Deployment Challenges and Developments in Wireless Sensor Networks Clustering

Krishna Doddapaneni, Fredrick A. Omondi
School of Science & Technology
Middlesex University
London, UK
Email:{k.doddapaneni, f.adero}@mdx.ac.uk

Enver Ever
Middle East Technical University
Northern Cyprus Campus
Guzelyurt, Mersin 10, Turkey
Email:{eever@metu.edu.tr}

Purav Shah, Orhan Gemikonakli
School of Science & Technology
Middlesex University
London, UK
Email:{p.shah, o.gemikonakli}@mdx.ac.uk

Roberto Gagliardi
School of Science and Technology
University of Camerino
Camerino, 62032, Italy
Email:{roberto.gagliardi@unicam.it}

***Abstract*—Clustering techniques for wireless sensor networks (WSNs) have been extensively studied and they have proven to improve the network lifetime, a primary metric, used for performance evaluation of sensor networks. Although introduction of clustering techniques has the potential to reduce energy consumption and extend the lifetime of the network by decreasing the contention through either power control or node scheduling, scalability remains an issue. Therefore, the optimality of the cluster size still needs to be thoroughly investigated. In this paper, a single cluster head (CH) queuing model is presented. Using an event based simulation tool (Castalia), key issues that affect the practical deployment of clustering techniques in wireless sensor networks are analysed. These include identifying the bottlenecks in terms of cluster scalability and predicting the nature of data packets arrival distribution at the CH. Results presented show that this analysis can be used to specify the size of a cluster, when a specific flow of data is expected from the sensing nodes based on a particular application and also the distribution of the inter-arrival times of data packets at the CH follows exponential distribution.**

***Index Terms*—Wireless Sensor Networks, modelling, performance, cluster size, scalability, clustering, distribution**

I. Introduction

Wireless Sensor Networks (WSNs), with a wide range of applications are rapidly becoming an integral part of our lives. Recently, considerable amounts of research efforts have enabled the actual implementation and deployment of sensor networks tailored to the unique requirements of certain sensing and monitoring applications. The application of sensor networks are diverse, ranging from habitat monitoring to surveillance and physical intrusion detection and can be categorised into environment, health, military, home, disaster relief, space exploration and other commercial areas. The flexibility, fault tolerance, low cost, rapid deployment characteristics and high sensing fidelity of sensor networks create many new and exciting applications in the field of remote sensing. Energy efficiency is crucial because of the scale and application environments in which sensors are deployed [1]. WSN applications and communication protocols are tailored mainly to provide higher energy efficiency, as sensor nodes carry limited power sources. To provide extensive area coverage, a large number of nodes are required. Moreover, to provide a centralised management system of nodes, clustering algorithms are provided as an effective means to extend lifetime and manage WSN's [2]. Significant attention has been paid to clustering strategies and algorithms, yielding a large number of clustering protocols.

Performance modelling and evaluation should consider metrics, such as channel bandwidth and arrival distribution of data packets at the CH, and the introduction of new traffic attributes (such as variations in packet size, rate). Minimising the energy consumption and extending the lifetime of the network is possible with the introduction of clustering techniques as they decrease the contention through either power control or node scheduling, however, scalability is still an issue. Hence the optimality of the cluster size still needs to be thoroughly investigated. The objectives of this paper are twofold. The first is to identify the bottlenecks in the network, in terms of cluster size scalability, especially while addressing variety of high packet sending rate and real-time applications, such as wearable heart rate and physical activity monitors and holter monitors. This directly effects the volume of traffic, the CH can handle. The second objective is to predict the distribution of data packet arrivals at the CH, for performance modelling and evaluation and which can be incorporated in energy optimisation studies. In particular, a cluster network is implemented and analysed, using Castalia simulation tool, implemented on OMNET++.

The rest of the paper is organised as follows. Section II presents the background and related work, focussing on IEEE 802.15.4 standard, CC2420 radio and clustering. The description of the system and the assumptions considered are presented in section III. Section IV discusses the results obtained using simulation, along with analysis of the results. Finally, section V concludes the present work and provides the future directions.

II. Background and related work

In recent research focused on energy efficient WSN protocols, Clustering is considered as an approach for enabling communication between sensors in a field and the base station (BS) [3]. Group of sensor nodes create clusters with one sensor node amongst them acting as a CH. These CH's are responsible for communicating information from sensor nodes in their clusters to the BS, perhaps through other CH's. The data flow in a cluster network is presented in Figure 1. While the proposed approaches in the research literature are not very specific as to the functionality of the CH's (e.g. they are not like WPAN coordinators), in order to distribute the load across sensors in the clusters, cluster heads are periodically changed by an election process [4]. This improves the number of nodes in the network that are still alive after a given time, thus overall improving the network lifetime. In studies where clustering techniques were primarily proposed for energy efficiency purposes, e.g. LEACH, HEED, EEUC, [5], [6], [7], [8], [9], [10], [11], [12], [13], the network lifetime was significantly prolonged. Clustering is mainly considered to improve scalability and energy efficiency [7]. Besides achieving energy efficiency, clustering reduces channel contention and packet collisions, resulting in better network throughput under high load [14]. In the wireless domain, a high density of nodes has advantages in terms of connectivity and coverage as well as disadvantages in terms of increased collision and overhead for protocols that require neighbourhood information. As a result, optimising the size of the cluster is an issue in WSN protocols as the nodes increases. This is also related to how much traffic the CH will receive dependent upon it's cluster size. Although the protocols developed aim to decrease the contention through either power control or node scheduling, cluster size still is an issue.

Overall, clustering protocols have the following advantages in WSNs [15], [2]:

- Load balancing: Clustering schemes help to prolong the networks lifetime, by reducing energy usage in intra-cluster as well as in inter-cluster communication. It is intuitive to balance the load among them, so as to improve the performance goals [16].
- Network Longevity: In a cluster, the CHs are active most of the time, while other nodes wake up only in a specified interval to perform data transmission to the CH. Furthermore, by dynamically changing the CH functionalities among nodes, the energy consumption of the network can be significantly reduced. For the applications of WSNs in harsh conditions, networks lifetime is a major concern. It is imperative to minimize the energy for intra-cluster communication, when CHs are richer in resources than sensors [7]. Adaptive clustering is a possible solution for achieving network longevity [5], [6]. It is also ideal to place cluster heads closer to other sensors in its cluster [17], [18].
- Scalability: Cluster-based protocols limit the number of transmissions between nodes, thereby enabling a higher number of nodes to be deployed in the network.
- Collision reduction: Since most of the functionalities of nodes are carried out by the CHs, fewer nodes contend for channel access, improving the efficiency of channel access protocols.
- Local information: Intra-cluster information exchange between the CH and the nodes helps summarize the local network state and sensed information of the phenomenon state at the CH.
- Routing backbone: Cluster-based approaches enable efficient building of the routing backbone in the network, providing reliable paths from sensor nodes to the sink. Since the information to the sink is initiated only from CHs, route-thru traffic in the network is decreased.

The heterogeneity in the available sensor platforms result in compatibility issues for the realization of envisioned applications. Hence, the need for a standardisation for certain aspects of communication. The IEEE 802.15.4/Zigbee standard was formed to provide low data-rate wireless transceiver technology with low complexity and longer battery life. In most recent platforms such as CC1000, which is used in Mica2 motes [1], the transmit-receive energy is almost equal. This trade-off has recently tilted toward receive energy consumption in the CC2420 and is therefore used in many platforms and complies with the IEEE 802.15.4 standard. Receiver electronics dominate amplifier energy consumption due to the increased complexity with spread-spectrum techniques, making receive energy consumption higher than that for transmission. The transmission range of the nodes is assumed to be 10-100 m with data rates of 20 to 250 kbps, depending on the purpose of their deployment. While the physical layer uses binary phase shift keying (BPSK) in the 868/915MHz bands, offset quadrature phase shift keying (O-QPSK) is used in the 2.4GHz band, Medium Access Control layer provides communication for star, mesh, and cluster tree-based topologies with controllers. The symbol rate is 62.5 kS/sec (with 4 bits/S, and bit rate 250 kb/sec). Each symbol is converted into a 32-chip sequence. Consequently, the chip rate is 2 Mc/sec [19], [20].

Sensor nodes have resource constraints in terms of limited energy, limited communication and computational capabilities, and limited memory. A sensor node may belong in one of four categories (1) Intel research developed a high bandwidth sensing device such as iMote, which has a much broader bandwidth than the earlier ones (Bluetooth-based radio) as well as a larger memory; (2) a specialized sensing platform (Spec), smaller in size and memory, and has a narrow communication bandwidth and a short radio distance; (3) a generic sensing platform such as the Bekeley mote, designed using off-the-shelf components and has a bandwidth of 100 kbps or so and more memory, compared to Spec; and (4) a gateway-

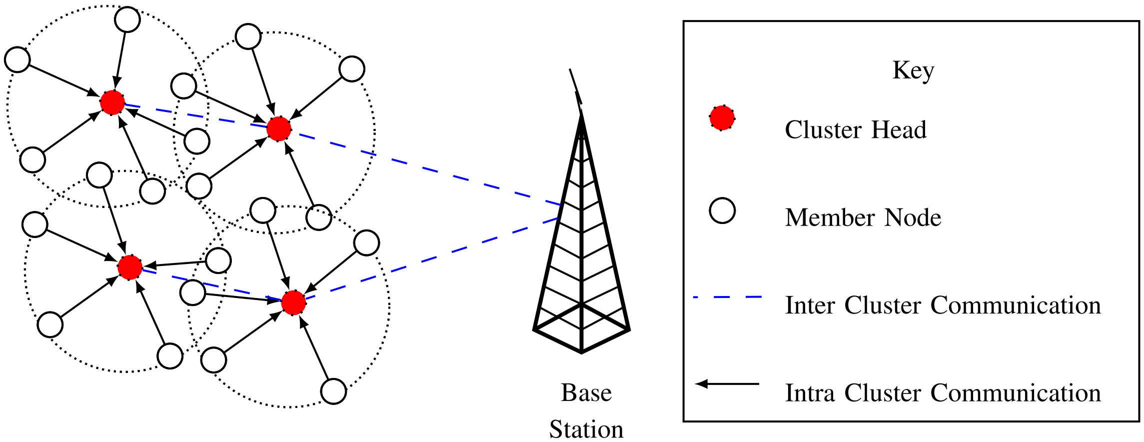

Fig. 1. Data flow in a clustered network

like sensor node such as Stargate, which is a gateway to directly connect mote (or iMote)-based devices. These sensor nodes have different levels of resources within them, but they all contain at least the following physical units: a radio unit (transceiver), a processing unit (micro-controller and a memory), a sensing unit, and a limited power supply unit [21].

III. System description and Assumptions

In this section, the system description along with the assumptions considered, as well as the queue model of the system is presented. A cluster network with one CH coordinating the cluster operations is considered. Sensed information at the nodes is forwarded to the CH which finalises cluster aggregation and transmit all the information to the sink either directly or through intermediary CHs. We assume that all the sensor nodes are connected directly to the CH, hence they communicate via their CH. It is assumed that at least one path always exists towards the sink [22].

The resulting job arrivals (packets sent to the base station) at the CH is the collection of jobs from random nodes, (λ_1, λ_2... λ_N) in the cluster. The behaviour of each CH is considered as an open queue network using M/M/1 queues [23]. A single CH's behaviour can be analysed with the queuing model shown in Figure 2. The internal arrivals to the CH ($k = 0$) are arrivals within the cluster (from nodes 0,1,2...N). The mean arrival rate (λ_k) at the CH can therefore be given by $\lambda_k = \sum_{n=0}^{N} \lambda_n$, where n = 0,1,....N. It is assumed that the resulting superposition of all the job arrivals at node k follows Poisson distribution with mean arrival rate λ_k packets/sec. The analysis performed in the next section shows that the assumption holds. The external arrivals are from the other CHs, forwarding their data to the sink, through node k. Once the jobs are processed at node k, they are transmitted directly or forwarded to the sink through an intermediary node r (node r represents the next CH towards the sink). Node 0 ($k = 0$) represents the CH. At node r the process is similar to that at node k. The external arrival to the CH (from other CHs) are arrivals with rate λ_r. The other arrivals are originated from the sensor nodes forwarding their data to CH. The rate of traffic from node r within the cluster to node k is $\lambda q_{(r,k)}$. The operation is assumed to be similar at all other CHs. Packets are handled on first come first served (FCFS) basis. In the following section, the validity of the model is analysed using Castalia simulator and results are presented.

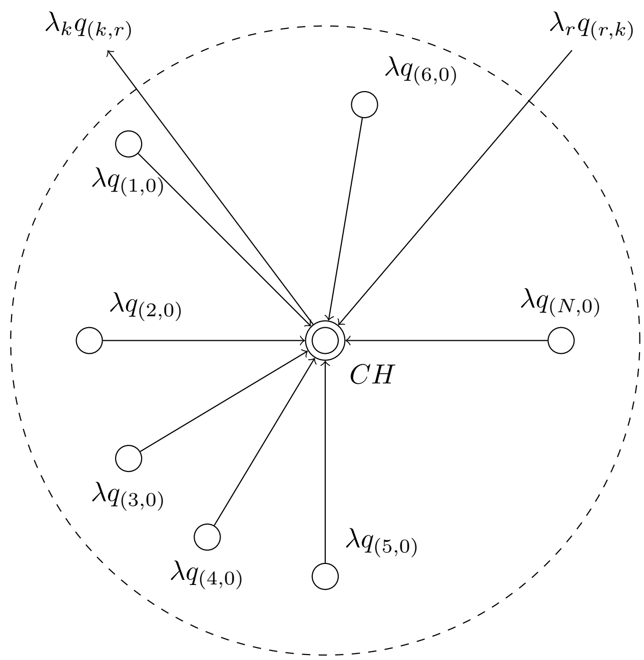

Fig. 2. Queuing model of a single CH

IV. Numerical results and discussions

In this section, numerical results are presented for the model considered. Castalia is a WSN simulator based on the OMNeT++ platform. It is mainly used for initial testing of protocols and/or algorithms with a realistic node behaviour, wireless channel and radio models. The OMNeT++ platform is an extensible, modular, component-based C++ simulation library and framework, primarily for building network simulators. Castalia is highly tunable, features an accurate radio model based on the work of the authors in [24]. It also features physical process model, considering clock drift, sensor energy consumption, CPU energy consumption, sensor bias etc. Unpredictability of the wireless channel, energy spent in transmission/receiving packets, performance degradation experienced by duty cycles, collisions are usually overlooked by simple simulators. However these details are well established in Castalia [25].

The following parameters are used throughout this section, unless otherwise stated. A CC2420 chip, compatible with 802.15.4, is used to provide wireless communication, operating at 2.4 GHz and providing a data rate of 250 kbps. The packet size is considered to be 105 bytes [26]. TMAC, which provides both collision avoidance and reliable transmission is used as the MAC protocol. Each simulation lasts for 100 sec to reach steady state and the results are taken over an average of 100 runs.

A. Case A

In order to analyse the effects of packet rate and number of nodes on the cluster, simulation studies are performed by varying the number of nodes and packet rates in each run. The number of nodes in the cluster are varied from 10 to 40. All the nodes are sending the sensed information to the CH at a constant rate which is varied from 1 packet/sec to 30 packets/sec. The results presented are in good agreement with the two different analytical solution approaches for performability modelling of a WSN cluster, presented by the authors in [23]. Figure 3 shows the average number of packets received by the CH, from each node as a function of varying nodes and packet rates. The results presented show that higher packet rates can be accommodated by fewer nodes to attain saturation (maximum utilization reached), as compared to the network with low data rates and higher number of nodes. In other words, to scale out the network for wider coverage a trade-off between number of nodes is appropriately considered against the aggregate packets received at the CH.

The saturation points for possible number of cluster nodes for packet rates chosen (Figure 4), are in agreement with the work presented by the authors in [23]. The number of nodes has been varied from 5 to 60, and the packet rates are varied from 1 packet/sec to 7 packets/sec. At low packet rates (say 1 packet/sec, 2 packets/sec), saturation levels are reached much later, with more number of nodes in the system. At higher packet rates, saturation levels are reached earlier with less number of nodes. Hence, this analysis can be used to specify the size of a cluster, when specific flow (data rates) is expected from different types of applications. In other words, it is a key issue that affects the practical deployment of clustering techniques in sensor network applications.

B. Case B

In order to predict the distribution of packet inter-arrival times (time between two consecutive arrivals) at the CH, a

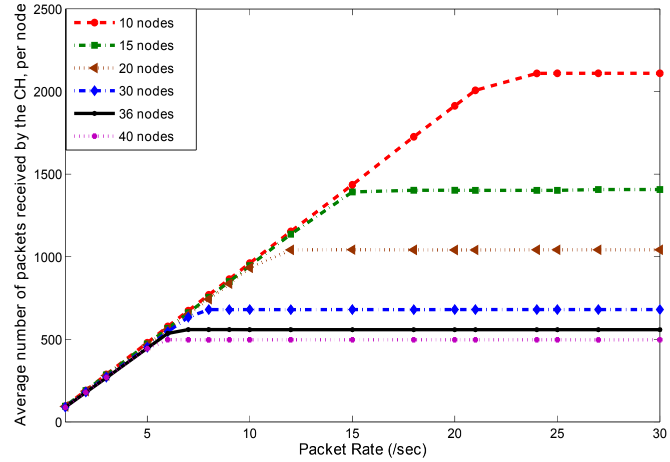

Fig. 3. Comparison of packet arrivals at CH vs Packet rate of cluster nodes

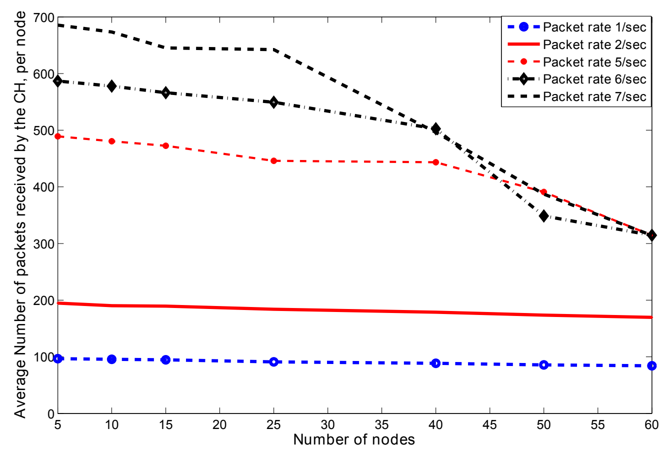

Fig. 4. Comparison of packet arrivals at CH vs number of cluster nodes

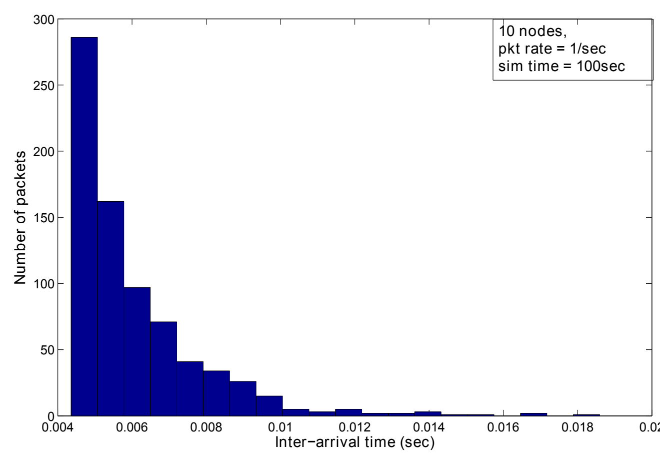

Fig. 5. Distribution of data packet inter arrivals at CH (10 nodes, 1pkt/sec)

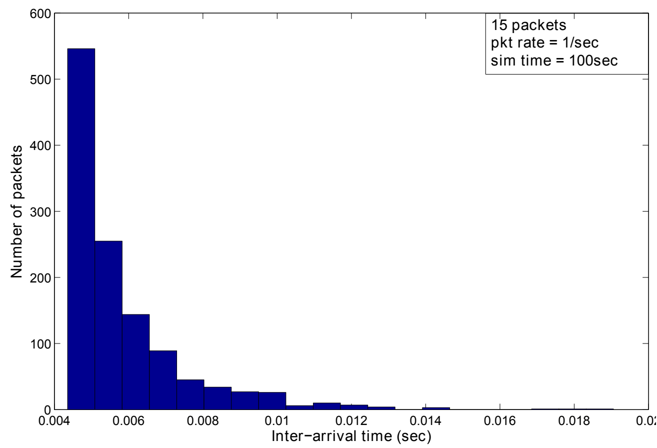

Fig. 6. Distribution of data packet inter arrivals at CH (15 nodes, 1pkt/sec)

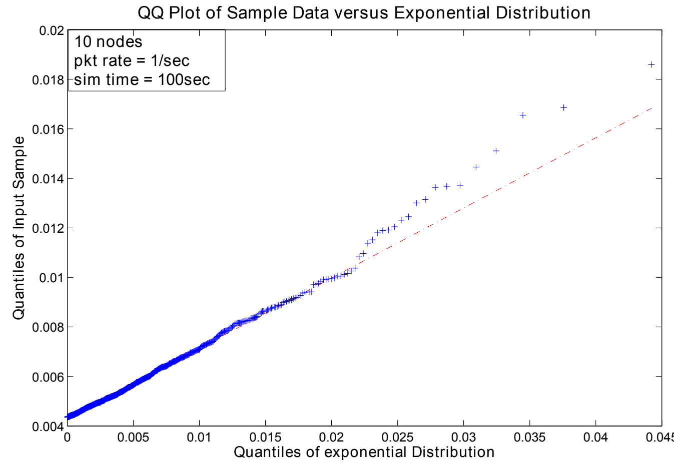

Fig. 8. QQ plot for fitting into an exponential distribution

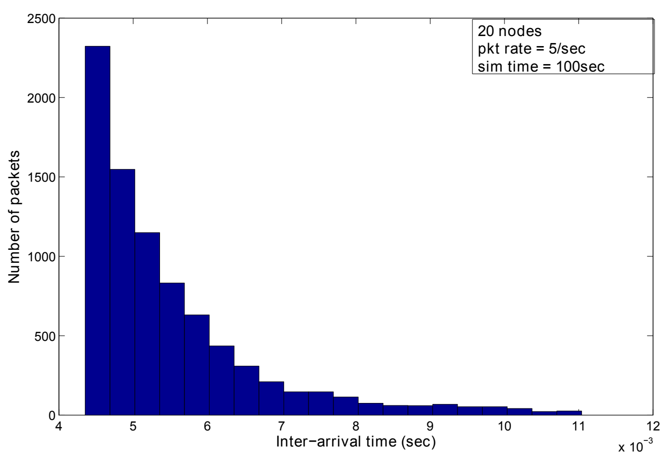

Fig. 7. Distribution of data packet inter arrivals at CH (20 nodes, 5pkt/sec)

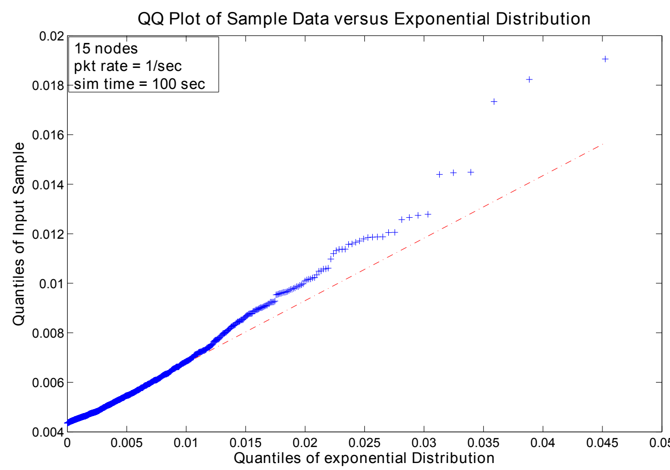

Fig. 9. QQ plot for fitting into an exponential distribution

cluster network of various sizes (from 10 nodes to 35 nodes) are considered. All the nodes in the cluster send data packets to the CH at a constant packet rate. The resulting job arrivals at the CH is a collection of jobs from the cluster nodes, the sensed information from the CH itself and the data forwarded from other CHs. The same has been performed for various packet sizes and simulation times. The mean service rate of the CH is assumed to be exponentially distributed with rate μ_k. Figures 5, 6 and 7 present the distribution of data packet inter-arrival time at the CH. The incoming packet inter-arrival times follow an exponential distribution which can be well approximated. In addition, regardless of the data packet rate or the simulation time, the distribution of inter-arrival times at the CH is exponential. Quantile-quantile plots compared in Figures 8, 9 and 10 follow the exponential distribution phenomenon, as the plot follows the $x = \beta\ y$, with skinny positive tail. The linearity of the points suggests that the inter-arrival times are exponentially distributed. The aim of all investigations in queueing theory is to get the main performance measures of the system which are the probabilistic properties of the distribution of inter-arrival times, service times, number of servers, capacity and service discipline. Therefore, following our findings from Figures 8, 9 and 10, the first parameter of Kendall's notation for the M/M/1 queuing system considered in this study is confirmed as Markovian.

V. Conclusions and Recommendations

Performance modelling and evaluation should consider metrics, such as channel bandwidth and arrival distribution of data packets at the CH, and the introduction of new traffic attributes. In this paper, key issues affecting the practical deployment of clustering techniques are thoroughly investigated. Though clustering techniques extend the lifetime of the network, scalability is still an issue, hence the optimality of the cluster size still needs to be thoroughly investigated. In order to have a wider coverage, a trade-off exists between the number

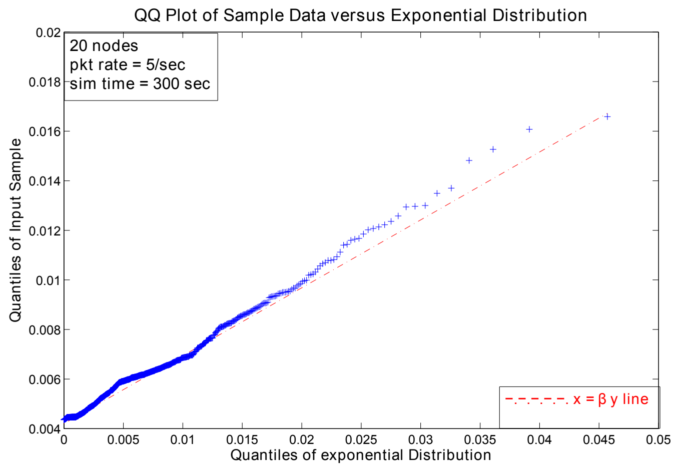

Fig. 10. QQ plot for fitting into an exponential distribution

of nodes in a cluster being considered and the aggregate packets received at the CH. The results presented show that this analysis can be used to specify the size of a cluster when a specified flow of data is expected from the sensing nodes. In other words, higher packet rates can be accommodated by fewer nodes to attain saturation, as compared to the network with low data rates and higher number of nodes. This directly affects the volume of traffic, the CH can handle. The results presented in order to predict the distribution of inter-arrival times at the CH follows the exponential distribution phenomenon, thus confirming the Markovian nature of the first parameter in Kendall's notation.

References

[1] I. Akyildiz and M. Vuran, *Wireless Sensor Networks*, ser. Advanced Texts in Communications and Networking. Wiley, 2010.

[2] O. Younis and S. Fahmy, "Distributed clustering in ad-hoc sensor networks: A hybrid, energy-efficient approach," in *Proc. of the 23rd IEEE Intl Conf. on Computer Communications (INFOCOM)*, 2004.

[3] O. Younis, M. Krunz, and S. Ramasubramanian, "Node clustering in wireless sensor networks: recent developments and deployment challenges," *Network, IEEE*, vol. 20, no. 3, pp. 20–25, 2006.

[4] A. A. Abbasi and M. F. Younis, "A survey on clustering algorithms for wireless sensor networks." *Computer Communications*, no. 14-15, pp. 2826–2841.

[5] T. Moscibroda and R. Wattenhofer, "Maximizing the lifetime of dominating sets," in *19th IEEE International Parallel and Distributed Processing Symposium, 2005. Proceedings.*

[6] R. Khanna, H. Liu, and H. H. Chen, "Self-organization of sensor networks using genetic algorithms," in *Communications, 2006. ICC '06. IEEE International Conference on*, vol. 8, 2006, pp. 3377–3382.

[7] M. Younis, M. Youssef, and K. Arisha, "Energy-aware management for cluster-based sensor networks," *Computer Networks*, pp. 649–668, 2003.

[8] W. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "Energy-efficient communication protocol for wireless microsensor networks," in *Proceedings of the 33rd Annual Hawaii International Conference on System Sciences, 2000*, p. 10 vol.2.

[9] S. Lindsey and C. Raghavendra, "Pegasis: Power-efficient gathering in sensor information systems," in *Aerospace Conference Proceedings, 2002. IEEE*, pp. 3–1125–3–1130.

[10] A. Manjeshwar and D. Agrawal, "Teen: a routing protocol for enhanced efficiency in wireless sensor networks," in *15th Internationa Parallel and Distributed Processing Symposium Proceedings l*, 2001, pp. 2009–2015.

[11] A. M. et.al., "Apteen : a hybrid protocol for efficient routing and comprehensive information retrieval in wireless," in *Parallel and Distributed Processing International Symposium Proceedings , IPDPS 2002*.

[12] R. S. Chang and C. J. Kuo, "An energy efficient routing mechanism for wireless sensor networks," in *20th International Conference onAdvanced Information Networking and Applications, 2006. AINA*, vol. 2.

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

[14] O. Younis, M. Krunz, and S. Ramasubramanian, "Node clustering in wireless sensor networks: Recent developments and deployment challenges," *IEEE Network Magazine*, vol. 20, pp. 20–25, 2006.

[15] O. Younis and S. Fahmy, "Heed: A hybrid, energy-efficient, distributed clustering approach for ad hoc sensor networks," *IEEE Transactions on Mobile Computing*, vol. 3, no. 4, pp. 366–379, 2004.

[16] M. Younis, K. Akkaya, and A. Kunjithapatham, "Optimization of task allocation in a cluster-based sensor network," in *Proceedings of the Eighth IEEE International Symposium on Computers and Communications*, ser. ISCC '03. Washington, DC, USA: IEEE Computer Society, 2003, pp. 329–336.

[17] E. Oyman and C. Ersoy, "Multiple sink network design problem in large scale wireless sensor networks," in *2004 IEEE International Conference on Communications*, vol. 6, 2004, pp. 3663–3667.

[18] Y. Hou, Y. Shi, H. Sherali, and S. Midkiff, "On energy provisioning and relay node placement for wireless sensor networks," *Wireless Communications, IEEE Transactions on*, vol. 4, no. 5, pp. 2579–2590, 2005.

[19] K. Pahlavan and P. Krishnamurthy, *Networking Fundamentals*. Chichester, UK: John Wiley and Sons, 2009.

[20] "IEEE Standard for Information Technology- Telecommunications and Information Exchange Between Systems- Local and Metropolitan Area Networks- Specific Requirements Part 15.4: Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications for Low-Rate Wireless Personal Area Networks (WPANs)," Tech. Rep., 2006. [Online]. Available: http://dx.doi.org/10.1109/ieeestd.2006.232110

[21] K. SOHRABY, *Wireless Sensor Networks Technology Protocols and Applications*. Wiley, 2007.

[22] C. Chiasserini and M. Garetto, "Modeling the performance of wireless sensor networks," in *INFOCOM 2004. Twenty-third AnnualJoint Conference of the IEEE Computer and Communications Societies*, vol. 1, march 2004, pp. 4 vol. (xxxv+2866).

[23] F. Omondi, E. Ever, P. Shah, and O. Gemikonakli, "Modelling wireless sensor networks for performability evaluation," in *Ad-hoc, Mobile, and Wireless Network*, ser. Lecture Notes in Computer Science, J. Cicho, M. Gbala, and M. Klonowski, Eds. Springer Berlin Heidelberg, 2013, vol. 7960.

[24] M. Zuniga and B. Krishnamachari, "Analyzing the transitional region in low power wireless links," in *First Annual IEEE Communications Society Conference on Sensor and Ad Hoc Communications and Networks, 2004. IEEE SECON 2004.*, pp. 517 – 526.

[25] C. Newport, D. Kotz, Y. Yuan, R. S. Gray, J. Liu, and C. Elliott, "Experimental evaluation of wireless simulation assumptions." ACM Press, 2004, pp. 78–82.

[26] B. Latr, P. D. Mil, I. Moerman, N. V. Dierdonck, B. Dhoedt, and P. Demeester, "Maximum throughput and minimum delay in ieee 802.15.4." in *MSN*, ser. Lecture Notes in Computer Science, vol. 3794. Springer, 2005, pp. 866–876.