The Communications Skin: Challenges and Dangers for Ubiquitous Data Management[1]

Dimitrios Katsaros[1,2] **Yannis Manolopoulos[1]**
Alexandros Nanopoulos[1] **Apostolos Papadopoulos[1]**

[1]Department of Informatics
Aristotle University of Thessaloniki
54124 Thessaloniki, GREECE

[2]Dept. of Computer & Communication Engineering
University of Thessaly
38221 Volos, GREECE

{dimitris,manolopo,alex,apostol}@ delab.csd.auth.gr

Abstract

The recent technological advances in distributed information gathering from a given location or geographical region by deploying a large number of tiny microsensor nodes, and the progress of wireless communications along with the convergence of Internet technologies and mobile computing have totally reshaped the research and development agenda. This progress has created the anticipation for a *communications skin*, which will cover significant parts of our planet, and, similar to the human's skin functionality, will gather information from its surroundings; it will process it, filter it, aggregate it, store it and finally make decisions based on it. The realization of such a vision requires the interplay of several disciplines, like hardware, communications and data management technologies.

In this paper, will present the recent developments related to various critical aspects of data management in wireless sensor network environments. The main goal of the paper is to highlight the most important dimensions of the new era for information manipulation, which resulted from this networking paradigm and also to present the specific dangers for the privacy, which come out from the deployment of sensor networks.

Keywords. *Enabling Technologies, Wireless Sensor Networks, Sensor/Stream Databases, Pervasive Computing.*

Introduction

The technological advances in distributed information gathering from a geographical region by deploying a large number of tiny microsensor nodes, and the progress of wireless communications along with the convergence of Internet technologies and mobile computing have totally reshaped the research and development agenda. This progress has created the anticipation for a *communications skin*, which will cover significant parts of our planet, and, similar to the human's skin functionality, will gather information from its surroundings; it will process it, filter it, aggregate it, store it and finally make decisions based on it. Central to such a vision, is the development and deployment of large-scale wireless sensor networks (Karl and Willig, 2005).

[1] Research supported by a ΓΓΕΤ grant in the context of the project "Data Management in Mobile Ad Hoc Networks", funded by ΠΥΘΑΓΟΡΑΣ national research program.

A wireless sensor network usually consists of a hundred or a few thousands of sensor nodes that are densely deployed inside or close to a phenomenon under observation. The position of sensor nodes need not be engineered or predetermined, but may be random, for instance, in inaccessible areas (ocean bottom or battle-fields) or disaster relief operations. Sensor networks can be considered as a type of ad hoc wireless networks (Basagni et al., 2004), where sensor nodes are, usually but necessarily, stationary. There are many similarities between ad hoc wireless networks and sensor networks, such as energy constraints and dynamic network topology. However, the number of sensor nodes can be several orders of magnitude higher than the nodes in an ad hoc wireless network, and hence sensor nodes are more densely deployed. Also, sensor nodes are more prone to failures than those of an ad hoc wireless network. Therefore, the topology of a sensor network is changed mainly by switch-on/off of sensor nodes, whereas the topology of an ad hoc wireless network is changed by the movement of mobile hosts. In addition, sensor nodes may not have global identification because of the large amount of overhead and large number of sensors. Finally, nodes in ad hoc wireless networks are strictly peer-to-peer, whereas nodes in sensor networks form (at least at the first years of development) a two-level hierarchy with the base station being the gathering point.

A unique feature of sensor networks is the cooperative behavior of sensor nodes. Sensor nodes are usually fitted with an onboard processor. Instead of sending the raw data to the nodes responsible for the fusion, they use their processing abilities to locally carry out simple computations and transmit only the required and partially processed data. A sensor system normally consists of a set of sensor nodes operated on a limited battery and a base system without any battery constraint. Typically, the base station serves as the gathering point for the collected data (through fusion). The base station also broadcasts different control commands to sensor nodes. A characteristic example of a sensor network is illustrated in Figure 1.

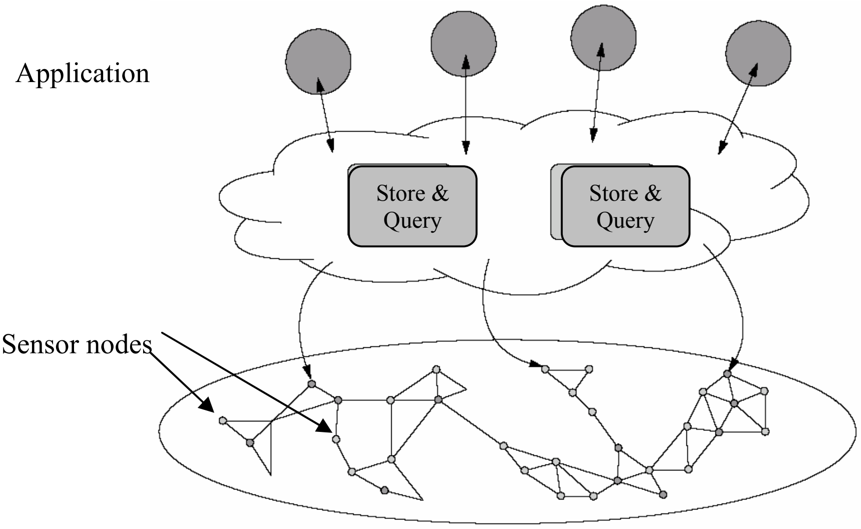

Figure 1. Typical sensor network.

The application areas of sensor networks include health, military, and civilian. In military application, the rapid deployment, self-organization, and fault-tolerance characteristics of sensor nodes make them a promising sensing technique for command, control, communication, computing, intelligence, surveillance, reconnaissance, and targeting systems. In health care, sensor nodes can be used to monitor patients and assist disabled patients. Other applications include managing inventory, monitoring product quality, and monitoring disaster areas.

In this paper, we address the technologies and the most important data management issues involved towards realizing such a communications skin. Firstly, the paper will describe the generic issue of routing information in a wireless sensor network, i.e., the issues involved in discovering relevant data in a large network of tiny storage devices, and explain the novel paradigm of *data-centric networking*, as opposed to the traditional address-based networking. Delving into the most critical aspects of the data-centric networking area, the paper will discuss the issues of broadcasting and network clustering for sensor networks. Taking also an orthogonal direction, the paper will also focus on the storage issues arising in the context of sensor networks, describing the novel architecture of *sensor database systems*, which extend the traditional centralized and distributed databases. Finally, the paper will touch the privacy and security risks involved for individuals due to ubiquitous presence of sensors, using as an example the technology of the RFIDs. Apart from these aspects, sensor networks introduce additional data management problems (Zhao and Guibas, 2004; Wu, 2005), but we have a strong feeling that these form the fundamental ones at this stage of development.

Data-centric Networking

Methods for routing and discovering information in traditional (wired or wireless) networks was developed around the notion and existence of a Unique-ID, associated with each node and also with the exploitation of the knowledge about where the information that someone sought was stored. Such assumptions are no longer valid in wireless sensor networks, where each sensor node has not anymore a globally (in the whole sensor network) unique ID, but a local one, it has an ID at all. Moreover, when searching for information in a wireless network, we may not know or may not be interested in where the relevant data reside. Therefore, we need novel and (most of the time) application dependent techniques for retrieving the information we are seeking for, which must consider the energy-constraint nature of sensor nodes and also the volatile topology of the network. These requirements caused the turn from the traditional address-based networking to the novel and network-adaptive paradigm of *data-centric networking*.

Intanagonwiwat et al. (2003) were the first who laid the foundational for implementing this networking paradigm. They introduced a data dissemination paradigm called *Directed Diffusion* for sensor networks, based on a flat network topology. The query is disseminated (flooded) throughout the network with the queried node acting as a *source* and gradients are set up toward the requesting node (*sink*) to find the data satisfying the query. As one can observe in Figure 2, the query is propagated toward the requesting node along multiple paths shown by the light lines. The arcs show how the query is directed toward the event of interest, similar to a ripple effect. Events (data) start flowing toward the requesting node along multiple paths. To prevent further flooding, a small number of paths can be reinforced (shown by dark lines in the figure) among a large number of paths initially explored to form the multi-hop routing infrastructure so as to prevent further flooding. A significant advantage of this protocol is the ease of creating multiple paths between communicating nodes, thereby alleviating congestion and providing robustness in the presence of failures.

Apart from this pioneering approach, newer, sophisticated protocols were proposed in the sequel, which addressed more successfully the challenges of the wireless sensor networks (Kim, Abdelzaher and Kwon, 2005; Luo, Ye, Cheng, Lu and Zhang, 2005; Ye, Zhong, Lu, and Zhang, 2005).

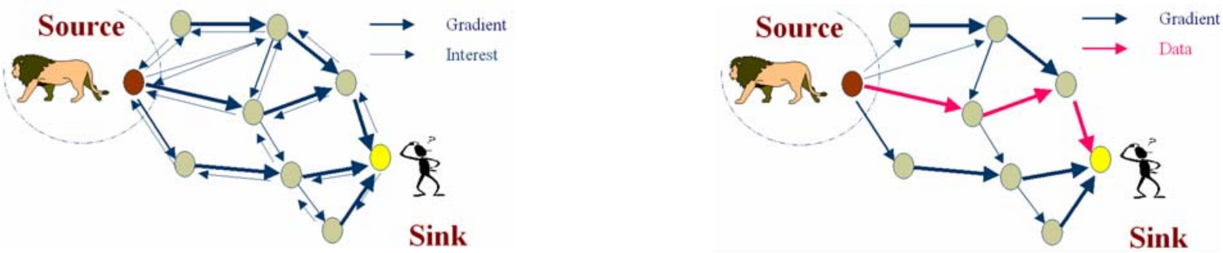

Figure 2. Directed Diffusion in wireless sensor networks.

Broadcasting and Network Clustering

In such unstructured wireless networks, like mobile ad hoc and sensor nets, efficient mechanisms need to be devised which will guarantee that any message transmitted from a source will eventually be delivered to its final destination, provided that the network is not partitioned and making minimum possible number of transmission, because the penalty of communication can be severe. Energy conservation is a vital goal in wireless networks for prolonging the longevity of the sensor network or for guaranteeing as much power-independence as possible for the mobile hosts. To achieve energy savings, mobile nodes support two generic modes of operation, the active mode, which is a fully operational state, and the doze mode, which is a power saving state. The ratio of energy consumption between the two modes is usually an order of magnitude (Viredaz, Brakmo and Hamburgen, 2003). Similarly, sensor nodes can be in one of three active states — transmit, receive, idle — or in sleep state; a sensor in the sleep state consumes 7–20 times less energy than when it is in the idle state (Feeney, 2004).

Broadcasting is the task where a source node sends the same message to all the nodes in the network. In the *one-to-all model*, which is the most common, transmission by each node can reach all nodes that are within radius distance from it, whereas in the *one-to-one model*, each transmission is directed toward only one neighbor (using narrow-beam directional antennas or separate frequencies for each node). The traditional solution to the broadcasting problem is *blind flooding*, whereby each node receiving the message will retransmit it to all its neighbors. The only "optimization" applied to this solution is that nodes remember messages received for flooding, and do not act when receiving repeated copies of the same message. However, blind flooding causes unnecessary collisions and bandwidth waste, with many nodes not receiving the message as a consequence.

An important requirement for a broadcasting protocol is the *reliability*, that is the ability of a broadcast protocol to reach all the nodes in the network, and *saved rebroadcasts*, that is the ratio of the number of rebroadcasting nodes under the broadcasting protocol to the that number under a naïve blind flooding broadcasting protocol. Towards achieving these goals a number of broadcasting protocols have been proposed which rely on the concept of *(connected) dominating sets, CDS*. A CDS of a graph is a set of nodes, such that any node of the graph either belongs to the

CDS or is a neighbor of a node of the CDS. Nodes that belong to a dominating set will be called *internal nodes*. Abstracting a wireless sensor network as a graph, Wu and Li, 2001, were among the first that used this notion for reducing the number of rebroadcasting nodes. They introduced the concept of an *intermediate node*, which formed the connected dominating set. A node A is an intermediate node if there exist two neighbors, B and C, of A that are not direct neighbors themselves. For example, in Figure 3, nodes C and K in are not intermediate nodes, while the other nodes are. Wu and Li also introduced two rules that considerably reduce the number of internal nodes in the network. Rule 1 is as follows. Consider two intermediate neighboring nodes v and u. If every neighbor of v is also a neighbor of u, and ID(v) < ID(u), then node v could be eliminated from the CDS. We may also say that node v is “covered” by node u. Observe that retransmission by v, in this case, is covered by retransmission of u, since any node that might receive the message from v will receive it instead from u. This rule can be generalized to exploit pair of nodes covering another node (Rule 2) or even triples of nodes and so on, which cover a node. After finding the intermediate nodes, these will be the only rebroadcasing nodes. Later, more improved techniques were proposed based on the concept of CDS, which can be found in (Basagni et al., 2004; Stojmenovic et al., 2002; Stojmenovic, 2005).

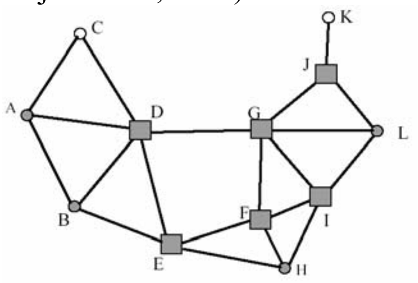

Figure 3. The concept of intermediate nodes.

Another way of minimizing the data transmissions over long distances is to cluster the sensor network so that signaling and control overheads can be reduced, while critical functions such as media access, routing, and connection setup could be improved. While all nodes typically function as switches or routers, one node in each cluster is designated as the cluster head (CH) and traffic between nodes of different clusters must always be routed through their respective CHs or gateway nodes that are responsible for maintaining connectivity among neighboring CHs. The number of tiers within the network can vary according to the number of nodes, resulting in hierarchical network architecture as shown in Figure 4.

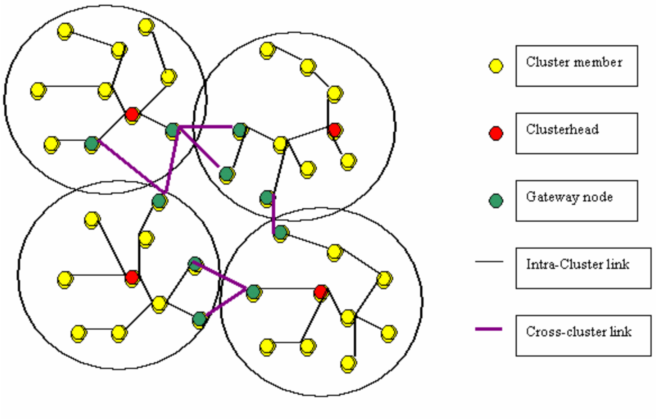

Figure 4. Clustered wireless sensor network.

Various methods have been developed for the formation and maintenance of clusters in ad hoc and in wireless sensor networks. Among them the protocols described in (Basagni, 1999; Amis et al., 2000; Heinzelman, Chandrakasan, and Balakrishnan, 2002) are very popular in the research literature. The LEACH clustering and dissemination protocol (Heinzelman et al., 2002) is one of the initial data-gathering protocols introduced by MIT researchers Heinzelman et al. (2000). Each cluster has a CH that periodically collects data from its cluster members, aggregates it, and sends it to an upper-level CH. Only he CH needs to perform additional data computations such as aggregation, etc., and the rest of the nodes sleep unless they have to communicate with the CH. To evenly distribute this energy consumption, all the nodes in a neighborhood take turns to become the CH for a time interval called the cluster period.

Sensor Databases

There have been a spectrum of approaches for constructing sensor storage systems. In the simplest, sensors stream data or events to a server for long-term archival storage (Bonnet, Gehrke and Seshadri, 2000), where the server often indexes the data to permit efficient access at a later time. Since sensors may be several hops from the nearest base station, network costs are incurred; however, once data is indexed and archived, subsequent data accesses can be handled locally at the server without incurring network overhead. In this approach, the storage is centralized, reads are efficient and cheap, while writes are expensive. Further, all data is propagated to the server, regardless of whether it is ever used by the application.

An alternate approach is to have each sensor store data or events locally (e.g., in flash memory), so that all writes are local and incur no communication overheads. A read request, such as whether an event was detected by a particular sensor, requires a message to be sent to the sensor for processing. More complex read requests are handled by flooding. Thus, in this approach, the storage is distributed, writes are local and inexpensive, while reads incur significant network overheads. Requests that require flooding, due to the lack of an index, are expensive and may waste precious sensor resources, even if no matching data is stored at those sensors. A hybrid approach is exploited in TSAR (Desnoyers, Ganesan and Shenoy, 2005), that reflects and exploits the multi-tier nature of emerging sensor networks, where the application is comprised of tens of tethered sensor proxies (or more), each controlling tens or hundreds of untethered sensors (see Figure 5).

Depending on the nature of the application, the types of queries injected in a sensor network can vary. Queries posed to a sensor network are usually classified as one-time queries or periodic queries. “One time” queries are injected at random times to obtain a snapshot view of the data attributes, but “periodic” queries retrieve data from the source nodes after regular time intervals. We now concentrate on periodic queries that are long running; that is, they retrieve data from the source nodes for a substantially long duration, possibly the entire lifetime of the network. We can classify queries into three different categories based on the nature of data processing demanded by the application.

- *Simple queries*. These are stand alone queries that expect an answer to a simple question from all or a set of nodes in the network. For example, “Report the value of temperature.”
- *Aggregate queries*. These queries require collaboration among sensor nodes in a neighborhood to aggregate sensor data. Queries are addressed to a target region consisting of many nodes in a geographically bounded area instead of individual nodes. For example, “Report the average temp of all nodes in region X.”

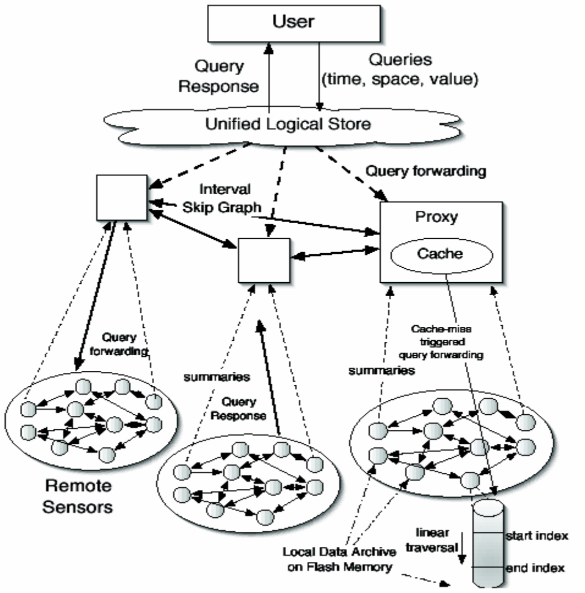

Figure 5. The TSAR storage architecture.

- *Approximate queries*. These are queries that require data summarization and rely on synopsis data structures to perform holistic data aggregation in the form of histograms, isobars, contour maps, tables, or plots. For example, “Report the contours of the pollutants in the region X.”
- *Complex queries*. If represented in SQL, these queries would consist of several joins nested or condition based sub-queries. Their computation hierarchy is better represented by a query tree. For example, “Among regions X and Y, report the average pressure of the region that has higher temperature.”

Security and Privacy Risks in Wireless Sensor Networks

One of the most well-known and practical application of wireless sensor networks are the RFID technology (Stanford, 2003). Radio Frequency Identification (RFID) tags are small, wireless devices that help identify objects and people. Thanks to dropping cost, they are likely to proliferate into the billions in the next several years — and eventually into the trillions. RFID tags track objects in supply chains, and are working their way into the pockets, belongings, and even the bodies of consumers. Although different RFID systems have been in use for years, popular accounts of RFID technology typically refer to the Electronic Product Code (EPC).

RFID technology poses unique privacy and security concerns because humans cannot sense the Radio Frequency (RF) radiation used to read tags, and the tags themselves typically maintain no history of past readings. As a result, tags are promiscuous: they can be read by entities other than their owners and without their owners’ knowledge. Further, both tags and readers can be covertly embedded in the environment; short-range readers can be small enough to fit into a cell phone. We will examine the security and privacy risks in the context of the EPC network.

We can identify several threats, related to both corporate and private security. Here we list the most important of them:

- Corporate espionage threat. Tagged objects in the supply chain make it easier for competitors to remotely gather supply chain data, which is some of industry's most confidential information.
- Competitive marketing threat. Tagged objects make it easier for competitors to gain unauthorized access to customer preferences and use the data in competitive marketing scenarios.
- Action threat. In this threat, an individual's behavior (or possibly his or her intent) is inferred by monitoring the action of a group of tags.
- Association threat. When a customer purchases an EPC-tagged item, the customer's identity can be associated with the item's electronic serial number even involuntary.
- Location threat. Placing covert readers at specific locations creates two types of privacy threats. First, individuals carrying unique tags can be monitored and their location revealed if the monitoring agency knows the tags associated with those individuals. Second, a tagged object's location — regardless of who (or what) is carrying it — is susceptible to unauthorized disclosure.
- Preference threat. With the EPC network, the tag on an item uniquely identifies the manufacturer, the product type, and the item's unique identity. This exposes otherwise unavailable customer preferences to competitive (and inquisitive) forces at low marginal cost.

Addressing the privacy and security challenges of RFID technology is not an easy task. Regulations for the RFID's use are needed, which could be legislated or adopted voluntarily. The fact that the debate about RFID systems' privacy and security is taking place far ahead of the actual ubiquitous deployment is a good sign.

Conclusions

Sensor networks represent a paradigm shift in computing. Traditional computing involves computers directly interacting with human operators. However, in the near future, hundreds of computers will be embedded deep in the world around us. In this environment, the computers themselves will interact more directly with the physical world. They will sense their environments directly, compute necessary responses, and execute them directly. In this paper, we presented the recent developments related to various critical aspects of data management in wireless sensor network environments. The main goal of the paper was to highlight the most important dimensions of the new era for information manipulation, which resulted from this networking paradigm and also to present the specific dangers for the privacy, which come out from the deployment of sensor networks.

References

Amis, A.D., Prakash, R., Huynh, D. and Vuong, T. (2000). "Max-min d-cluster formation in wireless ad hoc networks", *Proceedings IEEE Conference on Computer Communication (INFOCOM)*, pp. 32–41.

Basagni, S. (1999). "Distributed clustering for ad hoc networks", *Proceedings International Symposium on Parallel Architectures, Algorithms and Networks (ISPAN)*, pp. 310–315.

Basagni, S., Conti, M., Giordano, S. and Stojmenovic, I. (2004). "*Mobile Ad Hoc Networking*", IEEE/Wiley-Interscience.

Bonnet, P., Gehrke, J., Seshadri, P. (2000). "Querying the physical world", *IEEE Personal Communications*, 7(5), pp. 10–15.

Desnoyers, P., Ganesan, D. and Shenoy, P. (2005). "TSAR: a two tier sensor storage architecture using interval skip graphs", *Proceedings ACM International Conference on Embedded Networked Sensor Systems (SenSys)*, pp. 39–50.

Feeney, L.M. (2004). "Energy efficient communication in ad hoc wireless networks", In *Mobile Ad Hoc Networking* (Basagni, S., Conti, M., Giordano, S. and Stojmenovic, I. eds), IEEE/Wiley, pp. 301–327.

Garfinkel, S.L., Juels, A. and Pappu, R. (2005). "RFID privacy: an overview of problems and proposed solutions", *IEEE Security & Privacy*, pp. 34–43.

Heinzelman, W.B., Chandrakasan, A.P. and Balakrishnan, H. (2002). "An application-specific protocol architecture for wireless microsensor networks", *IEEE Transactions on Wireless Communications*, 1(4), pp. 660–670.

Intanagonwiwat, C., Govindan, R., Estrin, D., Heidemann, J.S. and Silva, F. (2003). "Directed diffusion for wireless sensor networking", *IEEE/ACM Transactions on Networking*, 11(1), pp. 2–16.

Juels, A. (2006). "RFID security and privacy: a research survey", *IEEE Journal on Selected Areas in Communications*, 24(2), pp. 381–394.

Karl, H. and Willig, A. (2005). "*Protocols and Architectures for Wireless Sensor Networks*", Wiley-Interscience.

Kim, H.S., Abdelzaher, T.F. and Kwon, W. H. (2005). "Dynamic delay-constrained minimum-energy dissemination in wireless sensor networks", *ACM Transactions on Embedded Computing Systems*, 4(3), pp. 679–706.

Luo, H., Ye, F., Cheng, J., Lu, S. and Zhang, L. (2005). "TTDD: Two-tier data dissemination in large-scale wireless sensor networks", *ACM/Kluwer Wireless Networks*, 11(1-2), pp. 161–175.

Stanford, V. (2003). "Pervasive computing goes the last hundred feet with RFID systems", *IEEE Pervasive Computing*, pp. 9–14.

Stojmenovic, I. (2005). "*Handbook of Sensor Networks: Algorithms and Architectures*", Wiley-Interscience.

Stojmenovic, I., Seddigh, M. and Zunic, J.D. (2002). "Dominating sets and neighbor elimination-based broadcasting algorithms in wireless networks", *IEEE Transactions on Parallel and Distributed Systems*, 13(1), pp. 14–25.

Viredaz, M.A., Brakmo, L.S. and Hamburgen, W.R. (2003). "Energy management on handheld devices", *ACM Queue*, 1(7), pp. 44–52.

Wu, J. (2005). "*Handbook of Theoretical and Algorithmic Aspects of Sensor, Ad Hoc Wireless, and Peer-to-Peer Networks*", Auerbach Publications.

Wu, J. and Li, H. (2001). "A dominating-set-based routing scheme in ad hoc wireless networks", *Telecommunication Systems*, 18(1–3), pp. 13–36.

Ye, F., Zhong, G., Lu, S. and Zhang, L. (2005). "GRAdient Broadcast: A robust data delivery protocol for large scale sensor networks", *ACM/Kluwer Wireless Networks*, 11(3), pp. 285–298.

Zhao, F. and Guibas, L. (2004). "*Wireless Sensor Networks: An Information Processing Approach*", Elsevier/Morgan Kaufmann.