An Empirical Study of Data Collection Protocols for Wireless Sensor Networks

Stephen Rothery
Autonomous Systems Laboratory
CSIRO ICT Centre, Australia
stevekr@gmail.com

Wen Hu
Autonomous Systems Laboratory
CSIRO ICT Centre, Australia
wen.hu@csiro.au

Peter Corke
Autonomous Systems Laboratory
CSIRO ICT Centre, Australia
peter.corke@csiro.au

Abstract

In the past few years, numerous data collection protocols have been developed for wireless sensor networks (WSNs). However, there has been no comparison of their relative performance in realistic environments. Here we report the results of an empirical study using a Fleck3 sensor network testbed for four different data collection protocols: One phase pull Directed Diffusion (DD), Expected Number of Transmissions (ETX), ETX with explicit acknowledgment (ETX-eAck), and ETX with implicit acknowledgment (ETX-iAck).

Our empirical study provides useful insights for future sensor network deployments. When the required application end-to-end reliability is not strict (e.g., 70%) and link quality is good, DD and ETX are the best options because of their simplicity and low routing overhead. Both ETX-eAck and ETX-iAck achieve more than 90% end-to-end reliability when the link quality is reasonable (less than 25% packet loss). When the link quality is good, ETX-iAck introduces significantly less routing overhead (up to 50%) than ETX-eAck. However, if the radio transceiver supports variable packet length, ETX-eAck can outperform ETX-iAck when the link quality is poor.

The important message from this paper is that *choice* of data collection protocol should come *after* the operating environment is *understood*. This understanding must include the characteristics of the radio transceiver, and link loss statistics from a long-term (across seasons and weather variation) radio survey of the site.

Categories and Subject Descriptors

C.2.2 [**Computer Communications Networks**]: Network Protocols

General Terms

Experimentation, Performance, Routing

Keywords

Sensor Networks, Data Collection

1 Introduction

In the past few years, there has been intensive research activities in the area of Wireless Sensor Network (WSN), which comprise numerous tiny, simple devices combining sensing, computing, and communication capabilities. Since a WSN comprises a large number of nodes cooperating with each other, an effective communication paradigm is of prime importance and this has been investigated by many, for example [7, 15, 16, 10, 6, 9, 12, 11].

In this paper, we study the performance of data collection protocols [7, 6, 15, 9], which many current WSN deployments (such as [1, 14, 5, 8, 4]) focus on. The design of each collection protocol is based on different assumptions about the scenario such as network density, link quality, and radio transceiver parameters. Although each protocol claims to solve some of the challenges in WSN data collection, little is known about the relative performance of these protocols, in particular in *realistic environments*.

To the best of our knowledge, this work is the *first* comprehensive empirical study of WSN data collection protocols. We do not attempt to find the best possible operational scenario for each protocol. Instead we focus on the direct comparison of four popular data collection protocols, which include Directed Diffusion (DD) [7], Expected number of transmissions (ETX), ETX with explicit acknowledgment (ETX-eAck) [15], and ETX with implicit acknowledgment (ETX-iAck) [9], under a set of scenarios. The contributions of this paper include:

- Useful insights into the different protocols' performance, such as routing overhead, end-to-end delivery rates, and node energy consumption.
- A key observations that the relative performance of a protocol varies significantly in different scenarios. It is therefore very important to perform a site survey to collect these scenario parameters before choosing a suitable data collection protocol for a new WSN deployments. When a site survey is expensive, it is desirable to design an adaptive protocol, which can dynamically choose the protocol based on real time scenario parameters.

The rest of this paper is organized as follows. We give a brief overview of the protocols evaluated in this work in

RealWSN'08, April 1, 2008, Glasgow, Scotland.

Section 2. We discuss protocol evaluation environment and metrics in Section 3. Section 4 discusses the results of performance comparisons. Finally, Section 5, concludes and discusses areas for future work.

2 Data collection protocols overview

In this section, we provide a brief overview of the WSN data collection protocols investigated in this paper.

2.1 Directed diffusion – One phase pull

Directed Diffusion (DD) [7] is a data-centric, reverse-path based communication paradigm for sensor networks. Sink nodes flood their interests into the network when they join the network. An interest is a query specifying the attributes of the information a sink wants a node to collect and respond to. Source nodes in turn flood the first few exploratory data packets, called gradients, into the network. Sinks select and reinforce the best paths and the sources use reverse best paths to deliver data back to the sinks. The sink chooses the neighbor from which the first discovery-packet is received as the immediate upstream node. Therefore, it can minimize the per-node network state maintenance and achieve highly-efficient data dissemination in sensor networks.

The initial packet-flooding is not efficient for dense and large scale networks and the *one-phase-pull* [6] algorithm was introduced to reduce flooding. The sources choose the best paths by judging the arrival of the interest packets.

2.2 Expected number of transmissions – ETX

Instead of using hop count as the routing metric, ETX makes use of link quality information gained from the Media Access Control (MAC) layer. For a path comprising h hops, ETX can be calculated as:

$$ETX = \sum_{i=1}^{h} \frac{1}{p_i} \tag{1}$$

where p_i is the link quality of hop i. The cost information for this summation is distributed down through the network in the routing beacons. Furthermore, in order to prevent frequent changes to the routing of the network, any new parent node needed to have a cost which was 20% better than the current cost.

2.3 ETX with explicit acknowledgment

ETX-eAck [15] or *Surge Reliable* extends the functionality of ETX with link layer retransmissions. A receiver sends an eAck after receiving a packet from a sender. Assuming a packet is lost, the sender will retransmit the packet if it has not received an eAck after a preset Retransmission TimeOut (RTO). For a path comprising h hops, the ETX for Surge Reliable can be calculated as:

$$ETX = \sum_{i=1}^{h} \frac{1}{p_i q_i} \tag{2}$$

where p_i is the upstream link quality, and q_i is the downstream link quality of hop i. This routing metric takes the transmission cost of both data and eAck packets into account.

2.4 ETX with implicit acknowledgment

Similar to ETX-eAck, ETX-iAck [9] also uses Equation (2) to calculate routing metrics. The key differences between ETX-iAck and ETX-eAck are how duplicate packets are handled and how long a node waits for its acknowledgment. ETX-iAck uses the following equation to calculate RTO for a node i:

$$RTO_i = RTO_{i-1} \frac{1}{q_{i-1}} \tag{3}$$

where q_{i-1} is the downstream link quality of node i's parent node (node $i-1$). So the ETX-iAck RTO for a given node is calculated by taking the RTO of the parent node, and multiplying that by the expected number of transmissions that it will require to successfully deliver the packet to the parent nodes' parent. The RTO is capped at 1 second in order to prevent excessively long RTOs. Due to the recursive nature of this calculation, the required link quality information is passed down through the network in the routing beacons.

Another consideration was the case where the child node may not successfully overhear the parent node forwarding its packet, causing it to continue to resend, waiting for an iAck that will never come due to the parent having already forwarded it. Therefore, in the case that a parent node receives a packet from its child that it has already forwarded, it will mark the packet as a duplicate and resend it. This will satisfy the child node for the iAck, and other nodes will detect the duplicate and drop the packet.

3 Metrics and Methodology

In this section, we discuss the evaluation methodology and metrics for our empirical study of WSN data collection protocols.

3.1 Methodology

In order to evaluate the performance of these protocols, we set up a test bed of 51 Fleck-3 nodes [13]. The Fleck-3 node has an 8-bit ATmega128L micro-controller, 4K-byte RAM, and a packet-based Nordic NRF905 transceiver[2]. This transceiver is quite different to the commonly used (such as Texas Instruments CC1000) devices in that it provides short (32 byte) constant length packets with automatically generated preamble and checksum.

The nodes were arranged in a 7×7 grid, with an additional node and the sink placed outside the grid, in the middle of one side. All of the nodes were spaced 5cm apart and were adjusted so that their effective transmit range was about 15cm. In order to test the routing protocols' performance with differing link quality, uniform random packet loss with a known mean was induced on each of the sensor nodes. This is in addition to the normal vagaries of node-to-node wireless communications.

For tests where the network was reduced in size, the network was not moved but the relevant nodes were deactivated in software and set to listen only. In the case of the 35 node grid, two rows of nodes on the edge furthest from the sink were turned off and also the individual node outside of the grid. In the case of the 20 node grid, an additional row on the far side was deactivated and two rows from the top of the grid were also deactivated.

We implemented all four protocols using the Fleck Operating System (FOS) [3] — a lightweight threaded environment. For all of these tests, the data rate for each node was 1 packet per 30 seconds, with routing beacons being sent every minute. In addition to this, the maximum number of resends was set to 7 for the protocols that make use of some form of acknowledgment, e.g., ETX-iAck and ETX-eAck. The tests themselves were set up to run for 40 minutes each. Each of these tests was run twice and the average of the results from the two runs was computed. To collect experiment statistics, each node transmitted its current transmission counts, routing status, link quality information and a sequence number in the data payload of each packet sent.

3.2 Metrics

We have chosen the following three metrics to evaluate the performance of WSN data collection protocols.

- *End-to-end reliability:* this metric represents the ratio between the number of unique data packets sent by the sources and the number of unique data packets received by the sink. Ideally, the end-to-end delivery rate should be equal to 1.
- *Total Energy Consumption:* since WSNs are expected to operate unattended for a long period of time after deployment using limited battery energy or low available environmental power, energy consumption is one of the most important metrics in WSN data collection protocols. In this paper, we consider the energy consumption by radio transmission (sending and receiving), and use the number of transmissions to represent the radio transceiver's energy consumption. The number of transmissions should be as small as possible.
- *Routing Overhead:*This metric represents the total number of routing packets (e.g., beacons, Acks and data retransmissions) transmitted during the experiment. This in turn indicates the scalability of a protocol, and how much effort for a protocol to achieve the end-to-end delivery rates. Ideally, routing overhead should be as small as possible.

4 Empirical Results

In this section, we evaluate the performance of different WSN data collection protocols in a variety of environment (e.g., different network sizes and different link quality).

4.1 End-to-end reliability

Figure 1 shows the end-to-end delivery rates for all protocols with different link quality in a 50-node network. ETX-eAck and ETX-iAck show very similar behaviour, as do DD and ETX. When the link quality is good (no induced packet losses), all protocol achieve good (more than 75%) end-to-end delivery rates. ETX with either eAck or iAck achieved more than 3-fold end-to-end delivery rates when three are medium (25%, 50%) induced packet losses. When the link quality is very bad, e.g., with 75% induced packet loss, the delivery rates of all protocols is less than 15%. Although the delivery rates of data packets can be improved by link layer retransmissions, the routing beacon messages are broadcast and hence have no link layer retransmission. Few end-to-end routing route can be set-up successfully when the link quality is very bad; therefore, the delivery rates of all protocols is poor.

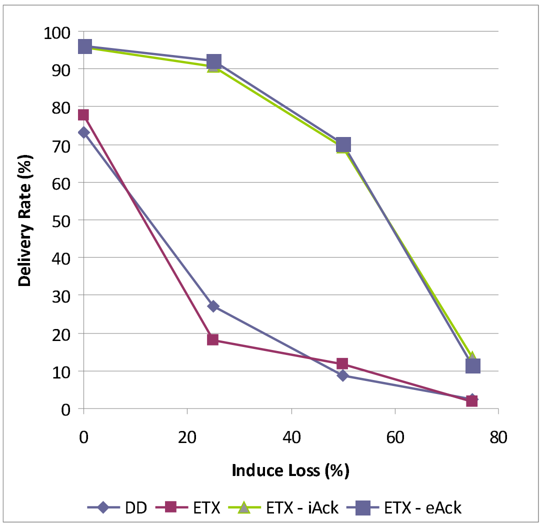

Figure 1. Reliability (50 nodes).

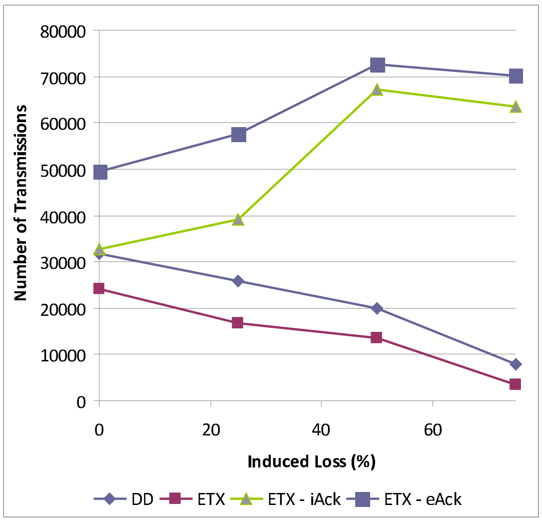

Figure 2. The total number of transmissions (50 nodes).

4.2 Total Energy Consumption

Figure 2 shows the total number of transmissions in a 50-node network, which include both data packets and routing overhead. When the link quality is reasonable (induced link losses are less than 25%), ETX-eAck transmits over 60% packets more than ETX-iAck while achieving similar end-to-end reliability (see Figure 1). However, when the link quality is poor (induced link losses are more than 50%), the benefits of ETX-iAck are reduced because excessive link dynamics cause significant number of data packet retransmissions. DD and ETX transmit fewer packets than ETX-iAck and ETX-eAck, but achieve lower end-to-end deliver rates (see Figure 1).

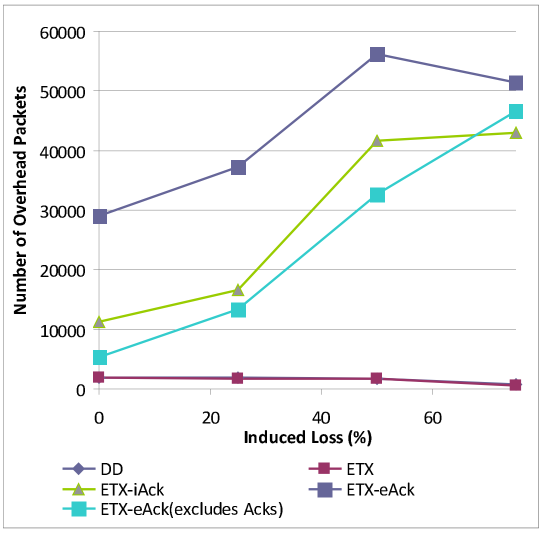

Figure 3. Routing overhead (50 nodes). Note that the lines of DD and ETX are overlapped in the figure.

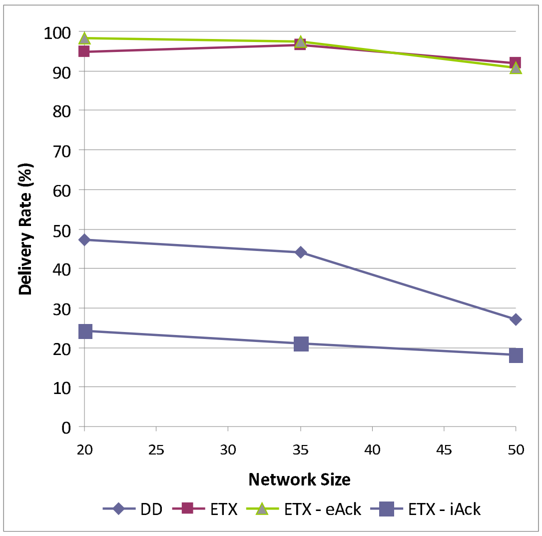

Figure 4. Reliability versus network size. Induced packet loss = 25%.

4.3 Routing Overhead

Figure 3 shows the routing overhead of all protocols in a 50-node network. DD and ETX generate very little routing overhead compared to ETX-iAck and ETX-eAck. When link quality is reasonable (induced link loss is 25% or less), ETX-iAck generates significantly (about 50%) less routing overhead than ETX-eAck, but it generates more data retransmissions (see ETX-eAck excludes Acks in Figure 3). When induced link loss is 25%, almost 70% of routing overhead for ETX-eAck is Acks.

For a fixed-length packet-based radio transceiver such as the Fleck-3's Nordic NRF905, there is no difference between Ack packets and normal data packets in term of transmission energy consumption. However, for a radio transceiver that can support variable packet lengths, ETX-eAck may transmit fewer bytes (and is therefore is more energy efficient) than ETX-iAck because the length of data packets is typically significantly larger than the length of Ack packets. For example, when induced link loss is 50%, ETX-eAck, compared to ETX-iAck, transfers about 1,000 less data packets, but 1,500 more Acks. If the length of a data packet is 1.5 times more than that of an Ack packet, ETX-eAck generates less routing overhead in bytes than ETX-iAck.

4.4 Scalability

Figure 4 shows the end-to-end delivery rates achieved by all protocols with different network sizes when induced link loss is 25%. Both ETX-iAck and ETX-eAck scale well with network size (achieving more than 90% end-to-end delivery rates). DD performs significantly better than ETX when network sizes are small.

Figure 5 shows the total number of transmissions for all protocols with different network sizes with induced link loss of 25%. As network size increases, the number of transmissions increases linearly in all protocols. When network size is 50, a significant number of packets are dropped in the forwarding processes of ETX, which results in very low delivery rate (less than 20%) and a small number of packet transmissions (almost the same as the number of transmissions when network size is 35).

We have also run the same experiment with different induced link loss rates (e.g., 0, 25%, 50%, and 70%), and have observed similar results.

4.5 Discussions and Deployment Suggestions

A key observation in our study is that the relative performance of a data collection protocol varies significantly with the operating scenario. When link quality is good (with no induced loss), DD and ETX provide good end-to-end reliability (70%) with minimum overhead. Without introducing extra protocol complexity, end-to-end reliability can be increased by over-sampling (increasing the sampling and transmission rates).

When link quality is poor, link layer retransmission can be used to improve end-to-end reliability but with heavy routing overhead cost. ETX-iAck outperforms ETX-eAck in most scenarios by embedding Acks into data packets. However, if the radio transceiver supports variable packet length, ETX-eAck can outperform ETX-iAck because the length of data packets is typically significantly larger than the length of Ack packets.

When link quality is very poor (less than 10%), no protocol can provide reasonable end-to-end reliability because routes are difficult to established and maintain. Because routing beacon messages are sent out by broadcast, MAC layer retransmission can no longer improve the reliable-delivery of beacon messages. The WSN designer should consider adding relay nodes, increasing antenna height or gain, or making use of better radio transceivers to improve link quality.

5 Conclusion and future work

In this paper, we reported the performance of four WSN data collection protocols in a variety of realistic environment: Directed Diffusion (DD), Expected number of transmissions (ETX), ETX with explicit acknowledgment (ETX-

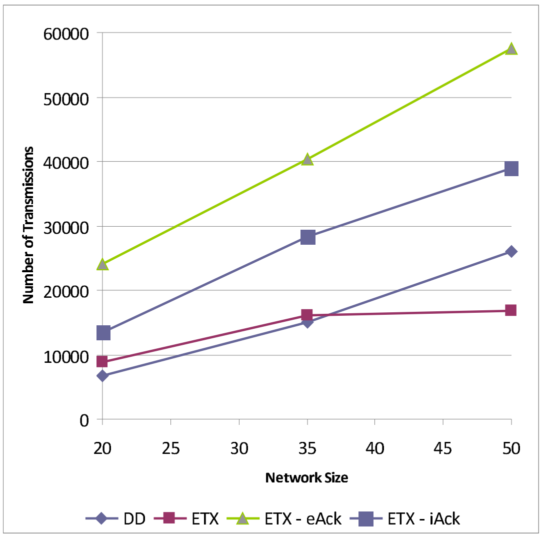

Figure 5. The total number of transmissions versus network size. Induced packet loss = 25%.

eAck), and ETX with implicit acknowledgment (RTX-iAck). Our comprehensive empirical study provides useful insights for future sensor network deployments. When the application's required end-to-end reliability is loose (e.g., 70%) and link quality is good, DD and ETX are the best options because of their simplicity and low routing overhead. Both ETX-eAck and ETX-iAck achieve more than 90% end-to-end reliability when the link quality is reasonable (less than 25% induced packet losses). When the link quality is good, ETX-iAck introduces significantly less routing overhead (up to 50%) than ETX-eAck. However, if the radio transceiver supports variable packet length, ETX-eAck can outperform ETX-iAck when the link quality is poor.

The important message from this paper is that *choice* of data collection protocol should come *after* the operating environment is *understood*. This understanding must include the characteristics of the radio transceiver, and link loss statistics from a long-term (across seasons and weather variation) radio survey of the site. When site survey is expensive, it is desirable to design an adaptive protocol, which can dynamically choose different data collection protocol based on real time scenario parameters.

Since the data collection protocol represents the energy consumption of one communication layer only, we plan to investigate the energy consumption of other WSN communication layers (e.g., Media Access Control) in the future.

6 Acknowledgments

The authors would like to thank Tuan Le Dinh for his help in the implementation of ETX-iACK protocol.

7 References

[1] Habitat monitoring on great duck island. http://www.greatduckisland.net/index.php.

[2] Nordic nrf905 radio transceiver. `http://www.nordicsemi.com/index.cfm?obj=product\&act=display\&pro=83\#`.

[3] P. Corke. Demo abstract: FOS — a new operating system for sensor networks. In *Fifth European conference on wireless sensor networks (EWSN 2008)*, Bologna, Italy, Jan, 2008.

[4] T. L. Dinh, W. Hu, P. Sikka, P. Corke, L. Overs, and S. Brosnan. Design and deployment of a remote robust sensor network: Experiences from an outdoor water quality monitoring network. In *Second IEEE Workshop on Practical Issues in Building Sensor Network Applications (SenseApp 2007)*, Dublin, Ireland, Oct. 2007.

[5] C. Hartung, R. Han, C. Seielstad, and S. Holbrook. Firewxnet: a multi-tiered portable wireless system for monitoring weather conditions in wildland fire environments. In *MobiSys 2006: Proceedings of the 4th international conference on Mobile systems, applications and services*, pages 28–41, New York, NY, USA, 2006. ACM Press.

[6] J. Heidemann, F. Silva, and D. Estrin. Matching data dissemination algorithms to application requirements. In *Proceedings of the First ACM Conference on Embedded Networked Sensor Systems (SenSys).*, 2003.

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

[8] K. Langendoen, A. Baggio, and O. Visser. Murphy loves potatoes: Experiences from a pilot sensor network deployment in precision agriculture. In *14th Int. Workshop on Parallel and Distributed Real-Time Systems (WPDRTS)*, Rhodes, Greece, 2006.

[9] T. Le, W. Hu, P. Corke, and S. Jha. ERTP: Energy-efficient and reliable transport protocol for data streaming in wireless sensor networks. *UNSW Computer Science and Engineering Technical Report no. UNSW-CSE-TR-0802*.

[10] J. Liu, F. Zhao, and D. Petrovic. Information-directed routing in ad hoc sensor networks. In *Proceedings of the 2nd ACM international conference on Wireless sensor networks and applications*, pages 88–97, San Diego, CA, USA, 2003. ACM Press.

[11] S. Madden, M. J. Franklin, J. M. Hellerstein, and W. Hong. Tag: a tiny aggregation service for ad-hoc sensor networks. *ACM SIGOPS Operating Systems Review*, 36(SI):131–146, 2002.

[12] A. Meliou, D. Chu, C. Guestrin, J. Hellerstein, and W. Hong. Data gathering tours in sensor networks. In *The Fifth International Conference on Information Processing in Sensor Networks*, pages 43–50, 2006.

[13] P. Sikka, P. Corke, and L. Overs. Wireless sensor devices for animal tracking and control. In *First IEEE Workshop on Embedded Networked Sensors in 29th Conference on Local Computer Networks*, pages 446–454, 2004.

[14] G. Tolle, J. Polstre, R. Szewczyk, D. Culler, N. Turner, K. Tu, S. Burgess, T. Dawson, P. Buonadonna, D. Gay, and W. Hong. A macroscope in the redwoods. In *Proceedings of the Third International Conference on Embedded Networked Sensor Systems (SenSys05)*, pages 51–63, 2005.

[15] A. Woo, T. Tong, and D. Culler. Taming the underlying challenges of reliable multihop routing in sensor networks. In *SenSys '03: Proceedings of the 1st international conference on Embedded networked sensor systems*, pages 14–27. ACM Press, 2003.

[16] F. Ye, H. Luo, J. Cheng, S. Lu, and L. Zhang. A two-tier data dissemination model for large-scale wireless sensor networks. In *Proceedings of the 8th annual international conference on Mobile computing and networking*, pages 148–159, Atlanta, Georgia, USA, 2002. ACM Press.