Managing your Trees: Insights from a Metropolitan-Scale Low-Power Wireless Network

Zhang Fu, Olaf Landsiedel, Magnus Almgren, Marina Papatriantafilou Email:{zhafu,olaf,magnus.almgren,ptrianta}@chalmers.se

Abstract—Low-power wireless, such as IEEE 802.15.4, is envisioned as one key technology for wireless control and communication. In the context of the Advanced Metering Infrastructure (AMI), it serves as an energy-efficient communication technology for both communications at building-scale networks and city-scale networks. Understanding real-world challenges and key properties of 802.15.4 based networks is an essential requirement for both the research community and practitioners: When deploying and operating low-power wireless networks at metropolitan-scale, a deep knowledge is essential to ensure network availability and performance at production-level quality. Similarly, researchers require realistic network models when developing new algorithms and protocols.

In this paper, we present new and real-world insights from a deployed metropolitan-scale low-power wireless network: It includes 300,000 individual wireless connected meters and covers a city with roughly 600,000 inhabitants. Our findings, for example, help to estimate real-world parameters such as the typical size of routing trees, their balance, and their dynamics over time. Moreover, these insights facilitate the understanding and the realistic calibration of simulation models in key properties such as reliability and throughput.

I. INTRODUCTION

A key vision in Wireless Sensor Network (WSN) research has always been large-scale deployment: tens of thousands of sensor nodes potentially covering large metropolitan areas shall sense and interact with the environment to enable new applications. With the recent trends towards Machine-To-Machine Communication (M2M) and Internet of Things (IoT) we gradually see these visions enfolding [3], [9], [11], [14].

In this paper, we present our insights from a metropolitanscale network of 300,000 wireless nodes, i.e. smart meters, covering a city with roughly 600,000 inhabitants. Meters employ IEEE 802.15.4 [7] and ZigBee [1] for wireless communication and provide the *Advanced Metering Infrastructure* (AMI) of the city. The deployment began in 2008 and the network has been operating at production level since 2009.

To our best knowledge, this paper is the first work to present insights into such a large scale low-power wireless network. We present common metrics for this network to characterize topological properties. In addition, we introduce new metrics, useful to describe the network dynamics. Our findings will (1) help researchers in developing topology models for simulation studies and (2) guide engineers when deploying large-scale networks to achieve high performance and ease network maintenance. In the following, we summarize our contributions.

• **Realistic Topology Characteristics:** We conduct holistic studies including 300,000 wireless nodes and present common topology characteristics such as distances, paths, etc. We

argue that based on our findings, detailed network models of low-power wireless networks can be derived. Such models enable the generation of synthetic network topologies with realistic properties and, as a result, ease protocol evaluation in controlled environments when real network data, especially on a large scale, is unavailable.

- Network Dynamics: Wireless link-dynamics impact the link quality between two devices and potentially lead to connectivity and topology changes. Such link dynamics are especially common in low-power wireless communication, e.g., IEEE 802.15.4, as their low transmission power makes its signals susceptible to interference and noise. We investigate such topological dynamics and study correlations between these dynamics and selected deployment factors, such as *received signal strength indication* (RSSI), and *geographical deployment density*. We believe that the network dynamics characterized in this paper can guide engineers when deploying and maintaining low-power wireless networks and ease their management.
- Routing Trees: In low-power networks, multi-hop routing is a crucial building block: it ensures that packets from a source are forwarded to the sink where processing takes place. In this setting, connectivity and load-balancing are crucial to ensure network reliability and performance. We investigate how packet forwarding tasks are distributed among the nodes in order to identify the hotspots, i.e., the critical points, in the network. Our findings shed light on how to choose network parameters when deploying low-power wireless networks and they provide insights for developing new routing protocols with further improved network reliability, reduced delay and increased energy efficiency.

The remainder of this paper is organized as follows: In Section II, we introduce the studied metropolitan-scale network and our data traces. In Section III, we begin by presenting basic characteristics of the network topology. Next, we conduct a deeper analysis and describe topology dynamics and load balancing in Section IV and Section V, respectively. We discuss related work in Section VI and conclude in Section VII.

II. DEPLOYED LOW-POWER WIRELESS NETWORK

The low-power wireless network discussed in this paper contains around 300,000 wireless nodes. These nodes are smart electricity meters belonging to the Advanced Metering Infrastructure (AMI) [6] of a city. The network covers a metropolitan area with roughly 600,000 inhabitants and about 450 km^2 in size. Its main task is to collect metering data from



Fig. 1: Average heights of all routing trees. Fig. 2: Average path lengths of all routing trees.

smart meters and forward it to the central server for processing and billing. In addition, it handles monitoring of the meters, including error-reporting and over-the-air configuration.

The 300,000 meters communicate with each other via IEEE 802.15.4 and a variant of ZigBee for routing. In addition, 7600 coordinator nodes are deployed throughout the city. Coordinators serve as data sinks and employ a cellular radio, commonly GPRS, next to the 802.15.4 radio. Meters connect to these coordinators either directly or use other meters as relays. Thus, throughout the city there are 7,600 *routing trees*, each with a coordinator node as the root. Nodes can connect and later change to any of these trees, but at any time point a meter can belong to only one tree. Coordinators and electrical meters have built-in power supplies. In addition, batteries are used in the presence of power outages.

The results presented in this paper are based on a data trace of nearly two months (56 days) during the year 2012. For each day, it contains a snapshot of the complete topology of all active nodes.¹ Table I gives the notations used in this paper.

III. BASIC TOPOLOGY CHARACTERISTICS

In this section, we present standard graph-related parameters of routing trees in the studied network, as these are useful for (i) registering these characteristics as parameters that can be used in simulation studies, (ii) getting implications for network performances (such as latencies) that can be affected by those parameters, (iii) using them in the analysis and study of higher level properties and measures, such as dynamics and balancing properties, as proposed and presented in the subsequent sections.

More specifically, we study **a**) *tree heights*, **b**) *path lengths* from nodes to the corresponding coordinators, and **c**) the *node distributions* at different tree levels. Since each of these measures depends on the number of nodes in the tree, we study the measures relative to the sizes of the corresponding routing trees.

A. Tree height and average path length

Tree height is an important topology characteristic indicating the depth of a tree. It can be used in measuring the tree balance [8], through the comparison to the tree size. If it has logarithmic relation to the tree size, then the tree has a good balancing property, which implies good message delivery latencies from leaf nodes to the root. However, tree height can be affected by changes of only a small part of the tree (i.e. adding a child node to a node in the maximum tree level is enough to make the tree height grow by 1), while average path length of a tree can imply the network latencies on average cases, which will not be affected by extreme values. Thus we study both of them: The average height of T_i is computed as $\frac{1}{D_i} \sum_{1 \le d \le D_i}$ (height of t_i^d); the average path length of T_i is computed as $\frac{1}{D_i} \sum_{1 \le d \le D_i} (average path length of <math>t_i^d$). Fig. 1 and Fig. 2 show the average tree heights and average path lengths of all routing trees, respectively. Based on the figures, we found the followings:

- In Fig. 1, we observe significant "step" growth of tree heights taking place at approximately network sizes 15, 50, 100, 200. This is an expected pattern in relatively balanced tree topology, as each tree level has its capacity of containing nodes and, the number of nodes at each level is a multiple of those at the previous level, thus the minimum number of tree levels (i.e. the lower bound of tree height) grows approximately logarithmically with the tree size.
- The variance of tree heights can be large for similar tree sizes. For example, in Fig. 1 when the average tree sizes are around 100, the corresponding average tree heights vary from 3 to 10. This is because the maximum distance between a node and the coordinator is highly depending on the physical locations of the meters, which may have large variations in real-world deployments.
- In Fig. 2, we observe that, in general, the average path lengths grow logarithmically with the tree sizes. We plot two lines: $\log_{10} x$ and $\log_3 x$, noting that most points are bounded by these two lines. This observation implies good performance of message delivery latencies for average cases.

B. Tree-level densities

Next we study the tree-level densities, i.e., the node distribution upon different tree levels, which is an important topology characteristic for generating synthetic topologies and can also give an explanation from this aspect for what we observed

¹Note that due to ongoing network maintenance, some coordinators and their child nodes are omitted for individual days.



Fig. 3: Proportion of nodes in each tree level with different categories of tree sizes



Fig. 4: Tree size dynamic ratios of all routing trees



Fig. 5: Reachability-distances (in km) of all coordinators plotted in log scale

in Fig. 1, and 2. The nodes distribution upon different tree levels were studied for the different categories of tree sizes, following the "step" pattern observed in Fig. 1. The results are shown in Fig. 3. We observe that:

- When there are a small number of nodes in a routing tree, then most of the nodes are connected directly to the coordinator (i.e., reside in l_1 in the tree).
- When the tree size is growing, the broadest tree level moves down. In Fig. 3 we can see that the broadest tree level is l_1 when the tree size is below 15; when the tree size grows, it moves to l_2 ; when tree size reaches 200, it is l_3 .
- The size of l_i does not always increase exponentially when i increases. This explains why the heights of many trees are not logarithmic to the tree sizes. Instead, the size of l_i increases with i increasing until it reaches the broadest level; then it decreases exponentially with i increasing. Therefore, the proportion of nodes with a large distance to the root is quite small, i.e. the trees are diamond-like in shape, thus explaining why the average path lengths are logarithmic to the tree sizes.

IV. ANALYSIS OF TOPOLOGICAL DYNAMICS

The dynamic nature of wireless sensor networks influences network management and communication reliability. In this study, we try to capture such dynamics from the aspects of inter-tree dynamics and intra-tree dynamics, indicating changes of tree participations and changes of parent-child relations among nodes, respectively.

A. Inter-tree dynamics

We propose the following measures to describe the inter-tree dynamics features:

Definition 1: (Tree size dynamic ratio) The tree size dynamic ratio of T_i is defined as the ratio of the standard deviation of its tree size to its average tree size.

Definition 2: (Core-set and Core-set ratio) The core-set of T_i is defined as:

$$Coreset_{T_i} \stackrel{\text{def}}{=} \left\{ m_j : \left| \left\{ d : m_j \in t_i^d \right\} \right| \ge \Phi \right\}$$
(1)

i.e. the set of nodes that have been part of this particular routing tree in at least Φ snapshots. In our study, we choose

 Φ as 50, meaning that if a meter belongs to more than 50 tree snapshots of T_i , then this meter is counted in the core-set of T_i . The core-set ratio of T_i is defined as:

Core-set ratio of
$$T_i \stackrel{\text{def}}{=} \frac{|Coreset_{T_i}|}{\text{average tree size of } T_i}$$
 (2)

The tree size dynamic ratio describes how the nodes are associated with the routing trees and how big the changes of such associations can be. Fig. 4 shows tree size dynamic ratios of all routing trees. We can see that most (95%) of routing trees do not have big size dynamic ratios (less than 0.2), which lead us to the following question:

Q 1: Can we induce that most of the nodes always stay in the same routing trees?

To have deeper insight of inter-tree dynamics, the core-set ratio can help us to see what is the proportion of nodes that are "faithful", i.e. that always stay in the same routing trees. Whether a meter chooses to change its routing tree depends on whether there are other routing trees nearby offering better communication performance. Thus, the density of neighbor coordinators of T_i can affect its core-set ratio. Taking this into consideration, we study the core-set ratio based on different geographical deployment densities of coordinators. We use a well-known density-based clustering algorithm, called OPTICS [2], to find the variance of deployment densities of the coordinators. In the OPTICS algorithm, the standard way to identify clusters is to plot the reachability-distances (RD) of all the points to their neighbor *cluster cores* in an algorithm-specified order, which is called *reachability-plots*. The reachability-distances describe how dense the points are located: smaller reachability-distances indicate denser deployment, while bigger ones indicate sparser deployment. Figure 5 shows the reachability-plots of coordinators in the studied network.

We identified three levels of reachability-distances based on which we conducted the study on core-set ratios of the routing trees. The results are shown in Fig. 6, from where we can see the followings:

• When the tree sizes are small (smaller than 10), the core-set ratios are usually high (above 0.8).



Fig. 6: Probability density function of bivariate distribution of <a verage tree size, core-set ratio> under the three reachability-distances levels of coordinators: (a) reachability-distances are smaller that 100m, (b) reachability-distances $\in [100m, 300m]$, (c) reachability-distances are bigger than 300m

• When the coordinators are deployed more sparsely, the proportion of routing trees who has high core-set ratios increases, e.g., in Fig. 6c, we see that most of the core-set ratios are concentrated between 0.6 and 0.8.

The observations above are what we expected: When the coordinators are deployed sparsely, meters do not find many routing trees nearby, it is more likely that they always stay at the same routing trees. For most of the small routing trees, coordinators and meters are physically located very closely (e.g., in the same room), to get the optimal routing service, the best choice for those meters is to stay in the routing trees of the coordinators co-located with them.

From Fig. 6, we know that it is not true that most of the meters always stay in the same routing trees (answer for question 1). And combined with the observation from Fig. 4, it is interesting to observe that even though the node sets of routing trees vary a lot, the relative sizes among routing trees are quite stable.

B. Intra-tree dynamics

After studying the inter-tree dynamics, we focus on the dynamics within routing trees. In particular, we study the *parent-child* associations among meters and give the following measure:

Definition 3: (**Parent-dynamics**) The parent-dynamics of meter m_i is defined as the number of different parent nodes that m_i has in the data trace.

Important factors which may affect the parent-dynamics are the coordinators deployment density, the received signal strength indication (RSSI) and the tree levels that meters belong to. Fig. 7 shows the probability density functions (PDFs) of parent-dynamics affected by the three factors. Based on that, we found the followings:

- Coordinator-deployment density does not have big influence to the parent-dynamics. It is observed from Fig. 7a that the three PDFs corresponding to different coordinatordeployment densities are quite similar.
- Higher RSSI does not necessarily lead to less parentdynamics. Based on Fig. 7b, it is not possible to conclude

that higher $RSSI^2$ (higher than -30dBm) implies higher probability of small parent-dynamics than lower RSSI, e.g., within (-70dBm, -30dBm).

• Meters which are close to the coordinator do not have less parent-dynamics. From Fig. 7c we observe that for a meter within 3 hops away from the coordinator, the probability that it has more than 15 parents is higher than the corresponding probability for a meter further (e.g. more than 3 hops) away from the coordinator.

C. Implications

Network dynamics may imply difficulties for network maintenance. For example, in the AMI system, communications between the control server and the meters are done through coordinators. If a meter changes its routing tree frequently, then it is difficult for the control server to track it when the server needs to communicate with the meter, which may impair the *two-way communication* ability of AMI [6]. In the meanwhile, within a routing tree, high parent-dynamics may cause *Orphan problems*[12]. For instance, a node may attract many nodes to connect to it due to the good signal strength, which can make it reach the limit of the number of children it can have. So the node cannot accept new connections anymore which may lead some nodes cannot join the network since that node is the only parent node that they can find.

Our findings suggest that sparser deployment of coordinators may ease the work of tracking meters. However, deploying less coordinators can increase the sizes of routing trees, thus affecting the network latencies (observed in Section III). This motivates further investigation on the optimal deployment of coordinators. Parent-dynamics seems to be inherent to the routing trees. Our findings provide insights of its behavior with different deployment factors. To control such dynamics, it can be also helpful to use some static routing strategies or parentchild associations with the combination of dynamic strategies, which was also suggested in [9] for large-scale WSNs.

²In our network, the maximum value of RSSI is 0dBm.



under different coordinator- with different RSSI levels deployment densities

Fig. 7: Parent-dynamics affected by different factors.

V. ANALYSIS OF LOAD-BALANCING PROPERTIES

A. Measure the task load

A node acting as a router in a routing tree carries responsibility for forwarding the messages for the nodes in the subtree rooted at it. The sizes of these subtrees have therefore implications on the robustness of the network, as large subtrees place large task loads to the routers. Therefore, we propose the following measure for analyzing the load-balancing properties of routing trees.

Definition 4: (Average biggest subtree proportion (ABSP)) The ABSP of T_i is defined as

$$ABSP_{T_i} \stackrel{\text{def}}{=} \frac{1}{|\mathcal{D}_i|} \sum_{1 \le d \le \mathcal{D}_i} \frac{\text{the biggest subtree size of } t_i^d}{\text{size of } t_i^d} \quad (3)$$

This measure describes the average biggest task load for routers in a routing tree. It is important, since usually a router in a tree is just an ordinary node who acts in the *full function* mode without having more computation power. Its performance is crucial for the message deliveries, especially when all the nodes in its subtree try to send messages simultaneously. Fig. 8 shows the ABSP with the average tree size for each routing tree. We observed the following:

• "Hot spots" do exist in many routing trees: Although there is a big proportion of routing trees having their ABSPs less than 0.4, there are still around 15% of routing trees with their ABSPs greater than 0.6. This means that in such a routing tree, there exists a node that has to forward messages for more than %60 of the nodes in the tree.

B. Connectivity responsibility

For fine-grained information of load-balancing, we investigate the *connectivity responsibility* of routers, which can be implied by the following question:

Q 2: Is there a small set of routers in l_i that always connect most of the nodes in l_{i+1} , or do all routers in l_i connect similar number of nodes in l_{i+1} ?

To capture such balance-related property, we compute the degree distribution of nodes in different tree levels. From the study in section III, we know that most of the nodes reside in the top 4 tree levels, so most of the routers reside in the top 3 levels. Therefore, we focus on the degree distribution

different hops prop

Fig. 8: Average biggest subtree proportion of all routing trees

of nodes in $l_{1,2,3}$. Since the tree sizes may also affect the nodes distribution of different tree levels which may also affect the degree distribution, we conduct our study for each of the tree size categories³: (16,50], (50,100], (100,200], (201,+ ∞]. Figure 9 shows the results. We found that:

- In all cases, the proportion of nodes have degree x decreases exponentially with the growth of x.
- In each tree level, most of the nodes are leaf nodes (having degree of 1)
- Answer to question 2: most of the nodes in l_{i+1} are connected to routers in l_i that have small degrees (e.g., between 2 and 5).

C. Implications

The energy consumption of a wireless node depends on its subtree size, as the wireless radio chip consumes by far the most power on a sensor node. We showed that nodes can have big subtrees with non-negligible probabilities. If such critical nodes run out of battery, then big part of nodes may be temporally disconnected from the network. This is not acceptable for AMI in smart grid, especially when timecritical messages needed to be delivered to the server. Based on our study, it is observed that the routing trees usually have good balance properties; however, critical nodes with big subtrees and degrees do exist. Therefore, our findings give the insights for controlling the sizes of routing trees and nodes deployment [5]. It also implies the need for rules to choose parent nodes and ensure tree-balancing when a new node is joining into the network. Moreover, besides the basic topological parameters given in Section III, the distributions of ABSP and nodes degrees can be used as parameters of topology models for WSN simulation studies.

VI. RELATED WORK

Large-scale deployments of wireless sensor nodes can be classified into two groups: (1) commercial, production-level deployments, (2) deployments for research purpose.

Commercial deployments often contain tens of thousands of nodes. For example, the City Center Hotel in Las Vegas

 $^{^{3}}$ We do not consider routing trees who have less than 15 nodes, since in such cases, most of the nodes are leaf nodes and are directly connected to the coordinators. Study the node-degree distribution for them is trivial.



Fig. 9: Degree distribution of nodes in different tree levels. Trees are categorized according to their sizes

has 4,200 wireless automated rooms and contains a total of 136,000 wireless ZigBee nodes [10]. The Jackson Memorial Hospital at the University of Florida has 14,000 Zigbee-based tags helping to track assets [4]. While these and many others are large-scale deployments, to our best knowledge, none of them have been analyzed from a research point of view as we do in this paper.

Next to these production-level deployments, there exist a large body of research deployments. While many of their network properties are discussed in scientific publications, these deployments are commonly limited to some thousands of nodes. For example, the Smart Santander deployment in Spanish city of Santander contains about 5000 wireless nodes [14]. Similarly, GreenOrbs [9] contains about 2000 nodes, CitySee about 1200 [11], and ExScal about 1000 [3]. In contrast to these, we present insights into a metropolitan-scale network, containing 300,000 nodes and covering a city of about 600,000 inhabitants. Furthermore, our study focuses on topological properties and dynamics which were not mainly considered in the previous work.

VII. CONCLUSION

In this paper, we presented a detailed study analyzing key properties of a metropolitan-scale low-power wireless network. It contains 300,000 wireless nodes (meters) and comprises several thousands of routing trees to which the wireless nodes connect. In particular, with our study we can characterize (i) the shape of the trees, (ii) proportions of the nodes that "migrate" among trees relative to the deployment density (inter-tree dynamics) and proportions of nodes that change parents within the same tree (intra-tree dynamics), and (iii) contention and balancing properties through distributions of node degrees and biggest subtree sizes. Analyzing those key properties and building corresponding models can ease network operation and management for practitioners and simulation projects for researchers. These new models are required as low-power networks differ in key aspects from other largescale deployments such as the Internet itself [15] or cellular networks [13].

ACKNOWLEDGMENT

This work has been partially supported by the European Commission Seventh Framework Programme (FP7/2007-2013) through the SysSec Project, under grant agreement 257007, through the FP7-SEC-285477-CRISALIS project, through the Swedish Energy Agency (Energimyndigheten), through the E-science research center of Chalmers and through the collaboration framework of Chalmers Energy Area of Advance.

REFERENCES

- Zigbee Alliance. Zigbee specification, http://www.zigbee.org/ Specifications.aspx, 2013.
- [2] Mihael Ankerst, Markus M. Breunig, Hans-Peter Kriegel, and Jörg Sander. Optics: ordering points to identify the clustering structure. In Proceedings of the 1999 ACM SIGMOD international conference on Management of data, SIGMOD '99, 1999.
- [3] A. Arora, R. Ramnath, E. Ertin, P. Sinha, S. Bapat, V. Naik, et al. Exscal: elements of an extreme scale wireless sensor network. In *Embedded* and Real-Time Computing Systems and Applications, 2005. Proceedings. 11th IEEE International Conference on, 2005.
- [4] Awarepoint Corporation. Awarepoint expands its real-time location solution (rtls) at jackson health system. Press Release, 2009.
- [5] Zhao Cheng, M. Perillo, and W.B. Heinzelman. General network lifetime and cost models for evaluating sensor network deployment strategies. *Mobile Computing, IEEE Transactions on*, 7(4):484–497, 2008.
- [6] H. Farhangi. The path of the smart grid. Power and Energy Magazine, IEEE, 8(1):18–28, 2010.
- [7] J.A. Gutierrez, M. Naeve, E. Callaway, M. Bourgeois, V. Mitter, and B. Heile. Ieee 802.15.4: a developing standard for low-power low-cost wireless personal area networks. *Network, IEEE*, 15(5):12–19, 2001.
- [8] P. L. Karlton, S. H. Fuller, R. E. Scroggs, and E. B. Kaehler. Performance of height-balanced trees. *Commun. ACM*, 19(1):23–28, January 1976.
- [9] Yunhao Liu, Yuan He, Mo Li, Jiliang Wang, Kebin Liu, Lufeng Mo, et al. Does wireless sensor network scale? a measurement study on greenorbs. In *INFOCOM*, 2011 Proceedings IEEE, 2011.
- [10] Ryan Maley. Zigbee standards and adoption. In Johnson Controls Eastern ABCS Conference, 2012.
- [11] Xufei Mao, Xin Miao, Yuan He, Xiang-Yang Li, and Yunhao Liu. Citysee: Urban co2 monitoring with sensors. In INFOCOM, 2012 Proceedings IEEE, 2012.
- [12] Meng-Shiuan Pan, Chia-Hung Tsai, and Yu-Chee Tseng. The orphan problem in zigbee wireless networks. *Mobile Computing, IEEE Transactions on*, 8(11):1573–1584, 2009.
- [13] U. Paul, A.P. Subramanian, M.M. Buddhikot, and S.R. Das. Understanding traffic dynamics in cellular data networks. In *INFOCOM*, 2011 *Proceedings IEEE*, pages 882–890, 2011.
- [14] Luis Sanchez, Jose Antonio Galache, Veronica Gutierrez, Jose Manuel Hernandez, Jesus Bernat, Alex Gluhak, et al. Smartsantander: The meeting point between future internet research and experimentation and the smart cities. In *Future Network & Mobile Summit (FutureNetw)*, 2011.
- [15] Georgos Siganos, Michalis Faloutsos, Petros Faloutsos, and Christos Faloutsos. Power laws and the as-level internet topology. *IEEE/ACM Trans. Netw.*, 11(4):514–524, aug 2003.