Kumar's, Zipf's and Other Laws: How to Structure an Optimum Large-Scale Wireless (Sensor) Network?

M. Dohler[1], T. Watteyne[1,2], D. Barthel[1], F. Valois[2], J. Lu[1,2]
[1]France Télécom R&D, Meylan Cedex, F-38243, France
[2]ARES INRIA / CITI, INSA-Lyon, F-69621, France
E-mail: mischa.dohler@orange-ftgroup.com

Abstract— **Networks with a very large number of nodes are known to suffer from scalability problems, influencing throughput, delay and other quality of service (QoS) parameters. Examples of such networks are cellular networks and recently emerged wireless sensor networks (WSNs). Cellular networks – often composed of several million of nodes – have mitigated the scalability problem by deploying a dense hierarchical and centralised infrastructure. WSNs – expected to be composed of several tens of thousands of nodes – will have to find a suitable deployment solution, trading performance against cost and energy consumption. Mainly applicable to WSNs, aims to give some fundamental indications on an optimum structuring approach for large-scale wireless networks, where optimality refers to various criteria to be exposed in the paper. To this end, various laws known from different domains will be invoked to draw knowledgable design guidelines. Optimum network structures are derived and discussed for a plethora of different scenarios.**

I. INTRODUCTION

The France Telecom Group constitutes one of the biggest integrated operators worldwide. It offers a variety of services to its clientele, including mobile and fixed telephony, wired and wireless internet, as well as integrated home and business solutions. It hence owns large scale wired and wireless networks, with the latter traditionally being composed of cellular networks and lately also of WSNs.

Through cellular systems, the France Telecom Group has already been offering 2G and 3+G wireless voice and date services for several years. Whilst subscriber numbers were low at the beginning, these have risen dramatically over the past years and hence triggered the need to continuously augment the capacity of the cellular network. The invoked solution consisted of introducing a hierarchical communication structure, i.e. several uses were, e.g., connected to a 3G Node B, several of these Node Bs to a radio network controller, etc.

Whilst this facilitated scalability, such a solution is clearly expensive; for instance, a Node B may cost several hundred thousand Euros. The question hence arises whether the approach taken is optimum or whether another solution would have been more appropriate. Whilst the answer depends on many factors and the limited scope of the paper prohibits all of these to be taken into account, we aim to give some indicative answers on the optimality of the hierarchical approach.

As for WSNs, the France Telecom Group hopes to offer more complete services by creating and facilitating ambient environments, which interface with incumbent and emerging services. For this reason, France Telecom has strong R&D activities in the area of WSNs – corroborated by the leadership of the ARESA project [1].

Sensor networks have been researched and deployed for decades already; their wireless extension, however, has witnessed a tremendous upsurge in recent years. This is mainly attributed to the unprecedented operating conditions of WSNs, i.e. a potentially enormous amount of sensor nodes reliably operating under stringent energy constraints. WSNs allow for an untethered sensing of the environment. It is anticipated that within a few years, sensors will be deployed in a variety of scenarios, ranging from environmental monitoring to health care, from the public to the private sector, etc. They will be battery-driven and deployed in great numbers at once in an ad hoc fashion, requiring communication protocols and embedded system components to run in an utmost energy efficient manner.

Prior to large-scale deployment, however, a gamut of problems has still to be solved which relates to various issues, such as the design of suitable software and hardware architectures, development of communication and organization protocols, validation and first steps of prototyping, until the actual commercialization. Of prime concern among industrialists, however, is currently the WSNs' scalability [2].

The aim of this paper is to give some insights into the scalability issue. To this end, the paper is structured as follows. In Section II, we briefly dwell on the definition of 'scalability' and 'optimum'. In Section III, we will discuss the role and importance various scaling laws. Section IV is dedicated to the application of these scaling laws to different communication structures. Finally, conclusions are drawn in Section V.

II. DEFINITIONS

Before embarking onto the quantification of optimum scalable architectures and various scaling laws, we shall subsequently define 'scalability' and 'optimality'.

A. Scalability

An algorithm or architecture is said to be scalable if it can handle increasingly bigger and complex problems. Whilst such basic notion is intuitive, the term 'scalability' has so far evaded a generally-accepted definition. To this end, Hill [3] claims that the "use of the term adds more to marketing potential than technical insight". He concludes that no rigorous definition is applicable to scalability and he challenges "the technical community to either rigorously define scalability or stop using it to describe systems."

To attempt a definition that is at least applicable to large-scale wireless systems, we will modify the approach taken by [3]. We first introduce

$$\eta_{12} = \frac{\mathfrak{F}_{\mathfrak{A}}(N_2, S_2)/N_2}{\mathfrak{F}_{\mathfrak{A}}(N_1, S_1)/N_1} \tag{1}$$

to be the relative efficiency between two systems obeying the same architecture $\mathfrak{A}$, consisting of N_1 and N_2 nodes, respectively, tackling some problems of size S_1 and S_2, respectively, and being gauged by some 'positive' average network-wide attribute $\mathfrak{F}$. This attribute could, for instance, be the total average network throughput, the inverse of the average end-to-end delay, etc. The problem size is determined by the 'problem' the system aims to solve; for instance, the problem of a network may be to deliver data from every node (cellular system), or to measure and deliver a fixed set of measurements (data aggregating WSN), or simply to deliver just one measurement (alert triggered WSN), etc.

To facilitate a definition, we assume that

- the difference between the two systems approaches infinity, i.e. with $N_1 = N$, $N_2 = N + \Delta$, $\Delta \to \infty$;
- N is sufficiently large such that the attribute $\mathfrak{F}$ holds with sufficiently high probability; and
- the problem size of the larger system does not decrease, i.e. $S_2 \geq S_1$.

With these assumptions, we now define an architecture $\mathfrak{A}$ to be scalable w.r.t. attribute $\mathfrak{F}$ if

$$\eta \triangleq \lim_{\Delta\to\infty} \frac{\mathfrak{F}_{\mathfrak{A}}(N+\Delta, S_2)/(N+\Delta)}{\mathfrak{F}_{\mathfrak{A}}(N, S_1)/N} \geq O(1). \tag{2}$$

In other words, this means that we consider an architecture scalable if the network-wide performance attribute, scaled by the number of nodes involved, does not decrease with an increasing number of nodes and a non decreasing problem space. The requirement on N being sufficiently large stems from the the fact that many network-wide attributes, such as average throughput, can only be quantified if the number of involved nodes is sufficiently large.

B. Optimality

We are primarily interested in the optimality of a given architecture, i.e. network topology with associated communication protocols. With this in mind, an optimal architecture $\hat{\mathfrak{A}}$ w.r.t. attribute $\mathfrak{F}$ is defined as the one which, over all possible architectures $\mathfrak{A}$, maximises η. Note that although this definition is intuitive, it is often difficult to prove optimality over all possible architectures $\mathfrak{A}$.

III. SCALING LAWS

The first question we pose is when a network has to be considered large. To exemplify this problem, we have presupposed systems with and without internal conflicts [4]. For instance, two systems without conflicts are our circle of true friends (comprised by a small number of elements) and the soldiers of an ant colony (comprised by a large number of elements). On the other hand, two systems with conflicts, frictions and competition are, for example, a few children left on their own (comprised by a small number of elements) and state without government (comprised by a large number of elements). As such, 'large' is hence not about size [4]. It is rather about managing existing and emerging conflicts, and hence the amount of overheads needed to facilitate (fair) communication.

This overhead is well reflected in the efficiency η, which needs to be maximised for a given attribute $\mathfrak{F}$. If the attribute is cost, then a centralised hierarchical solution is unlikely to be the optimum architecture; whereas, if it is throughput or delay, then a centralised entity is likely to be a good choice. Using different scaling laws, we will use different attributes to judge upon the scalability of considered architectures.

A. *Kumar & Gupta's Throughput Scaling Law*

This milestone contribution [5] quantifies the theoretically achievable per-node capacity assuming that every node wishes to communicate with every other node. The architecture is assumed to be flat and hence does not contain any structural elements, such as hierarchies or clusters. They have determined that, assuming random deployment of the nodes in a large network, the per-node capacity scales with $1/\sqrt{N \log N}$ and the average network capacity

$$\Theta \propto \frac{N}{\sqrt{N \log N}}. \tag{3}$$

With reference to (2), the architecture $\mathfrak{A}$ is flat, the attribute $\mathfrak{F}$ is Θ and the problem size S is related to the total number of nodes in the network N (thus certainly not decreasing with an increasing number of nodes). The relative efficiency can hence be calculated as

$$\eta = O(1/\sqrt{\Delta \log \Delta}) < O(1), \tag{4}$$

revealing that w.r.t. network capacity the architecture is not scalable. This seems to be at odds with Figure 1, which shows the average network capacity to be increasing with an increasing number of nodes. However, this increase is not fast enough to guarantee non-diminishing throughput per node (i.e., each problem in S to be attended to with the same resources). Indeed, the throughput per node decreases rapidly with an increasing number of nodes. In other words, no matter what we try, we cannot design a protocol for large networks which is scalable according to (2) w.r.t. the total average network throughput. Since [5] has proven above bounds to be the result of optimum communication protocols, we concur that the only way to achieve some form of scalability is to invoke architectures different from flat ad hoc.

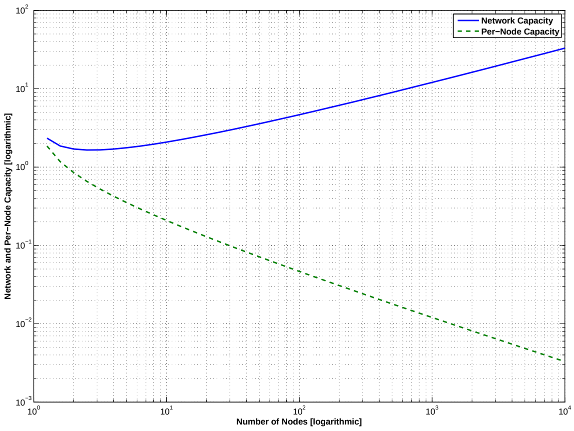

Fig. 1. Kumar & Gupta's network and per-node capacities.

B. *Odlyzko & Tilly's Value Scaling Law*

Mainly economically driven, various efforts in the past have been dedicated to establishing the value of a network in dependency of the number of its elements N. Sarnoff's Law quantifies the value of a broadcast network to be proportional to N [6]. Reed's Law claims that with N members you can form communities in 2^N possible ways; the value hence scales with 2^N [7]. Metcalfe's Law, unjustifiably blamed for many dot-com crashes, claims that N members can have $N(N-1)$ connections; the value hence scales with N^2.

Since a large-scale network – be it a cellular network or a WSN – is not truly of broadcast nature, nor do nodes form all possible communities, nor does every node communicate with every other node, another value scaling law is required to quantify the network's behaviour.

To this end, Odlyzko and Tilly have proposed a value scaling which is proportional to $N \log N$ [9]. Their argumentation bases on Zipf's Law [10], which states that if one orders a large collection of entities by size or popularity, the entity ranked k-th, will be in value about $1/k$ of the first one. The added value of a node to communicate with the remaining nodes is hence $1 + 1/2 + 1/3 + \cdots + 1/(N-1) \propto \log N$; the total value V of the network hence scales with

$$V \propto N \log N. \tag{5}$$

Among others, this law has been found to describe accurately the merging and partitioning of companies of unequal size [9]. It can also be used to describe the routing behaviour in WSNs, where the nodes of highest value are along the shortest projection of the routing path and nodes away from this routing path lose respectively in value.

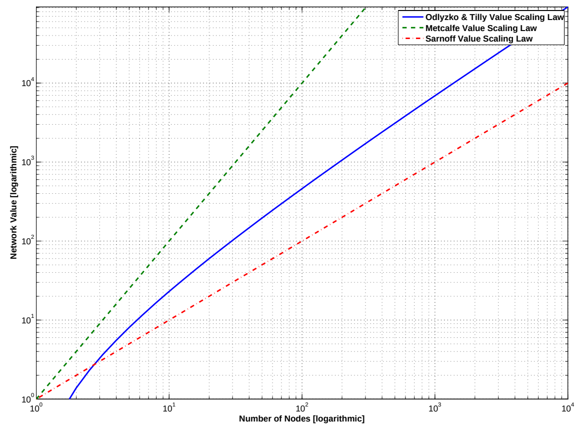

Fig. 2. Odlyzko & Tilly's fast increasing network value.

With reference to (2), the architecture $\mathfrak{A}$ has prioritised nodes, the network attribute $\mathfrak{F}$ is its value V and the problem size S is related to the total number of nodes in the network N (thus certainly not decreasing with an increasing number of nodes). The relative efficiency can hence be calculated as

$$\eta = O(\log \Delta) > O(1), \tag{6}$$

revealing that w.r.t. network value the architecture is scalable. Indeed, from Figure 2 depicting Sarnoff's, Odlyzko & Tilly's and Metcalfe's scaling laws, the network value increases rapidly with an increasing number of nodes. In other words, forming communities (e.g. chat rooms in the Internet community or clusters in a WSN) is favourable to the scaling of the network value.

IV. APPLICATION OF SCALING LAWS

The above fundamental scaling laws gave us the following insights:

1) w.r.t. (4), the network throughput decreases with an increasing amount of nodes due to the increasing amount of required links and hence counteracting scalability; and
2) w.r.t. (6), the network value increases with an increasing amount of nodes due to the increasing amount of facilitated links and hence enabling scalability.

This apparent discrepancy is due to both laws describing two inherently different but dependent attributes of an architecture. Indeed, the value of an architecture cannot be guaranteed if the throughput over the required links cannot be maintained. Subsequently, we hence aim at exploiting and trading this dependency to find architectures optimum under given assumptions.

Whilst not proven to be the optimum architectural solution $\hat{\mathfrak{A}}$, the question which naturally arises in this contents is whether clusters and hierarchies will improve the scalability of the architecture.

As described before, clusters in form of cells and hierarchies incorporating tiers of mobile stations, tiers of node Bs, tiers of radio network controllers, etc., have been used with success in cellular networks. Clustering has also been introduced as a means of structuring a wireless multihop network, such as WSNs.

It is commonly assumed that using such a hierarchical architecture yields better results than using a flat topology where all nodes have the same role. In particular, it is assumed that clusters can reduce the volume (but not contents!) of inter-node communication, and hence increase the network's lifetime.

To our knowledge, solely [11] has formally studied these aspects for WSNs only with focus on energy consumption. The authors have shown that clustered architectures out-perform non-clustered ones in a selected number of cases. In particular, the WSN's energy consumption is proven to be reduced only when data aggregation is performed by the cluster-heads.

Subsequently, we will hence examine a few selected clustering approaches and quantify their scalability w.r.t. some important attributes.

A. *Clustering with Odlyzko & Tilly's Law*

Based on Odlyzko and Tilly's value scaling law, we introduce a normalised network value V', which we define as the ratio between the value given in (5) and the number of links needed to support such connected community. This definition hence incorporates the required links into the value of the spanned network.

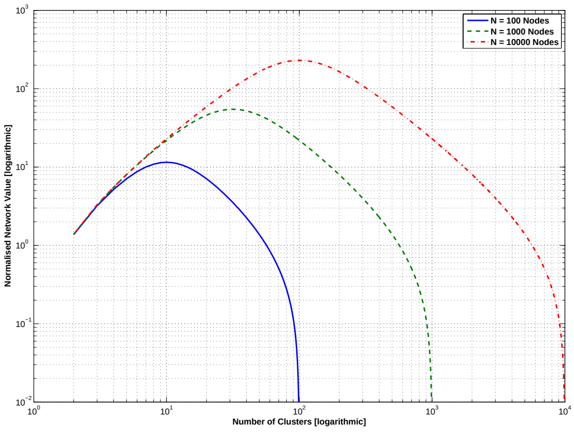

Fig. 3. Relative network value when using clusters.

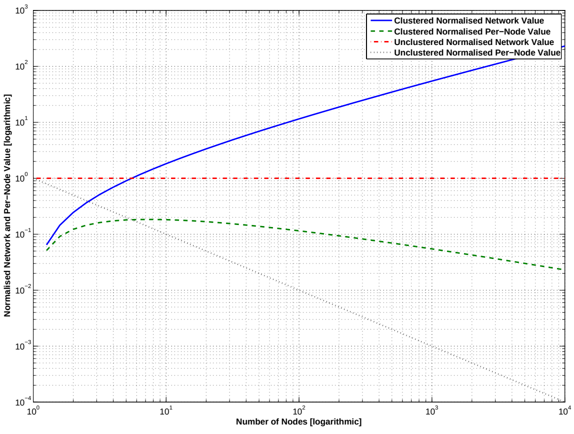

Fig. 4. Relative clustered and unclustered network value.

For an unclustered network, we can calculate the normalised network value V' as

$$V' = \frac{N \log N}{N \log N} = 1. \tag{7}$$

For clustered network, we assume C clusters and hence $M = N/C$ nodes per cluster. Assuming that the value of the nodes within a cluster as well as the cluster heads obeys Zipf's Law, the value per cluster is $M \log M$ and the value of the clustered network is $C \log C \cdot M \log M$. The average number of links needed to maintain all nodes and clusters at any time is $C \log C + M \log M$. We can hence calculate the normalised network value V' as

$$V' = \frac{C \log C \cdot M \log M}{C \log C + M \log M}. \tag{8}$$

The relative network value for different cluster sizes is depicted in Figure 3 with $N = \{100, 1000, 10000\}$ nodes. Clearly, clustering increases the normalized network value. For instance, if we assume a WSN with 10000 nodes, an optimal number of clusters would be 100 with about 100 nodes per cluster. We also observe that a different number of nodes yields a different optimum number of clusters and a different normalised network value. For instance, for $N = 1000$, the optimal cluster number would be about 32, in contrast to 100 for $N = 10000$. The optimum cluster size C can be derived from (8), i.e. $C = \sqrt{N}$, which constitutes the optimum architecture $\hat{\mathfrak{A}}$ from the reduced set of possible clustered architectures.

With reference to (2), the architecture $\mathfrak{A}$ is clustered, the attribute $\mathfrak{F}$ is V' and the problem size S is related to the total number of nodes in the network N (thus certainly not decreasing with an increasing number of nodes).

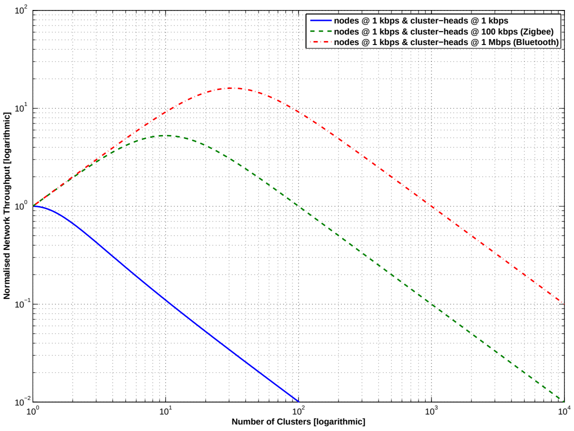

Fig. 5. Normalised network throughput for different data pipes.

The relative efficiency can hence be calculated as

$$\eta = O(\log \sqrt{N} / \sqrt{N}) < O(1), \tag{9}$$

revealing that w.r.t. the normalised network value the architecture is not scalable. However, if compared to an unclustered scheme, which yields an efficiency of $\eta = 1/N$, the clustered approach clearly exhibits a better scalability. This is corroborated by Figure 4.

B. *Clustering with 2-Tier Hierarchy*

Whilst previously we have characterised scalability, we now wish to shed light onto the requirements of the architectures data pipes. Among many possible topologies, we assume that each node communicates only with its respective cluster head and all cluster heads communicate among each other.

We hence assume a 2-tier hierarchy with N total nodes, C clusters and $M = N/C$ nodes per cluster. This 2-tier hierarchy requires two communication phases. In the first phase, all nodes communicate with their respective cluster-heads, and in the second phase, all cluster-heads communicate among each other. For subsequent analysis, we first assume all data pipes to have equal rates and then extend this to unequal pipes.

For equal data pipes, in the first phase, there shall be M time slots to transmit $c \cdot N$ bits, where c is a constant assumed to be one. In the second phase, there are hence $M \cdot C \cdot (C-1)$ time slots to transmit these N bits to every cluster head. The efficiency is hence $N/(M \cdot C \cdot (C-1))$; remember that no new information is injected in the second phase.

For unequal data pipes, let us assume the cluster-heads' pipes to be α times stronger than the data pipes between nodes towards the cluster heads or the cluster-heads performing data aggregation [13] leading to α-times less data to be forwarded. Therefore, in the first phase, there are again M time slots to transmit N bits; and in the second phase there are now $M \cdot C \cdot (C-1)/\alpha$ time slots to transmit these N bits. The efficiency is hence $N/(M \cdot C \cdot (C-1)/\alpha)$.

The normalised network throughput for different cluster sizes is depicted in Figure 5 with $N = 10000$ nodes. Clearly, clustering increases the normalized network throughput only if the data pipes among the cluster heads are stronger or data aggregation is performed to decimate the information shared among cluster heads. For instance, if we assume a WSN with 10000 nodes, then an optimal cluster number is 12 assuming the cluster heads' data pipes to be 1000 times stronger. If not all cluster heads communicated, as in the previous example, then the optimal cluster number would be larger.

Above quantification of the normalised network throughput and value of a large network hence stipulate the use of clustered approaches. This is corroborated by real-world roll-outs, all of which use hierarchical and/or clustered network topologies with stronger data pipes between cluster heads. For example, the currently functioning meter reading application of Coronis uses a hierarchical approach [14] and so does Intel's WSN.

V. CONCLUDING REMARKS

The aim of this paper was to expose some crucial issues related to the scalability and design of large wireless networks. Using some well established scaling laws from communications, i.e. Kumar & Gupta's throughput scaling law, and economics, i.e. Odlyzko & Tilly's value scaling law (which is based on Zipf's law), we have established that large scale networks generally do not scale w.r.t. some key attributes with an increasing number of nodes.

To quantify scalability, we have introduced the notion of architectural efficiency and defined an architecture to be scalable if this efficiency is at least of the order of $O(1)$. This definition has facilitated a comparison between the scalability of unclustered and clustered architectures.

Whilst both unclustered and clustered architecture was shown not to be scalable, a clustered approach – based on some given assumptions – has shown to exhibited a better scalability than its unclustered counterpart. We have also shown that, if clustering is used, the data pipes between the cluster-heads need either to be stronger or data aggregation needs to be performed at the cluster-heads. Such clustered architectures can be built using self-organising and self-healing algorithms, such as introduced in [15].

These results are clearly indicative only, where different assumptions on attributes $\mathfrak{F}$, inclusion of energy consumption, choice of hierarchy, choice of data flows (e.g., directed towards WSN sink), etc., will yield different absolute results. Nonetheless, the results expose tendencies which, so we hope, are of use for large-scale system designers and hence for emerging and future real-world applications, such as data collection, remote control solutions, wireless telemetry, automatic monitoring, metering solutions and smart environments such as homes, hospitals, and buildings of all kinds.

ACKNOWLEDGEMENT

This work has been performed in the framework of the RNRT project ARESA, the funding of which is gratefully acknowledged.

REFERENCES

[1] RNRT ARESA, https://aresa.citi.insa-lyon.fr.

[2] Privat communication with various industrials at *International Conference on WSNs,* November 2006, Paris, France.

[3] Mark D. Hill, "What is scalability?," *ACM SIGARCH Computer Architecture News*, December 1990, Volume 18 Issue 4, pages 18-21.

[4] M. Dohler, D. Barthel, S. Aubert, C. Dugas, F. Maraninchi, M. Laurent, A. Buhrig, F. Paugnat, M. Renaudin, A. Duda, M. Heusse, F. Valois, "The ARESA Project: Facilitating Research, Development and Commercialization of WSNs," *IEEE SECON 2007,* Conference CD-ROM, San Diego, USA, June 2007.

[5] P. Gupta, P. R. Kumar, "The capacity of wireless networks," *IEEE Transactions on Information Theory,* vol. 46, pp. 388-404, 2000.

[6] K.M. Bilby, *The General: David Sarnoff and the Rise of the Communications Industry,* New York: Harper & Row, 1986.

[7] D.P. Reed, "Weapon of math destruction: A simple formula explains why the Internet is wreaking havoc on business models," *Context Magazine,* Spring 1999.

[8] G. Gilder, "Metcalfe's Law and Legacy," *Forbes ASAP,* 1993.

[9] B. Briscoe, A. Odlyzko, B. Tilly, "Metcalfe's Law is Wrong," *IEEE Spectrum Magazine,* July 2006.

[10] G.K. Zipf, "Some determinants of the circulation of information," *Amer. J. Psychology,* vol 59, 1946, pp. 401-421.

[11] N. Vlajic, D. Xia, "Wireless sensor networks: To cluster or not to cluster?," *International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM06),* 2006, pp. 258-268.

[12] W. B. Heinzelman, A. P. Chandrakasan, H. Balakrishnan, "An application-specific protocol architecture for wireless microsensor networks, *IEEE Transactions on Wireless Communications,* vol. 1, no. 4, pp. 660-670, 2002.

[13] H. Chen, H. Mineno, T. Mizuno, "A Meta-Data-Based Data Aggregation Scheme in Clustering Wireless Sensor Networks," *MDM'06,* 2006.

[14] C. Dugas, "Configuring and managing a large-scale monitoring network: solving real world challenges for ultra-lowpowered and long-range wireless mesh networks," *Int. J. Network Mgmt,* 2005, vol 15, pp. 269-282.

[15] J.-L. Lu, F. Valois, D. Barthel, "Low-Energy Self-Organization Scheme for Wireless Ad Hoc Sensor Network," *IEEE WONS,* Conference CD-ROM, January 2007.