. Also given that entry costs for manufacturing objects with IoT technology will be high, certain businesses may dominate. There is therefore a risk that proprietary protocols and narrow application interfaces will become de-facto standards, taking away from the grassroots, engagements the Internet has seen. Other crucial factors related to standards are the energy and environmental requirements of the IoT. Further technological advances are required to produce energy-efficient devices and electronics that can be manufactured cleanly, incorporated into the existing production processes, and recycled.

There are many promising application areas, some of which are already appreciating the benefits of a limited IoT, provided through Internet-enabled RFID tags or wireless sensor networks. Application areas include: manufacturing, supply chain management, energy, health, automotive, and insurance. However, a number of technical challenges remain, which need to be overcome, before the full IoT vision becomes a reality. High among these challenges are internet scalability, identification, and addressability, heterogeneity, and service paradigms, as well as technologies for security, privacy, and trust mechanisms. Classic processing also becomes an issue, for example, matters of concurrency, transaction processing, and state and interprocess communication. A key feature of the IoT is the vast quantity of data that it generates or attracts. Appropriate data management is therefore a pivotal factor to perceive whether the vision of the IoT becomes a reality. Thirty years of research in traditional and more latterly new database areas may provide some pointers to this area.

In this article we describe the challenges we see facing the reality of the IoT, with particular emphasis on data management. In section 2 we describe the type of data in the IoT. In section 3 we discuss database issues in the IoT. In section 4 we set out some technical priorities and a roadmap for progress to the IoT. Section 5 offers some brief conclusions.

2. Types of Data in the Internet of Things

It is useful to classify data of the IoT into a number of categories. Some data is discrete and some continuous, some automatically generated and some an input by humans. We have categorized the data into the following areas: RFID, address/unique identifiers, descriptive data, positional and environmental data, sensor data, historical data, physics models, and command data.

2.1 Radio Frequency Identification

Radio Frequency Identification refers to identification and tracking using radio waves and is becoming a common place technology. RFID tags can be inserted into objects and used to transmit and receive information. The origins of RFID can be traced back to 1948 [5], emerging from technology advances of World War II. Some commercial applications appeared in the 1980s and standards started to emerge in the 1990s, with wider deployment, until now, where it is a part of everyday life.

The RFID tag is comprised of an integrated circuit that can store information and an antenna to receive and transmit signals. The tags can be miniaturized to a few millimeters in length and width, thus enabling their ubiquitous use for everyday objects. The tag is activated by radio waves emitted from an RFID reader. The reader communicates wirelessly with the tag. Once activated, the tag sends data stored in its memory relating to the item back to the reader. Some RFID tags are active and some are passive. Active RFID tags contain a battery and can operate autonomously. Passive RFID tags are only activated when they receive a radio wave sent by an RFID reader.

The technology is now used in many areas, for example passports, livestock tracking, road tolling, supply chain management, logistics, stock control, and healthcare. It has also been adopted in larger libraries and some retail outfits. It is a cheap technology with chips available for a few cents, but conversion from bar coding and electronic tagging can be expensive and thus many establishments have resisted its adoption, preferring to remain with barcodes. There are also some concerns about privacy and security [6,7]. One security concern is that RFID tags can be read from a distance without the carrier or owner of the tagged item being aware.

2.2 Addresses/Unique Identifiers

Objects of the IoT will need to be uniquely identified with IP addresses. As the number of objects in the IoT grows, the number of required identifiers will grow. It is with the application of the IoT in mind, among other concerns, that Internet Protocol version 6 (IPv6) was specified to replace the current widely deployed Internet Protocol version 4 (IPv4). IPv4 is suffering from IP address exhaustion, which is the decreasing supply of unallocated IPv4 addresses. This depletion has been a concern since the 1980s, when the Internet started to experience dramatic growth. As a result efforts have been made to expand the number of addresses IPv4 can hold, as also to develop a new version of the Internet Protocol, namely IPv6. IPv6 was defined in December 1998, by the Internet Engineering Task Force (IETF) [8]. IPv6, which uses a 128-bit address, has a much larger address space than IPv4, which uses only 32 bits. IPv6 has, however, not been widely deployed yet, with IPv4 remaining the most widespread protocol.

It is likely that identification will be enhanced through a hierarchical naming structure, with local identifiers qualified by domain names in a global naming hierarchy, in a similar way that domain names are used in the current Internet. The Internet has already developed a well-established naming practice, with the Internet Assigned Numbers Authority (IANA), which oversees global IP address allocation, root zone management, domain names, and other Internet Protocol-related assignments. There are subordinate administrations such as Regional Internet Registries, to deal with the current major Internet naming systems, such as IP addresses and domain names. IANA has its roots in the Advanced Research Projects Agency Network (ARPANET) [9].

At a lower level, other forms of unique identifiers may be used in the IoT, for example, Universally Unique Identifiers (UUIDs). Applications based on Bluetooth and other technologies can take advantage of this simplified form of addressing, for local applications. In 2005 [10], the IETF produced a *Request for Comments* (RFC) on a specification for a Uniform Resource Name namespace for UUIDs, which is also known as GUIDs (Globally Unique Identifiers).

2.3 Descriptive Data about Objects, Processes, and Systems

Much of the power of the IoT will come from the data or metadata that will be recorded on the participating objects, processes, and systems. Metadata is data about data and is essential to enable users to find and access the appropriate data. We can distinguish between data and metadata by the following example. Object "34.672.673.982" might have data "Cat", "Black," and "Fluffy" and the corresponding metadata "Type", "Color," and "Name". Other types of metadata may be recorded in a planned or ad-hoc manner, such as, provenance information on why data was captured, who captured it and when [11]. For instance, one might pick up an object and upon request be able to know what it is, the materials it is made of, and its origin. It can be seen that there are various types of data and metadata about objects that can be usefully recorded. In fact, in one system a piece of data might be considered to be data and in another system the same piece of data might be considered to be metadata. The distinction is not always clear and does not always need to be. Input of such data, initially by humans, is likely to be a very labor-intensive task. There are research questions on how the data will be stored, represented, and validated, to ensure maximum efficiency and non-repudiation in its retrieval and update. The objects will need to be self-describing and have the ability to report on dynamic characteristics, to maximize sharing.

Not only data about basic objects, but data about processes and systems will also need to be stored. Systems and processes can be regarded as special types of objects, albeit objects of a more complex nature than basic objects. It is important to store data about objects, processes, and systems, so that users know how to take advantage of the services and facilities offered on the IoT. Services of the IoT will allow one to see in which processes and systems an object participates so that one might utilize such an instance of the process. For example, in a domestic setting, environmental data on electricity usage might have been collected over a period of time. A process may be available on the IoT that calculates average usage with peaks and troughs over a given time period. Such a process or service may be among thousands offered in the IoT, and therefore, difficult to locate. Metadata about processes and self-describing data, together with good indexing systems, will be helpful here. An interesting research area will be the development of suitable representation schemes to capture different types of objects and their metadata, to maximize usage and usability.

2.4 Positional Data and Pervasive Environmental Data

Positional data provides the location of a particular tagged object either within a global positioning system (GPS) or within a local positioning system. GPS is implemented [12] with multiple satellites sending signals to a controlling unit from which objects can ascertain their position through triangulation. Local positioning systems work in a similar way, with smaller coverage. Examples of local technology are cellular base stations, Wi-Fi access points, and TV towers. Local positioning systems can be used in collaboration with GPS or sometimes instead of GPS. They can be used inside buildings or heavily built-up areas. In smaller areas, such as a room in a house, positional data can come from locally placed sensors and transmitters. Multiple sensors send signals to a smart object, which can then work out its location or the location of a collaborating object. Positional data will play an important part in the IoT given that its components may be static or mobile [13]. A very fine grained positioning remains a challenge and is being actively addressed by researchers [14,15].

A new type of information that will be part of the new Internet will be pervasive location information. This is information about each environment that is unobtrusively available, which will enhance and support our interactions with our surroundings. The information will be location dependent. Not only is there the phenomenon of IoT, there is also the new concept of the Internet of Places, from where information specific to places can be readily picked up by devices and users in specific locations. Eventually all locations will be incorporated. Relevant work in this area is work in mobile computing, geographical information systems, and ambient technologies.

2.5 Sensor Data — Multidimensional Time Series Data

One of the routes by which data enters the IoT is through wireless sensor networks. Advances in electronics have made it relatively easy to set up wireless sensor networks for monitoring all sorts of environmental phenomena, for example, weather, temperature, and noise. The Zigbee Alliance [16] has produced standards to support the setting up of Wireless Sensor Networks. Various interesting aspects emerge from this technology. Decisions need to be made on how frequently the data should be captured, for instance, continuously, at regular intervals, or only when queried. Questions arise on how we ensure we obtain a representative sample in an efficient manner and how much of the data capture we should archive. Sensors and grid technology have made it possible to capture vast amounts of data very quickly, but querying and mining these can be problematic, particularly when the analysis must be achieved in real-time. Techniques have been proposed to deal with these issues [17,18].

2.6 Historical Data

Petabytes and more of the data will be captured by sensors in the IoT. Such data may be required to be stored. As time passes the data becomes historical. Volume becomes a challenge. Application-oriented design decisions need to be made on how the data should be retained and which data should be retained. Some will be kept in active data warehouses for frequent querying, and some may be needed less frequently and may be archived in less accessible structures. In either case, the time needs to be captured alongside the data, along with location information, ownership, and method of capture. Issues in data archiving, such as, loss of data, inaccurate recording, missing information, and dependence on obsolete technology have been recognized for quite some time [19]. The database community has offered some solutions toward successful data archiving [20]. Such solutions may be adapted to the IoT.

2.7 Physics Models — Models that are Templates for Reality

Physics models will need to be represented so that they can be accessed and used in algorithms as needed by the applications of the IoT. Physics Models are templates for reality, for instance gravity, force, light, sound, and magnetism. Representing such models will allow modeling and simulation of physical scenarios. Physics models are in widespread use in the games- and computer-aided engineering arenas [21]. Their incorporation into the IoT will enhance its functionality.

2.8 State of Actuators and Command Data for Control

The IoT will be used to control remote devices. Thus there is a need to have a feedback showing the current state of the actuators of the devices. The representation of the state of actuators will be an interesting challenge, particularly because of the real-time nature of some of the relevant applications. Closely related technological advances in miniaturization and energy-efficient electronics, including reduced-power microcomputers and communication methods, energy-harvesting transducers, and improved micro-batteries, are needed to support the increased usage of Internet enabled actuators [4].

Some of the data entering the IoT will be the command data, to control devices. For instance one might be unexpectedly arriving home within 30 minutes and wish to switch on the heating in readiness for arrival. One will do this through the Internet using an appropriate interface. Users will need to control devices in the IoT and will need a language for doing this. It is probable that different systems participating in the IoT will have been developed independently and therefore have varying origins. It is therefore unlikely that the command interface will be the same across such systems. Work will need to be done in standardizing the command/control data and interfaces.

3. Database Issues in the Internet of Things

In this section we set out the areas of challenge for database management in the IoT and point out where developments in traditional database and other areas may provide solutions.

3.1 Size, Scale and Indexing

The size and scale of the data in the IoT will be vast. Data will need to be managed via responsible local ownership. Local owners will decide which data and services to make available to the global network. Thus, the IoT may operate on more than one level: private and public. Users may join groups for access to certain privately owned data or may, on the other hand, access data publicly available over the public Internet. There may be differences in quality of data depending on ownership and level of care. Gradually trust and reputation systems will provide information to users on the quality of the data.

In the global space there will be a need for a central authority, for managing addresses and identifiers, as there is with the current Internet. Indexing will be a major challenge. Finding a particular item in a world where all physical objects have an IP address will not be easy, unless we can devise suitable indexing methods. Work in the library catalog management might provide some pointers on how to do this, but the IoT will encompass many different types of objects. Creating a catalog of everything in the world, readable across countries and languages, is a daunting task. Some objects will be publicly accessible, some will need various levels of access control, and some may be private to the owner. At first the IoT is likely to develop through local systems that can be indexed coherently within a bounded domain. As local systems merge with global systems, new indexing methods will need to be developed. Categories of things will need to be defined together with subcategories. Specialized search sites may provide access to certain categories. For instance, if one wants to find a particular car part, one may go to a search site that specializes in that type of product, and from there be guided to a specific IP address.

3.2 Query Languages

Current popular query languages in database systems rely on structured data. *Structured Query Language* (SQL) is the most prominent example. Over the last few years, however, there have been proposals for query languages for semi-structured data, which is more typical of the data held on the Internet [22-28]. The quantities of data are so vast that it would be unrealistic to expect any sort of uniform structure, except perhaps that of the loosest variety, to be imposed on the IoT.

Extensible Markup Language (XML) offers a means of representing less structured as well as structured data, together with some level of self description. It is a well-accepted technology that supports interoperability at a technical rather than a semantic level. XQuery has been developed by the World Wide Web Consortium (W3C [29]. XQuery, a language for querying XML, can combine documents, web pages, and links to relational databases [30]. Query languages for semi-structured data usually adopt an underlying hierarchical data model, for instance a unidirectional graph. Required objects are specified by providing a path expression in a language that is usually quite intuitive. There are, however, inherent problems with hierarchical data models, such as, difficulty in representing many-to-many relationships. In spite of this, the hierarchical data model has been embraced by the web community as a useful, intuitive, and practical structure.

The IoT will have various sorts of users: casual users that briefly visit a site to pick up some data or information, expert users that know exactly what data they need and where to find it, and users that lie somewhere in between. In fact in different contexts the same person can be any one of these different types of users. It therefore seems necessary that different types of data access facilities be available. Casual users will need to access the IoT via a user-friendly graphical user interface (GUI), with detailed explanation available on any object, and more flexible, powerful, and efficient access interfaces will be needed for expert users. Services can be used to provide both types of access. Work in query languages for semi-structured data will be relevant for these developments.

3.3 Process Modeling and Transactions

It is likely that most processes will be developed and supplied as services on the IoT. Service Oriented Architecture (SOA) is becoming an important means of supporting interoperability in web-based systems [31]. The central idea is that independent outfits offer services in a uniform manner, which other users can then take up. Thus implementation details are hidden from the users of the services. Application processes will typically be made up of a number of lower level transactions. Transactions in turn will be made up of lower level operations or services. Therefore, the question of transaction processing in the IoT arises.

In the traditional database systems the matter of concurrent transaction processing has been handled through the maintenance of ACID properties through timestamping, locking, and a two-phase commit. ACID properties are atomicity, consistency, isolation, and durability. A transaction must complete in its entirety or not at all, a transaction must leave the database in a consistent state, transactions should not show other transactions, and intermediate results and changes made by a transaction must be permanent. In distributed database systems a two-phase commit is used to preserve consistency. All participating sites must confirm their readiness to commit before the commit command is issued by the coordinating site and written to the database log.

It has been recognized that the ACID properties do not fit web transaction processing well [32,33]. This is because the individual web services are essentially autonomous and must independently preserve consistency. This requirement might conflict with a consistency requirement of a user's global transaction. For instance a user sees booking a holiday as a transaction consisting of two operations: booking a flight from one operator and booking a hotel room from another operator. As far as the user is concerned the transaction should complete in its entirety or fail. However, the operators of the two booking systems may be independent and one booking may be successful while the other is not. As the underlying systems in this example are independent it is not feasible to impose a two-phase commit procedure. New methods and models have been produced for web-based transaction handling. These include the use of compensating transactions [34] and transaction systems that relax various ACID properties [35-39]. It has been found that the maintenance of ACID properties is not required by all applications. Sometimes, increased throughput of transactions may be a greater priority than preserving ACID properties.

The IoT, by its dynamic nature and vastness, makes ACID properties and a two-phase commit difficult. Within a bounded scope these traditional methods can apply. Even in the case of a domain of loosely coupled sites, it would be possible for the participating sites to agree to a two-phase commit protocol. A service could be executed to coordinate the protocol for the purpose of executing a particular transaction. Alternatively a method developed for web-based transaction handling could be applied. Relaxation of the ACID properties may be appropriate. In general the nature of the application will determine the level of transaction support needed.

3.4 Heterogeneity and Integration

Section 2 has outlined the many different types of data that will need to be handled in the IoT. The IoT will furthermore consist of billions of independent nodes, which will have their own systems for holding the data. Interoperability will not be achieved without a standard approach at some level of abstraction.

In the context of databases the areas of heterogeneity and integration have been researched since the 1980s, once it was considered useful to achieve interoperability across heterogeneous systems [40-44]. Considering that one might have a personnel system stored at one company in a relational database system, and in another company a similar system might be held in a network database system or even a different relational database system, questions arise as to how to integrate such data. Various solutions have been offered [45-49]. Some promising solutions suggest the use of a canonical data model, for instance a functional or binary data model [50,51]. However, it seems that often the solutions offered do not warrant the efforts needed to achieve them.

Now with the abundance of data and different systems on the web the problems of heterogeneity and interoperability arise anew. XML has played its part in offering a solution to some degree. It has offered a technical, practical, and efficient means to pass data from one system to another. However, XML does not solve the semantic problem. For instance, does the data item "student" in one system mean exactly the same as data item "student" in a different system? This question cannot be answered accurately without domain knowledge. Efforts to capture domain knowledge revolve around the concepts of ontology and the semantic web [52-54]. Ontologies define concepts and the relationships that exist between them. Current work has roots in the *Artificial Intelligence* (AI) and knowledge representation work of previous decades [55,56]. OWL (Web Ontology Language) is a family of languages for representing ontology on the web (W3C 2004) [57]. The idea is that communities will agree on common technologies and represent these using an OWL system, which will in turn provide the necessary support for semantic interoperability. In the future, agents will play a role in using semantic information to support the improved use of the IoT [58,59].

Rellermeyer *et al.* [60] consider the lack of a scalable model to develop and deploy applications atop a heterogeneous collection of ubiquitous devices as one of the biggest challenges in making the IoT a reality. They propose a model based on an extension of the ideas already in use for modular software development. It is likely that the SOA will play an important role in providing a fabric into which the heterogeneous applications of the IoT can be weaved.

3.5 Time Series Aggregation

Time series aggregation is an interesting area, which has been noted as raising challenges in various application domains [61,62]. It has been recognized that inappropriate time aggregations can give rise to spurious causality [63]. The problem revolves around the ability to select the optimal sampling period for continuous data. Trade-offs include processing time and storage space against accuracy and realistic representation [64].

In the database field, interesting work has been carried out in the stream data capture. It has been recognized that new models were needed for data streams [65,66]. A number of articles on the topics of database systems, data streams, stream mining, classification, and summarization have been produced [67]. It has been recognized that traditional query languages such as SQL are not suitable for querying time series data [68]. Other work has considered how missing data, which might occur through exceptions such as power breaks, can be estimated [69]. These developments will be important for intelligent data streaming capture systems in the IoT.

In the IoT, the optimal time sampling period will depend very much on the nature of the data and the application area. Suitable querying facilities will need to be defined. These are questions that will need to be addressed by data owners who will offer sampling services on their continuous data. Work that has been ongoing in data streaming will contribute to the streaming services of the IoT.

An interesting new idea is data-centric middleware for context-aware pervasive computing, where contextual data drives both application behavior and service adaptation inside the middleware system where sensors are treated as data stream publishers [70]. Adaptable schemas could become part of the solution.

3.6 Archiving

In recent times there has been much interest in archiving the Internet. A non-profit organization, Internet Archive, was founded in 1996, to build an Internet library, which included an archive of all web pages. The aim of the organization was to offer "permanent access for researchers, historians, and scholars to historical collections that exist in digital format and to stop digital publications disappearing" [71]. The web archive can be accessed via the 'Wayback' machine. As the Internet is so vast, archiving is often done automatically with crawlers, which take copies of web pages at defined intervals. Internet Archive is supported by Alexa Information, the National Science Foundation, the Library of Congress, and other institutions.

Database archiving is a long-established technique that involves taking copies of entire databases at specified intervals and keeping them in a secure store. In a web context archived databases can be converted to XML and basic querying can be permitted. Processes can also be archived. This is achieved by using the capture software that takes copies of each request and response entering a service, be that a web service, a database, or any other software system. Such techniques will be applicable to the IoT.

Interesting areas arise when we consider the nature of some of the data and the magnitude of the IoT. The solution is likely to be local management of archived data with good indexing and discovery facilities. In the archived data, the main operation will be retrieval, update will only be necessary in exceptional cases. This simplifies the problem space. Interesting areas will revolve around efficient storage, querying, and performance. Work in data warehousing and data mining may offer some directions [72-76].

3.7 Data Protection

The IoT is likely to hold much more personal information than is held on the present Internet. Furthermore, access to such information technically is borderless from the point of view of national boundaries. Many countries such as the UK have Data Protection laws. The UK Data Protection Act [77] requires all organizations which handle personal information to comply with a number of important principles regarding privacy and disclosure. Unless exempted, all UK data controllers of personal information have to register with the Information Commissioner's Office [78]. Countries in the EC operate along principles similar to the UK law, the main principle of which is that data much be processed fairly and legitimately. Another is that suitable safeguards must in place to protect the data. There are eight principles altogether. Personal data in the IoT will be harder to protect because of its sheer quantity and the interconnectedness of systems. Outside the EC, other countries also have similar laws. For the IoT to be a success, however, protection mechanisms and a legal framework that can work across national borders is essential. Some work has been carried out in developing data privacy taxonomy [79].

4. Roadmap to Progress and Technical Priorities

It has been said that the development of the IoT will depend on the existence of relevant business needs. In other words the business needs will drive the development. Already Internet-enabled RFID is being used in inventory control, logistics, and healthcare. As enabled objects become more pervasive, application support will begin to be built into common devices and appliances.

The National Intelligence Center (NIC) of the US government, informed by SRI Business Consulting intelligence [4], sees the technological road map of the IoT as follows. In the 2000s demand for expedited logistics has led to RFID tags for facilitating routing, inventory, and loss prevention. Cost reduction will lead to a new wave of applications in the 2010s. In the early 2010s these will include surveillance, security, healthcare, transport, food safety, and document management. In the late 2010s the ability of devices located indoors to receive geolocation signals will lead to further applications, including locating people and everyday objects. Around 2020, miniaturization, power efficient electronics, and available spectrum will lead to applications of teleoperation and telepresence, namely, the ability to monitor and control distant objects. Software agents and advanced sensor fusion will further enhance the power of the IoT.

Likewise, advances in general database research will have direct application to the IoT. In the recent Claremont Report of Database Research [80], there has been a consensus among leading database researchers that there is a resurgence in fundamental database research moving away from the traditional relational models. Important core topics that are useful in the IoT have been identified:

- Exploiting remote RAM and Flash as persistent media.
- Compressing and encrypting data at the storage layer.
- Designing systems that embrace non-relational models.
- Trading off consistency and availability for better performance.
- Designing power-aware DBMSs that limit energy costs.

In order to realize this vision, appropriate systems must be in place, for the handling of data. In recent years the database community has addressed new areas and suggested new methods and mechanisms for working with data on the web. Distributed system methods have been proposed to overcome some of the limitations of the central data base architectures; these include new models, new query languages, new methods of transaction handling, and new methods of analyzing continuous data, known as data streams. Many of these results will contribute to the future data management systems of the IoT. However, some of the methods will require adaption to the specific challenges of the IoT.

Some technical priorities for the development of the IoT are:

- Process modeling and Interoperability — A fabric for interoperability is needed. At a technical level, SOA looks to be a promising direction for interoperability. The characteristics of the IoT need to be thoroughly evaluated, to see if the general SAO approach works and if so, whether and how it would need to be adapted.
- Methods of indexing — The magnitude of data in the IoT will be something that we have not experienced before. Will our current method of indexing work? At a high level, we need to establish a taxonomy of things from which general indices can be built, At a low level we need to assess standard indexing and organization approaches, to see how and if they can be adapted to the IoT.
- Archiving — Standard methods of archiving need to be established. Different approaches may be suitable for different applications. Taxonomy of approaches mapped against applications would be a useful aid for the IoT.
- Actuator Control — Much of the IoT will be actuator controlled. A standard for control and command data that links with things and actuators would be helpful.
- Transaction Management — There has been much discussion about transaction management and the state of the IoT. Some applications will need strong transaction management, others may require none. A study needs to be made of application types and recommendations for relevant transaction management. Then systems or services can be built that offer various types and levels of transaction management.
- Intelligent Interoperability — Interoperability can be achieved at a technical level through approaches like SOA and XML. At a semantic level more is needed. AI techniques, such as knowledge representation, expressed through ontologies and other means, might be used to make processes and things more intelligent. A study needs to be made of applications and suggestions made regarding the knowledge representation systems that need to be developed. Organic systems should be developed, which provide ontologies for specified domains.
- Agents — To remove the need for human interaction for various tasks, intelligent agents should be developed. These will work with the ontologies established for specified domains.

5. Conclusion

In this article we have discussed the IoT in the context of database challenges. We have considered the type of data that will be part of the IoT and have discussed areas of concern for data management. We conclude that some interesting database research results of recent years could prove useful in developing the technologies needed for data management in the IoT, for instance methods for querying semi-structured data, data streaming, sampling continuous data, and data mining. Such approaches would need to be adapted to the particular requirements of the IoT. The major priority is finding suitable methods for storing, indexing, accessing, and enabling self-description of trillions of objects that will be part of the new IoT.

References

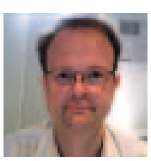
1. S. Haller, S. Karnouskos, and C. Schroth, “The Internet of Things in an enterprise context”, in Future Internet Systems (FIS), LCNS, vol. 5468. Springer, 2008, pp. 14-8.
2. C.Y. Wong, “Integration of Auto-id tagging system with holonic manufacturing systems”, White Article, Auto-id Labs, University of Cambridge, 2001 (available at http://www.autoidlabs.org/single-view/dir/article/6/198/page.html).
3. ITU, The Internet of Things, ITU Internet Reports 2005, Executive Summary, 2005 (available at http://www.itu.int/osg/spu/publications/internetofthings/InternetofThings_summary.pdf).
4. SRI consulting business intelligence, technology roadmap: The internet of things, NCI, Directorate of National Intelligence, US (available at http://www.dni.gov/nic/PDF_GIF_confreports/disruptivetech/appendix_F.pdf).
5. J. Landt, Shrouds of time: The history of RFID, AIM, Inc. (available at http://www.transcore.com/pdf/AIM%20shrouds_of_time.pdf).
6. K. Albrecht, and L. McIntyre, Spychips: How Major Corporations and Government Plan to Track Your Every Purchase and Watch Your Every Move, Plume, 2006.
7. R. Van Krannenburg, The internet of things: A critique of ambient technology and the all-seeing Network of RFID, Institute of Network Cultures, 2008 (available at http://www.networkcultures.

org/_uploads/notebook2_theinternetofthings.pdf).

8. Internet Protocol, Version 6 (IPv6) Specification (available at http://tools.ietf.org/html/rfc2460).
9. P.T. Kirstein, "Early experiences with the Arpanet and Internet in the United Kingdom", Annals of the History of Computing, IEEE, vol. 21, 1, 1999, pp. 38-44.
10. RFC 4122, A Universally Unique IDentifier (UUID) URN Namespace, http://tools.ietf.org/html/rfc4122.
11. P. Buneman, S. Khanna, and W.C. Tan, "Why and where: a characterisation of data provenance", in ICDT, LNCS, vol. 1973, Springer, 2001, pp. 316-30.
12. J.M. Zogg, GPS Basics, u-blox ag, 2002, (available at http://geology.isu.edu/geostac/Field_Exercise/GPS/GPS_basics_u_blox_en.pdf).
13. P. Steggles, P. Batty, and T. Sails, Local Positioning Systems: New Possibilities for Urban Combat Training, Ubisense White Article, 2003, (available at http://www.ubisense.net/media/pdf/Local_positioning_systems-_New_possibilities_for_urban_combat_training.pdf).
14. X. Ji, and H. Zha, "Sensor positioning in wireless ad-hoc sensor networks with multidimensional scaling", in INFOCOM, 2004 (available at http://www.ieee-infocom.org/2004/Articles/55_2.pdf).
15. A. Savvides, C.C. Han, and M. Srivastava, "Dynamic fine-grained localization in ad-hoc networks of sensors", in 7th ACM Int. Conf. on Mobile Computing and Networking (Mobicom), Rome, Italy, Jul. 2001, pp. 166-79.
16. Zigbee Alliance, Website 2009, http://www.zigbee.org/.
17. X. Zhang, T. Kurc, J. Saltz, and S. Parthsarathy, "Design and analysis of a multi-dimensional data sampling service for large scale data analysis applications", in 20th IEEE International Parallel and Distributed Processing Symposium (IPDPS), Apr., 2006, pp. 58.
18. A. Deligiannakis, and Y. Kotidis, "Exploiting spatio-temporal correlations for data processing in sensor networks", in Geosensor Networks, LNCS, vol. 4540, Springer, 2008, pp. 45-65.
19. C. Huc, T. Levoir, and M. Nononlatapie, Long-Term Archiving and Data Access: Modelling and Standards, Modelling and Standards, 1996.
20. P. Buneman, S. Khanna, K. Tajima, and W.C. Tan, "Archiving scientific data", in ACM SIGMOD International Conference on Management of Data, Jun. 2002.
21. J.R. Ellis, and D. Conger, Physics Modeling for Game Programmers, Premier Press, 2004.
22. S. Abiteboul, "Querying semi-structured data", in 6th International Conference on Database Theory, ICDT'97, Delphi, Greece, LNCS, vol. 1186, Springer,1997, pp. 1-18.
23. D. Chamberlin, J. Robie, and D. Florescu, "Quilt: an XML query language for heterogeneous data sources", in WebDB (Informal Proceedings), 2000, pp. 53-62 (available at http://www.research.att.com/conf/webdb2000/ARTICLES/4.pdf).
24. A. Heuer, and D. Priebe, "Integrating a query language for structured and semi-structured data and IR techniques", in 11th International Workshop Database and Expert Systems Applications, IEEE, 2000.
25. L. Cardelli, and G. Ghelli, "A query language based on the ambient logic", in ESOP'01 (invited article), (available at http://www.luca.demon.co.uk).
26. W. Ni, and T.W. Ling, "Glass: a graphical query Language for semi-structured data", in DASFAA, 2003, pp. 363.
27. L. Liu, C. Du, X. Shi, H. Song, and Z. Lin, "RDF-Based representation and query of information in digital library", in First IEEE International Symposium on Information Technologies and Applications in Education, ISITAE '07, 2007.
28. Z. Cao, Z. Wu, and Y. Wang, "UMQL: A unified multimedia query language", in Third International IEEE Conference on Signal-Image Technologies and Internet-Based System, SITIS '07, 2007.
29. W3C, XML Query (XQuery) 1.1 Requirements, W3C Working Draft 23 Mar. 2007 (available at http://www.w3.org/TR/xquery-11-requirements).
30. J. Melton, and S. Buxton, Querying XML: XQuery, XPath and SQL/XML in Context, Data Management Systems Series, Morgan Kaufmann, 2006.
31. E. Newcomer, and G. Lomow, Understanding SOA with Web Services, Addison Wesley, 2005.
32. M. Younas, and K. Chao, "A tentative commit protocol for composite web services", Joural of Computer System Science, vol. 72, 2006, pp. 1226-37.
33. K. Tshoganetso, A. James, and R. Iqbal, "Adaptive user defined transaction relaxing approach for CSCW", in 13th International Conference on Computer Supported Cooperative Work in Design, Santiago Chile, 2009.
34. M. Chessell, C. Griffin, D. Vines, M. Butley, C. Ferreira, and P. Henderson, "Extending the concept of transaction compensation", IBM Systems Journal, vol. 41, 4, 2002.
35. S. Choi, H. Jang, J. Kim, S.M. Kim, J. Song, and Y. Lee, "Maintaining consistency under isolation relaxation of web services transactions", in Web Information Engineering, WISE, LNCS, vol. 3806, Springer, 2005, pp. 245-57.
36. S. Bottcher, L. Gruenwald, and S. Obermeier, "Reducing sub-transaction abort and blocking time within atomic commit protocols", in British National Conference on Databases (BNCOD), LNCS, vol. 4042, Springer, 2006, pp. 59-72.
37. X. Ding, J. Wei, and T. Huang, "User-defined atomicity constraints: a more flexible transaction model for reliable service composition", in Formal methods and Software Engineering, LNCS, vol. 4260, Springer, 2006, pp. 168-84.
38. W. Xu, W. Cheng, and W. Liu, "A transaction-aware coordination protocol for Web Services composition", in 7th International Conference on Web Information Systems Engineering, WISE 2006, pp. 126-31.
39. W. Zhao, L.E. Moser, and P.M. Melliar-Smith, "A reservation-based extended transaction protocol", IEEE Trans. Parallel and Distributed Systems, vol. 19, 2, 2008.
40. C. Batini, M. Lenzerini, and S.B. Navathe, "A Comparative Analysis of Methodologies for Database Scheme Integration", ACM Computing Surveys, vol. 18, 4, 1986, pp. 324-64.
41. A.P. Sheth, and J.A. Larson, "Federated database systems and managing distributed, heterogeneous and autonomous databases", ACM Computing Surveys, vol. 22, 3, 1990, pp. 183-226.
42. W. Kim, and J. Seo, "Classifying schematic and data heterogeneity in multdatabase systems", IEEE Computer, vol. 24, 12, 1991, pp. 12-8.
43. W. Litwin, Multidatabase Systems, Prentice-Hall, Englewood Cliffs, N.J., 1994.
44. A. Bouguettaya, B. Benatallah, and A. Elmagarmid, Interconnecting Heterogeneous Information Systems, Kluwer, 1998.
45. R. Ahmed, P. Smedt, W. Du, W. Kent, M. Ketabchi, W. Litwin, A. Rafii, and M. Shan, "The Pegasus heterogeneous, multidatabase system", IEEE Computer, vol. 24, 12, 1991, pp. 19-27.
46. E. Bertino, R. Gagliaro, G. Pelagatti, and L. Sbattella, "The COMANDOS integration system: an object oriented approach to the interconnection of heterogeneous applications", in International Workshop on Object-oriented Database Systems, LNCS, vol. 334, Springer, 1988, pp. 213-8.
47. M. Kaul, K. Drosten, and E. Neuhold, "Viewsystem: integrating heterogeneous information bases by object-oriented views", in IEEE International Conference on Data Engineering, 1991, pp. 2-10.
48. Y. Arens, C. Chee, C. Hsu, and C. Knoblock, "Retrieving and integrating data from multiple information sources", International Journal of Intelligent and Cooperative Systems, vol. 2, 2, 1993, pp. 127-57.
49. L. Serafini, and G. Ghidini, "Using wrapper agents to answer queries in distributed information systems", in Fourth International Conference on Multi Agent Systems, 2000.
50. D.W. Shipman, "The functional data model and the data language DAPLEX", ACM Transactions on Database Systems, vol. 6, 1, 1981, pp. 140-73.

51. A.E. James, "A database integration system and an example of its application", International Workshop on Database Re-engineering and Interoperability, Hong Kong Computer Society, Hong Kong, Plenum, 1995, pp. 207-308.
52. D. Allemang, and J. Hendler, Semantic Web for the Working Ontologist: Effective Modeling in RDFS and OWL, Morgan Kaufmann, 2008.
53. P. Hitzler, M. Kroetzsch, S. Rudolph, A. McGettrick, and J. Impagliazzo, Foundations of Semantic Web Technologies, Chapman and Hall, 2009.
54. M. Lytras, and A. Sheth (eds), Progressive Concepts for Semantic Web Evolution: Applications and Developments, Advances in Semantic Web and Information Systems (ASWIS) Book Series, Information Science Reference, 2009.
55. R. King, and R. Hull, "Semantic database modelling: survey, applications and research issues", ACM Computing Surveys, vol. 13, 3, 1987.
56. A.E. Robinson, "Current ideas in knowledge-base management systems", Journal of Information and Software Technology, Butterworth-Heinemann, 32, 4, 1990.
57. W3C, OWL Web Ontology Language,W3C Recommendation 10 Feb. 2004, (available at http://www.w3.org/TR/2004/REC-owl-ref-20040210/).
58. M. Huhns, "Agents on the Web", in IEEE Internet Computing, vol. 1, 3, 1997.
59. N. Gibbins, S. Harris, and N. Shadbolt, "Agent-based semantic web services", WWW 2003, pp. 710-717.
60. J.S. Rellermeyer, M. Duller, K. Gilmer, D. Maragkos, D. Papageorgiou, and G. Alonso, "The software fabric for the Internet of Things", in IoT 2008: First International Conference on the Internet of Things, Zurich, Switzerland, LNCS, vol4952., Springer, 2008, pp. 87-104.
61. A.P. Levich (ed.), On the way to understanding the Time Phenomenon: the Constructions of Time in Natural Science, Part I, World Scientific, New York, 1995.
62. D. Ceballos, and M. Sorrosal, "Time Aggregation Problems in Financial Time Series", *MS'2002 International Conference on Modelling and Simulation in Technical and Social Sciences*, Working Article of Institute of Time Nature Exploration, Girona 25-27/6, 2002, pp. 243-52.
63. J. Breitung, and N.R. Swanson, "Temporal Aggregation and Spurious Instantaneous Causality in Multiple Time Series Models", Journal of Time Series Analysis, vol. 23, 2002, pp. 651-66.
64. Y.N. Law, and C. Zaniolo, "Improving the accuracy of continuous aggregates and mining queries on data streams under load shedding", IJBIDM, vol. 3, 1, 2008, pp. 99-117.
65. B. Babcock, S. Babu, M. Datar, R. Motwani, and J. Widom, "Models and issues in data stream systems", Invited talk, in 21st ACM Symposium on Principles of Database Systems, ACM 2002, pp. 1-16.
66. D.J. Abadi, D. Carney, U. Cetintemel, M. Cherniack, C. Convey, S. Lee, *et al.*, "Aurora: a new model and architecture for data stream management", VLDB Journal, vol. 12, 2, pp. 120-39.
67. C. Aggarwal, (ed.), Data Streams: Models and Algorithms, Advances in Data Base Systems Series, vol. 31, Springer, 2007.
68. Y.N. Law, H. Wang, and C. Zaniolo, "Query languages and data models for database sequences and data streams", in International Conference on Very Large Databases (VLDB), 2004, pp. 492-503.
69. L. Gruenwald, H. Chok, and M. Aboukhamis, "Using data mining to estimate missing sensor data", in Seventh IEEE International Conference on Data Mining Workshops (ICDMW 07), Oct. 2007, pp. 207-12.
70. G. Chen, and D. Kotz, "Data-centric middleware for context-aware pervasive computing", Journal of Pervasive and Mobile Computing (PMC), vol. 4, 2, Apr. 2008, pp. 216-53.
71. Internet Archive, Home Page, www.archive.org/index.php.
72. S. Chaudhuri, and U Dayal, "An overview of data warehousing and OLAP technology", ACM SIGMOD Record, vol. 26,1, 1997, pp. 65-74.
73. A. Abelló, and C Martín, "The data warehouse: an object-oriented temporal database", in JISBD, 2003, pp. 675-84.
74. M. Rifaie, K. Kianmehr, R. Alhajj, and M.J. Ridley, "Data warehouse architecture and design" International Conference on Information Reuse and Integration, IRI 2008, IEEE, 2008.
75. J. Hendler, "Agents and the semantic web", Intelligent Systems, IEEE, vol. 16, 2, 2001, pp. 30-7.
76. J. Han, "Data mining for web intelligence", IEEE Computer, vol. 35, 2002, pp. 64-70.
77. Office of Public Sector Information (OPSI), The Data Protection Act 1998, UK (available at http://www.opsi.gov.uk/acts/acts1998/ukpga_19980029_en_1).
78. Information Commissioner's Office (ICO), UK, http://www.ico.gov.uk/what_we_cover/data_protection.asp.
79. K. Barker, M. Askari, M. Banerjee, K. Ghazinour, B. Mackas, M. Majedi, *et al.*, "A data privacy taxonomy", in 26th British National Conference on Databases, BNCOD, Birmingham, Springer, LNCS 5588, 2009, pp. 42-54.
80. Claremont 2008, The Claremont Report on Database Research, Berkeley University, (available at http://db.cs.berkeley.edu/claremont/claremontreport08.pdf).

AUTHORS



Joshua Cooper is the CEO of Hildebrand, a technology and innovation consultancy. Joshua joined Hildebrand from Boltblue, UK's largest mobile portal, where he was the Founder, CTO, and VP of Product Development. Under his leadership, Boltblue grew at an unprecedented speed, culminating in over 12 million registered users and ground-breaking product offerings in the mobile arena. With over 15 years of experience in software development, Joshua is well known for his expertise in Internet technologies. Prior to Boltblue, he was involved with two start-up firms in the United States (Infominers. com and Flow Management Technologies), working in the data warehousing and process management industries, and was a team leader for software development at Nortel Networks. Joshua holds a BSc in Biomedical Engineering from Rensselaer in New York, USA.

E-mail: jcooper@hildebrand.co.uk

Anne James is a Professor of Data Systems Architecture at Coventry University, UK. She obtained her BSc degree at Aston University, UK, in 1980, and her PhD at the University of Wolverhampton, UK, in 1986. She has been involved in database research for many years since completing her PhD in the area of database models. She has authored over 100 journals and conference research articles in the field; many with a focus on data heterogeneity, integration, and cooperative systems. Anne has contributed to many international conference committees and steering groups. Her current research interests are distributed systems and applications of data management methods to new areas of computing. She is also interested in new methods of teaching in higher education, particularly in the database area.

E-mail: a.james@coventry.ac.uk

DOI: 10.4103/0256-4602.55275; Paper No TR 226_09; Copyright © 2009 by the IETE