Systematic Experimental Analysis and Evaluation of Routing Protocol in Wireless Sensor Networks

Tiong Hoo Lim[1], **Iain Bate**[1], **Jon Timmis**[1,2]

[1] Department of Computer Science,
[2]Department of Electronics,
University of York, York
North Yorkshire, YO10 5DG
United Kingdom
Email: thlim, iain.bate, jtimmis@cs.york.ac.uk

Keywords: Wireless Sensor Networks, Statistical Analysis, Statistical Confidence, Scientific Confidence, Significant Tests, Routing Protocol

Abstract

The use of multihop routing protocols in wireless sensor network (WSNs) is common, as sensor nodes are usually deployed across a geographical area with minimal transmission range to conserve energy. However, existing routing protocols cannot tolerate multiple failures, for example, NST-AODV performs well in sporadic and temporal link failure but cannot cope with longer periodic failure. In this paper, we present a novel Multi-modal Routing Protocol (MRP) which address these issues, and validate its performance against two existing reactive routing protocols in both simulated and testbed environments. We propose the use of a systematic experimental analysis and evaluation technique to produce results with an acceptable level of confidence and improve the credibility of our experiments. The main contributions of this paper are (1) to evaluate whether the previous simulated results are valid in a real hardware, (2) the application of statistical evaluation techniques to ensure the results are both scientifically and statistically significant. Both experimental and simulation approaches have produced credible performance improvement in term packet delivery for MRP and yielded a lower routing control overhead than AODV and NST-AODV.

1 Introduction

With the advancement of micro-chip technology and cheaply available sensor nodes, Wireless sensor networks (WSNs) have been used with success in application scenarios such as remote patient health monitoring, fire search and rescue operation, structural monitoring of engineering structures, and military surveillance. For many of these applications, the dependability is an important factor. A number of sensor motes can be used to monitor and collect information from the environment and send the information to a central location over a geographical area. Each mote can autonomously communicate and interact with each other over the wireless medium via multihop routing protocol, to ensure that critical information are routed and received by the user.

However, real world deployments of WSNs are usually hard to control and difficult to deploy. It is not always practical to test a new algorithm design in a live network or real world deployment as the algorithm usually needs to be reconfigured and finetuned. The motes once deployed are usually difficult to access and reconfigure. In order to address this, WSNs research community sometimes relies on simulation tools to test and evaluate their new algorithm or protocol as it allows significant levels of testing to be performed at reasonable cost. Unfortunately, many current WSN simulations are developed with simplifying assumption about the underlying simulation models, network protocols, wireless communication and environment that do not provide the same result and behaviour in real work deployment. An alternative approach to validate and evaluate WSNs is the use of real hardware testbed experiments in a controlled environment. However, it is difficult to test a system sufficiently to have confidence that it will work in practice and pilot studies in the laboratory are not the same as the *real* environment. This has lead to many reports of failures when WSNs are deployed for real [15, 16].

In Lim et al. [6][1], a Multi-modal Routing Protocol (MRP) was proposed that showed improvements in terms of performance and reliability of message delivery over Adhoc On-Demand Distance Vector (AODV) [9] and Not So Tiny-AODV (NST-AODV) [2]. The principal weaknesses of this work are that the evaluation is very limited in terms of the number of tests performed, the fact the tests were only performed in simulation makes it unclear how well the approach would work on real hardware, and only limited analysis of the results was found. In this paper an experimental framework is proposed that is useful for evaluating MRP and other routing protocols. The framework is based on performing extensive testing in simulation and more limited testing on the real hardware. The larger amounts of testing in simulation allow us to show, using state of the art statistical techniques, whether there are improvements with greater confidence than the real world testing. The real world testing allows us to confirm the trends of simulation and understand the degree of similarity (validity) between

---

[1] http://rtslab.wikispaces.com/file/view/mrp.tar

the two. The main objective of this paper is to provide a systematic approach to improve the credibility of an experimental study. To the best of our knowledge, this is the first time that a comprehensive analysis and evaluation approach has been proposed to evaluate the credibility of the study of WSNs.

The main contributions of this paper are:

- Formulation of a systematic experimental approach to improve the reliability and validity of a WSNs experiment.
- Quantitative evaluation of an experiment to verify the performance of the Multimodal Routing Protocol (MRP) [6] is significantly better than Adhoc Ondemand Distance Vector (AODV) [9] and Not So Tiny-AODV (NST-AODV) [2].

The rest of this paper is organised as follows: We discuss the limitation of existing works in Section 2 followed by the introduction and technical details of MRP in Section 3, that is used in this paper to test our experimental framework. We describe our experimental framework in Section 4. The results are analysed and discussed in section 5, before we present our conclusion in section 6.

2 Related Work

Limited work has been done in WSNs to validate the credibility of the results obtained from hardware or software experiments using statistical hypothesis techniques. Work by Ivanov et al. [4] has validated the performance of a link-state ad-hoc routing protocol using the results of 16 real wireless ad-hoc nodes experiment with the results of the ns-2 wireless simulator. The results have shown that the simulated packet delivery ratio is very close (error rate of 1%) to the real emulated results, but the latency results show much deviation from the real experiment due to delay introduced by the hardware and operating systems. However, it is difficult to see the significance of the results due to the lack of statistical tests applied on the results.

In Pham et al. [10], the Castalia WSN simulator is used to evaluate the performance of tunable MAC protocol and validate with the results collected from 9 TelosB motes deployed in a 3 by 3 grid. The validation process is performed by calculating the average of packet reception rate (PRR) for all the links using 50 random runs obtained from simulator and compared against the PRR obtained from real hardware run. The study has found that, even with the complex radio model available in the Castalia simulator, the results obtained are still significantly difference. However, the degree of differences is not validated using statistical techniques and the number of run in hardware experiment is small (one run).

In Lim et al. [6][2], the MRP has been proposed to operate in different routing protocols during network void. Individual node in the network can make its own routing decision to switch between routing modes autonomously with minimal network disruption. Significant performance improvements in terms of reliability and efficiency of message delivery over AODV and NST-AODV. The weaknesses of this work are that the evaluation is only performed in simulation makes it unclear how well the approach would work on real hardware, and limited statistical analysis of the results was found.

3 Multi-Modal Routing Protocol

The use of multi-modal routing in WSNs has been proposed in various papers over the last 10 years due to its autonomous characteristic and the ability of each protocol to deal with different situations, e.g [3, 5]. It has been applied in WSNs to cluster the nodes according to the complexity of the tasks and capabilities of the node and allow them to operate in various modes [1, 7]. Haas and Pearlman [3] applied proactive routing to operate within a cluster of nodes and reactive routing between clusters but unable to tolerate topological changes. Kim et al. [5] used a rule based to select between geographical diffusion and AODV depending on the packet priority. This rule based approach uses a centralised decision based the current network statistics and threshold level. However, it is subject to single point of failure.

The MRP is proposed in Lim et al. [6] to tolerate failure caused by external interference where an individual node will make its own routing decision to switch between routing modes autonomously with minimal network disruption. It is a distributed approach based purely on reactive routing protocols, integrating existing features from AODV and NST-AODV. It consists of a Route Selection Module (RSM), a set of routing protocols and conditional rules, and threshold as shown in Figure 1. The routing decision made by each node is independent from other nodes. During initialisation process, each of routing features is associated with a cost value based on the energy required to execute them. Retransmission (RT) from NST-AODV will have the lowest cost as it required only single communication, while global discovery will have the highest cost due to network-wide Route Discovery (RD) process, the propagation and multiplication of Route Request (RREQ), and packet dropped.

In MRP, the Link Layer Notification (LLN) is enabled by default. When a packet is received by a node, it will be buffered and fed into the RSM. The RSM module will check the routing table for the available route to forward the packet. If a route is not available, RD will be initiated. If a route is available, it will transmit the packet. When a packet cannot be sent, a failure notification will be reported by the link layer. Using the LLN, the RSM will update and adjust routing cost accordingly in real time as shown in the flow diagram in Figure 2. For each failure, the cost of the routing feature executed will be increased by one until it reaches a maximum value. As the network returns to normal, these costs will decrease exponentially over a time period. The cost will also be decreased for every 5 successful transmissions for RT. The node will also select the routing recovery

---

[2] downloadable at http://rtslab.wikispaces.com/file/view/mrp.tar

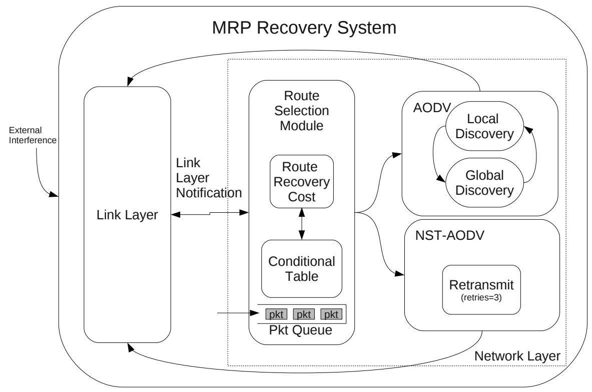


Figure 1: The Architecture of Multimodal Routing Protocol [6]

feature with the lowest cost during transmission failure to transmit the failed packet. If the cost of current feature is higher than the next routing feature, or above a threshold value, where the next routing feature will be selected. This approach will allow a more effective and efficient routing strategy to be executed by individual nodes and different routing protocols maybe used for different packets by different nodes depending on its cost at that time instance.

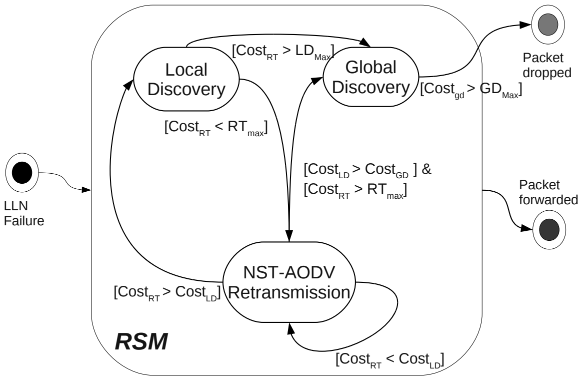


Figure 2: Flow Diagram for MRP during LLN

4 Experimental Framework and Setup

In this section, we discuss the objectives of this paper and present the experimental work to demonstrate that MRP has a better network reliability with lower latency and greater energy efficiency by formalising a set of hypotheses.

Hypothesis 1 (H_1): MRP is able to deliver more packets than NST-AODV and AODV when failures occur for different durations.

Hypothesis 2 (H_2): There are no significant differences in the latency between MRP and AODV, and MRP and NST-AODV.

Hypothesis 3 (H_3): The number of packet generated to determine the route to a destination is significantly less in MRP than AODV and NST-AODV.

Hypothesis 4 (H_4): There are no differences in the results obtained from the simulation and the real hardware implementation.

4.1 Representation of Results

In any scientific experiment, before any statistical test can be applied to test its significance, the raw data result collected has to be pre-processed and analysed to better understand the nature of the data, and determine the central tendency of the results. The mean value is widely used to measure the best central tendency as all the values are accounted and any variation will affect the calculated mean value. However, if the data are skewed or have outlier, the mean may be pulled toward the skewed data or outlier, and loses its ability to represent the central tendency. In this situation, taking the median will be a better statistics.

In our experiment, the median is used as we do not need to make any assumptions of the underlying distribution and is less affected by outliers and skewed data. The median is computed by arranging the data in the order of magnitude and is represented by the midpoint of the data set. We also determine the 1st quartile (Q1) and 3rd quartile (Q3) of the data sets by identifying the first quarter of the data set as Q1 and third quarter of the data set as Q3. The quartiles show how the data are distributed on either side of the median and the difference between Q3 and Q1 is known as the interquartile range (IQR). As a general rule of thumb, any results outside 1.5 times the IQR can be identified as outliers.

4.2 Conceptual Statistical Test Framework

Once we have obtained meaningful representation of the results, statistical tests can be run to verify its significance. We have presented a conceptual statistical test framework in Figure 3 to analyse and verify the improvement observed in MRP are significant and have scientific values. The components of this framework are discussed in the following subsection where subsections 1, 2, 3 correspond to the labels in the Figure, i.e. (1, 2, 3).

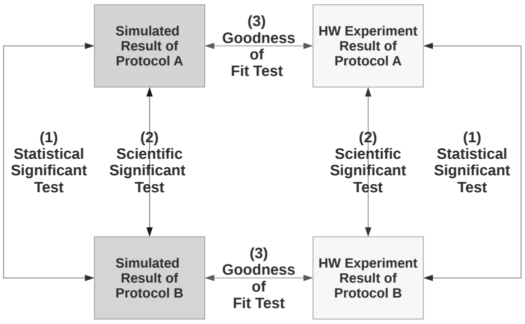


Figure 3: Conceptual Statistical Test Framework

4.2.1 Statistical Significance

Statistical significance test can be used to determine whether the difference in performance observed in the results is likely to have occurred due to random chance with the samples available, i.e. whether protocol X is really better than Y or whether the results are so close than differences are purely random. In order to determine and compare the relationship between the two samples collected from the experiments, Mann-Whitney-Wilcoxon, also known as rank-sum test, [14] is applied. This non-parametric test is used as it does not make any particular assumptions about the distribution of the result, avoiding the need to verify the data conform to the test assumption. Another benefit of using the rank-sum test is the statistics generated from this test can be used to perform scientific significant test.

To perform rank-sum test, a null hypothesis H_0 is defined. The H_0 states that the results have identical distributions; the alternative hypothesis H_a states that the distributions are different. Using a 5% significance level, H_0 can be rejected if the p-value of the test is $< 5\%$, indicating that any observed difference in the results is unlikely to occur by chance, and our results are statistically significant.

4.2.2 Scientific Significance

It is possible for the observed performance improvement between the protocols to be statistically significant but underlying differences are small, i.e. unimportant, and only noticeable due to the large amounts of data obtained. It is also important to examine the scientific significance of results, to measure the difference or the effect size between the protocols. Another non-parametric test known as the Vargha-Delaney A-statistic is used to measure the effect size [13]. A-statistic in the range $[0, 1]$ is obtained using the parameters collected from the previous rank-sum test. Using the guidelines given by Vargha and Delaney in [13], A-statistic value of 0.5 shows no significant difference in effect size for the protocol performance. A-statistic < 0.29 and > 0.71 is required as it indicates a large effect size.

4.2.3 Goodness of Fit

The goodness of fit allows us to compare the relationship between the hardware and simulated results. It measures the discrepancy in the results and allows us to deduce the similarity of the simulation and hardware experiments. We apply the Kolmogorov-Smirnov (K-S test) test to determine whether two samples are drawn from identical distributions to measure the goodness of fit. The K-S test is a non-parametric test for the equality of continuous to compare two samples by quantifying the distance between the empirical distribution functions of two results. The null hypothesis $H4_0$ for the K-S test states that the samples are drawn from the same distribution; the alternative hypothesis $H4_a$ states that the distributions are different. We reject H_0 if the p-value of the test is $< 5\%$ at 5% significance level and scientific significance is invalid.

4.3 Implementation

We have implemented and evaluated the MRP algorithm in both TelosB motes programmed using TinyOS and NS-2 simulator. TelosB with TinyOS 2.1.1 is selected for our experiments since NST-AODV was tested and evaluated using the same platform by Gomez et al. [2]. With the source available for download, it can easily be modified to AODV and extended to support MRP-AODV. NS-2.34 is selected in our experiments since it has included the IEEE 802.15.4 module, developed by Zheng et al. [17]. Further extension was also implemented to support NST-AODV and MRP-AODV and available to download[3].

4.3.1 TelosB Experimental Setup

We have set up a testbed network as shown in Figure 4 in a grid topology using six TelosB motes. A small number of nodes are chosen so that greater control can be provided for the experiments. That is, the experiments are performed in the centre of a large room relatively free from uncontrolled radio sources, however a larger physical network would then be closer to uncontrolled noise sources. Node 1 in the network is configured to collect temperature reading from the sensor and transmit the packet to node 6, at regular intervals of 250ms via the intermediate nodes using multihop routing protocol. Each node is placed in a position of 2 feet from the other node so that it can only communicate with its immediate neighbours. The transmission power in each node is also set to a range of about 2.5 feet to avoid any interference to other non-neighbouring nodes. Radio channel 26 is used to avoid any interference with other Wi-Fi operating in the same band. An acknowledgement for each packet transmitted is also enabled for LLN operation. Each node also logs its network activities on the on-board flash memory, that are later retrieved for analysis. An extra mote, assigned as the controller, is used to synchronise the clock and collect the network statistics from the flash memory of the nodes. A simple time synchronisation algorithm based on flooding mechanism [8] is implemented in the controller. The controller will communicate with all the nodes on the network without using any multihop routing. A 5s initialisation period is allowed for synchronisation before the actual data communication begins.

4.3.2 NS-2.34 Simulation Setup

In order to compare the testbed against simulation, extensive simulations were performed using Network Simulator (NS2), based on the same network, using the parameters shown in Table 2. Previously, we have evaluated MRP and have achieved better performance on a larger network. For this simulation, we have designed a controlled experiment based on 6 static nodes placed in a 2 x 2 grid that mirrors the real world deployment in Figure 4. Node 0 is configured to transmit to node 5 at every 250ms and 35 repeated run was conducted.

---

[3] http://rtslab.wikispaces.com/file/view/mrp.tar

Figure 4: TelosB network setup

Table 1: Tinyos Configurations

Parameters	**Values**
Tx interval:	250ms
Tx Channel:	26
MAC:	802.15.4 (CSMA/CA)
Route Protocol:	AODV, NST-AODV, MRP
Data Queue:	1 (AODV), 1 (NST-AODV), 6 (MRP)
Control Queue:	0 (AODV), 5 (NST-AODV), 0 (MRP)
RREQ Attempts:	1 (AODV), 3 (NST-AODV), 1-3 (MRP)

Table 2: NS Simulation Parameters

Parameters	**Values**
Tx interval:	250ms
MAC:	802.15.4 (CSMA/CA)
Routing Protocol:	AODV, NST-AODV, MRP
IFQ Size:	10 packets

4.4 Simulating Different Transient Failures

WSNs are susceptible to network disruption due to external interference. Depending on the nature of the interference, it can either cause a permanent or periodic failure that can be different in durations. To investigate how these failures can affect the behaviour of the nodes performance, failures are injected to the network by switching off the radio of an active node along the route at different intervals and durations. Five different failure durations mainly: 0.25s, 1s, 2s, and 5s, with different intervals were used to represent different types of network activities that can interfere with the motes radio communication. 15 sets of results are collected for each experiment. The effects of these failure durations are analysed and evaluated.

4.5 Evaluation Metrics

In order to compare the different routing protocol, the same metrics proposed in [6] are used as the input for the statistical tests to evaluate the significance of the routing protocols. These metrics were chosen , based on the work of the extensive models of Tate [11,12], as they represent the reliability, performance and efficiency of the protocol running in the networks.

- **Packet delivery ratio:** Packet delivery ratio (PDR) is used to measure the network reliability and is represented as the percentage of the number successful packet received to the total number of packet transmitted.
- **Average Delay:** Average delay measures the network performance and is calculated as the sum of the time required to send each packet over the total number of packets received. For better performance, a low average delay is required.
- **Normalised Routing Overhead:** Routing overhead is calculated as the normalised ratio of the sum of the total number of routing packets send to the total data packet received. It is used as an indicator to measure the amount of energy used to a data packet. A low value is desirable as it can represent small amount of energy is wasted for communication during route discovery.

5 Results and Analysis

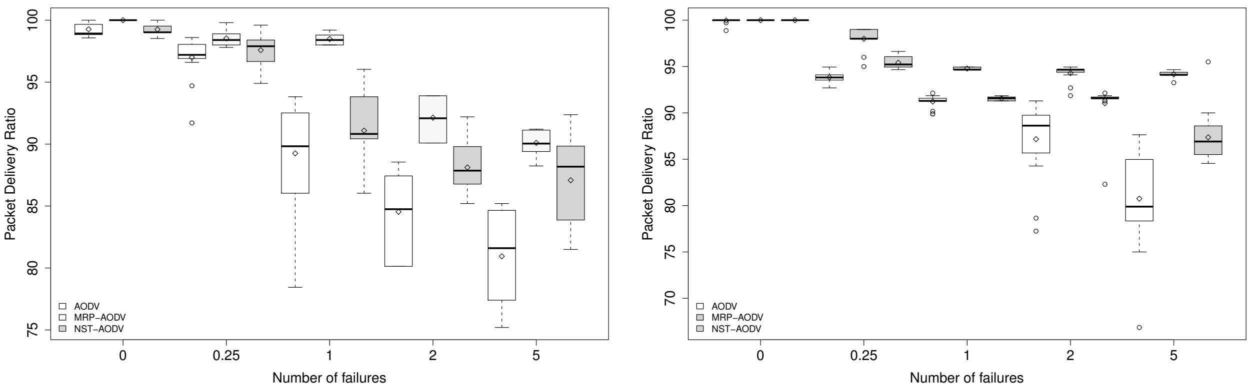
In this section, the results obtained from both simulation and hardware experiments are presented and analysed using boxplots as shown in 5. Boxplots have the capability to show the differences between results using the median and IQR without making any assumptions of the underlying distribution as they are non-parametric. The centre line in boxplot represents the median while the bottom and top edges of the box show the first and third quartiles respectively. Outliers are presented by dots outside the whiskers. The mean value is also plotted in a boxplot, represented by a diamond point. In addition, we run statistical tests discussed in 4.2 to analyse and verify that the results obtained are significant and have scientific values.

5.1 PDR Evaluation

In order to compare the reliability of each routing protocol, the median and mean value of the PDR is presented using boxplots in Figure 5a and 5b.

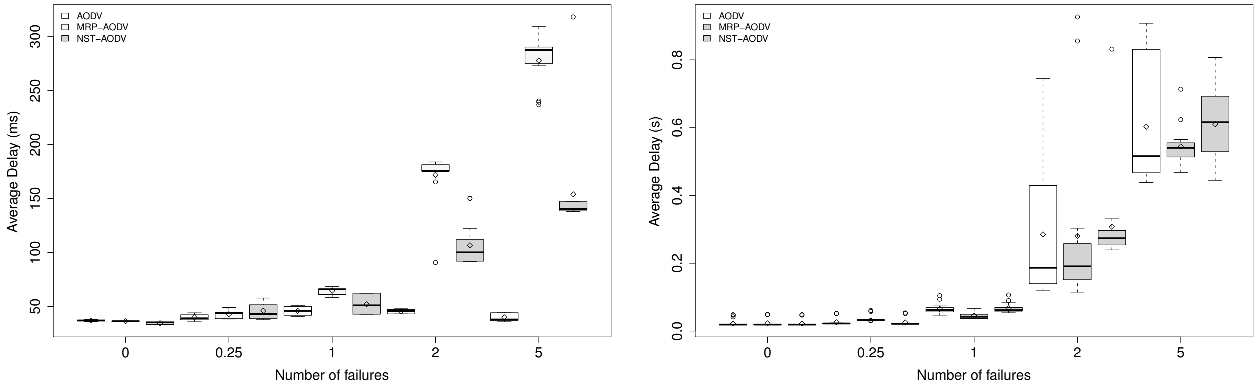
5.1.1 TelosB Experiment

In all failure scenarios, above 90% success rate has been achieved by MRP. This performance improvement is significantly different and higher than AODV and NST-AODV as the A-value given in column 4 of Table 4 is greater than 0.93 and the p-value < 0.05. Before failures were injected into the network, about 3% of the packets sent were dropped in AODV and 1% in NST-AODV. These packets were dropped during route discovery, where a significant



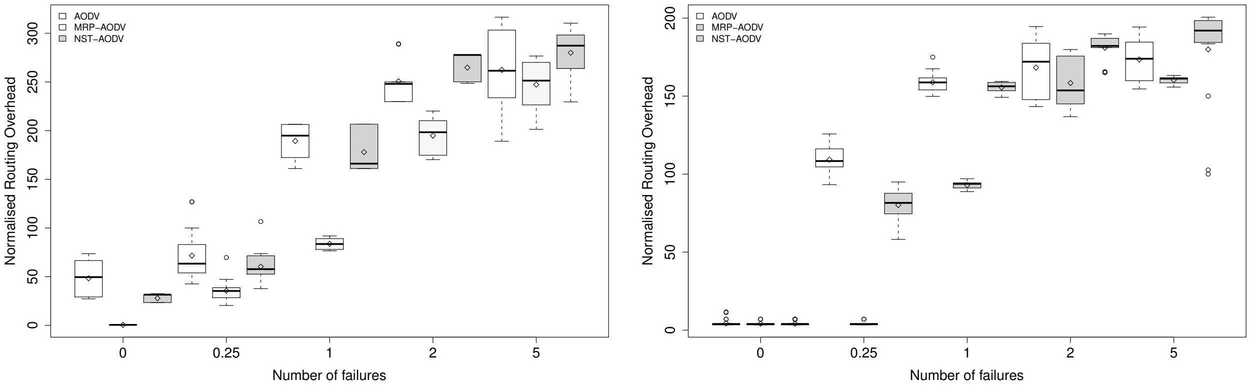
(a) TelosB PDR

(b) NS-2 PDR



(c) TelosB Average Packet Delay

(d) NS-2 Average Packet Delay



(e) TelosB Normalised Routing Overhead

(f) NS-2 Normalised Routing Overhead

Figure 5: Box-Whisker plot with Mean, Median and Inter-quartile range

number of RREQ packets were observed. As we have conducted this experiment under a controlled environment in a small network, we believe these packet loss is due to random errors between the motes e.g. through collisions. When the radios of node 4 and 5 are turned off at regular interval for 0.25s, MRP managed to deliver around 98% packet compared to 96% for AODV. More routing packets were observed in AODV causing other node to drop their data packets due to collision. When we gradually increased the failure duration from 0.25s to 5s, MRP has managed to prevent less than 10% packet loss compared to 15% in AODV, and 11% in NST-AODV as shown in Figure 5a.

5.1.2 NS-2 Simulation

Figure 5b shows the boxplot plot of PDR obtained NS-2 simulations. During normal operation, when no fault is injected to the networks, all the three protocols have achieved

100% PDR. When faults are injected into the node, by turning off an active node along the two possible paths, AODV starts to drop packets by 5% due to unavailability of next hop neighbour. These numbers keep on increasing as we gradually increase the duration of failure in all the three routing protocols. With our RSM in MRP, we have maintained over 90% packet received in MRP compare to 80% in AODV and 87% in NST-AODV on average. In terms of performance improvement, this is over 10% in AODV and 5% in NST-AODV at 5s. p-values in Table 3 column 3 has demonstrated statistical significance differences between the routing algorithms. The performance of MRP is significantly more reliable than AODV and NST-AODV. A large effect size is also achieved for all failures.

5.1.3 Comparing TelosB against NS-2

We also tried to validate NS-2 simulator with the TelosB motes using the goodness of fit statistics, based Kolmogorov-Smirnov Test, on the samples collected from hardware and simulation. From Table 5, a small test cases has shown similarity in PDR with a high p-value > 0.7 supporting H_4. We believe this low number of similarity could be caused by random radio noise in hardware that is not modelled in NS-2. During error-free condition, a few packets have failed to reach the destination in our hardware experiments in AODV and NST-AODV that was not observed in NS-2. The real sensor motes are sensitive to communication failures that can be tolerated by MRP but not NST-AODV and AODV.

5.1.4 Discussion

The statistical analyses between MRP and NST, and MRP and AODV have shown that the performance improvement between the two algorithms is both scientifically and statistically significance with a large effect size and a small p-value supporting H_1. This improvement is due to the RSM switching mechanism that has increased the probability of the packet being delivered. The result obtained from hardware is also better than NS-2 as MRP has maintained a higher packet delivery ratio during short errors. In term of similarity between hardware and software, PDR can only provide partial evidence for supporting H_4.

5.2 Routing Overhead

In WSNs, additional routing packets are required to establish a route to a destination when a link is unavailable. These additional packets are known to create additional overhead in the network and node. We have analysed the routing overhead between the three routing algorithms using boxplots in Figure 5e and 5f.

5.2.1 TelosB Experiment

When we increase the radio failure duration in the active node from 0.25 to 5s in the testbed experiments, the routing overhead increases in AODV. This overhead is less in MRP during small failure durations as shown in Figure 5e. The switching mechanism in MRP can abort RT and switch immediately to RT when the next hop neighbour is available for communication making it capable to operate more efficiently during transient random error. The MRP routing overhead gradually increases with failure duration until it is similar to AODV at 2s. When the cost of RT is higher than RD, MRP switches to RD after successive failures in RT. As for NST-AODV, significantly different routing overheads were observed after 2s with a p-value<0.01 due to additional RREQ packets generated during failure. These differences are statistically significant as represented by the small p-value in Table 4 at 5% confidence level.

5.2.2 NS-2.34 Simulation

The simulated routing overhead is shown in Figure 5f. During normal condition, each successful packet received requires only 7 routing packets to be sent on average for all the routing protocols. This number increased linearly for NST-AODV and AODV with failure duration where they peaked at 2s. However, the overhead in MRP is less than AODV and NST-AODV as shown by the mean and median in Figure 5f. It is also statistically significant and support H_3 as indicated by a low p-value in 3.

5.2.3 Comparing TelosB against NS-2

By analysing the mean and median values in Figure 5f and 5e, each failure scenario has shown differences in routing overhead between the hardware and simulator. Based Kolmogorov-Smirnov Test in Table 5, all the failure scenarios have shown a small p-values validating the differences between experimental and simulation results and rejecting H_4.

5.2.4 Discussion

From Figure 5e, the median and mean of routing overhead in MRP is smaller than AODV and NST-AODV. This difference is both statistically and scientifically significant as shown by the low p-value and high effect value of A-value. Hence, the results support H_3. However, there is no evidence from the KS-Test available to support H_4 as the overheads from the simulation are less than hardware.

5.3 Average Packet Delay

Figure 5c and 5d show that the average time delay increases with the failure duration for RD in all the three protocols.

5.3.1 TelosB Experiment

The delay observed in the testbed experiment in Figure 5c is smaller in AODV than MRP and is significantly different as shown by the small p-value in Table 4. There are two factors contributing to these observations. First, RD is the only delay in AODV, and this delay is minimal in this small network compared to waiting time incurred by MRP and NST during RT. Secondly, the packets that cannot be transmitted in AODV, were dropped and not accounted for in the

Failure Duration	Routing Protocols	PDR		AVG DELAY		Routing Overhead	
		p-value	*A*-value	*p*-value	*A*-value	*p*-value	*A*-value
Normal	MRP/AODV	0.3564	0.5278	0.9210	0.5095	0.9210	0.5095
	MRP/NST	1	0.5000	0.9210	0.5095	0.9210	0.5095
	NST/AODV	0.1602	0.5278	0.9865	0.5015	0.9865	0.5015
0.25 sec	MRP/AODV	**1.0306e-06**	**1**	**8.1995e-05**	**0.9102**	**1.7680e-04**	**0.8906**
	MRP/NST	**2.2483e-05**	**0.9336**	**1.7680e-04**	**0.8906**	**1.7680e-04**	**0.8906**
	NST/AODV	**3.8289e-06**	**0.9766**	0.0734	0.3125	0.9249	0.5117
1 sec	MRP/AODV	**8.9372e-07**	**1**	**8.1995e-05**	**0.9102**	**3.6860e-04**	**0.1289**
	MRP/NST	**8.4342e-07**	**1**	**5.0878e-05**	**0.9219**	**1.1195e-04**	**0.9023**
	NST/AODV	0.1097	0.6602	0.9249	0.5117	0.8358	0.5234
2 sec	MRP/AODV	**1.3217e-06**	**1**	**1.5449e-06**	**1**	0.7203	0.5391
	MRP/NST	**1.5946e-06**	**0.9902**	**1.5449e-06**	**1**	**0.0136**	**0.7578**
	NST/AODV	**2.0841e-05**	**0.9375**	0.3365	0.6016	0.1809	*0.6406*
5 sec	MRP/AODV	**1.0306e-06**	**1**	0.5847	0.5586	**0.0014**	**0.8320**
	MRP/NST	**1.9077e-05**	**0.9375**	**0.0400**	**0.2852**	**5.0878e-05**	**0.0781**
	NST/AODV	**1.2815e-04**	**0.8984**	0.0935	*0.7148*	**3.2464e-06**	**0.9844**

Table 3: *p* and *A* values for Simulation (Bold highlights significance value)

Failure Duration	Routing Protocols	PDR		AVG DELAY		Routing Overhead	
		p-value	*A*-value	*p*-value	*A*-value	*p*-value	*A*-value
Normal	MRP/AODV	**2.6433e-05**	**0.9000**	**0.0121**	**0.2311**	**6.2648e-07**	**0**
	MRP/NST	**7.4129e-06**	**0.9333**	**2.6191e-06**	**1**	**4.2844e-07**	**0**
	NST/AODV	0.9332	0.5111	**2.6744e-06**	**0**	**0.0179**	**0.2489**
0.25 sec	MRP/AODV	**0.0038**	**0.8111**	0.0457	0.7156	0.2980	0.3867
	MRP/NST	0.0860	0.6844	0.3182	0.3911	0.1832	0.3556
	NST/AODV	0.4288	0.5867	**0.0375**	**0.7244**	0.8029	0.5289
1	MRP/AODV	**2.7362e-06**	**1**	**3.0894e-06**	**1**	**3.0035e-06**	**0**
	MRP/NST	**3.8641e-06**	**1**	**1.9659e-04**	**0.9048**	**4.0074e-06**	**0**
	NST/AODV1	0.3663	0.5958	**0.0155**	**0.7500**	0.4275	0.4167
2 sec	MRP/AODV	**2.5246e-06**	**1**	**2.7592e-06**	**1**	0.7670	0.5333
	MRP/NST	**3.3530e-05**	**0.9422**	**5.2013e-05**	**0.9333**	0.7671	0.5333
	NST/AODV	**0.0052**	**0.8000**	**3.0194e-06**	**1**	**0.0330**	**0.7289**
5 sec	MRP/AODV	**3.1529e-06**	**1**	**3.2829e-06**	**1**	0.5057	0.4267
	MRP/NST	**0.0303**	**0.7333**	**5.1811e-05**	**0.9333**	0.1217	0.3333
	NST/AODV	**4.6293e-04**	**0.8756**	**2.9130e-06**	**1**	0.9334	0.4889

Table 4: *p* and *A* values for Hardware Experiment (Bold highlights signicance value)

delay calculation. The average delay observed in MRP is due to switching and buffering that occur during the fault period. If we normalise the average delay, by assigning a delay value to the dropped packets in all the protocols, the delay in MRP-AODV is less than AODV and NST-AODV in the testbed experiment as seen in Figure 6.

5.3.2 NS-2.34 Simulation

In simulation, the results show a different delay pattern was observed when the failure duration is increased in AODV and NST-AODV in Figure 5d. This increased delay pattern is due to the additional queueing of the outgoing packet between link and mac layer in NS-2 simulation. During failure, the packets to be transmitted are placed in the queue. This queue size is set to 10 in simulation, while in TelosB, it is set to 1. When a 5s failure duration is injected, over 10 packets with delays between 2 - 9.55s, were found in the simulation during each failure duration in AODV. This

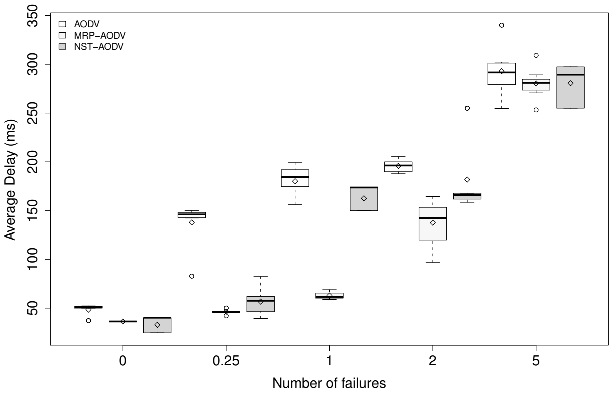


Figure 6: Testbed Delay with Normalised Delay

packet delay is 2s when we change the failure duration to

2s. Hence, the increased delay in the average packet delay, that was not observed in the hardware experiment, is due to this queue. If we set the IFQ value to 1, no packet was received at the sink.

5.3.3 Comparing TelosB against NS-2

There is significant difference between the hardware and simulated results as shown by the small p-values (6×10^{-8}) in Table 5 due to the design and configuration of the NS-2 implementation. In NS-2, failed packets are usually buffered and transmitted when the network is up. In the simulation, if we set the data buffer similar to the hardware, and no packet can be delivered. Hence, the average packet delivery metric was not able to verify H_4.

Routing Protocol	F_{rate}	PDR	DLY	RT
MRP	0	**1**	6.0708e-08	6.0708e-08
	0.25	**0.7925**	6.0708e-08	6.0708e-08
	1	1.0778e-07	1.0778e-07	1.0778e-07
	2	3.5093e-06	6.0708e-08	3.5093e-06
	5	6.0708e-08	6.0708e-08	3.5093e-06
NST	0	5.3056e-08	1.6377e-10	1.6377e-10
	0.25	1.8119e-04	6.0708e-08	6.0708e-08
	1	0.0039	6.0708e-08	6.0708e-08
	2	3.9803e-06	6.0708e-08	6.0708e-08
	5	**0.2395**	6.0708e-08	0.0023
AODV	0	5.1399e-06	1.6377e-10	1.6377e-10
	0.25	3.9803e-06	6.0708e-08	6.0708e-08
	1	0.0071	3.5455e-08	3.5455e-08
	2	0.0264	6.0708e-08	6.0708e-08
	5	**0.8467**	6.0708e-08	4.9386e-07

Table 5: KS Test p-values (Significance highlighted in Bold) between NS2 and TinyOS

5.3.4 Discussion

From the analysis, the large p-value and small effect size allow us to support H_2. However, we cannot accept H_4 due to a very small p-value. In order to bring the result closer, the NS-2 needs to be reprogrammed and finetuned to incorporate the module available in TinyOS.

6 Conclusion

In this paper, we have improved the credibility of our experiments by using a structured statistical and experimental approach to analyse the data taken from repeated experiments with different test scenarios. Using statistical hypothesis test on both experimental and simulated results, we have shown the performance improvement of MRP is both statistically and scientifically significant with 95% confidence. We have demonstrated that MRP have improved the network reliability without increasing the latency or routing overhead. Due to the low number of similarity between real world and simulation, further work is required to look at how to make the simulation model more realistic to minimise the goodness of fit between hardware and software.

References

[1] D. Bein, Y. Wen, S. Phoha, B. B. Madan, and A. Ray. Distributed network control for mobile multi-modal wireless sensor networks. *Journal of Parallel and Distributed Computing*, 71(3):460 – 470, 2011.

[2] C. Gomez, P. Salvatella, O. Alonso, and J. Paradells. Adapting AODV for IEEE 802.15.4 mesh sensor networks: Theoretical discussion and performance evaluation in a real environment. In *Proc. IEEE International Symposium on World of Wireless, Mobile and Multimedia Networks*, pages 159–170, 2006.

[3] Z. J. Haas and M. R. Pearlman. The performance of query control schemes for the zone routing protocol. *IEEE/ACM Transactions of Networks*, 9:427–438, 2001.

[4] S. Ivanovn, A. Herms, and G. Lukas. Experimental validation of the ns-2 wireless model using simulation, emulation, and real network. *ITG-GI Conference on Communication in Distributed Systems*, pages 1 – 12, 2007.

[5] S. Kim, S. Lee, H. Ju, D. Ko, and S. An. Priority-based hybrid routing in wireless sensor networks. In *IEEE Wireless Communications and Networking Conference*, pages 1–6, 2010.

[6] T. Lim, I. Bate, and J. Timmis. Multi-modal routing to tolerate failures. In *Proc. of the 7th International Conference on Intelligent Sensors, Sensor Networks and Information Processing*, pages 211–216, 2011.

[7] C. E. Lopes, F. D. Linhares, M. M. Santos, and L. B. Ruiz. A multi-tier, multimodal wireless sensor network for environmental monitoring. In *Proc. of the 4th international conference on Ubiquitous Intelligence and Computing*, pages 589–598, 2007.

[8] M. Miklós, K. Branislav, S. Gyula, and L. Ákos. The flooding time synchronization protocol. In *Proc. of the 2nd international conference on Embedded networked sensor systems*, pages 39–49, 2004.

[9] C. E. Perkins and E. M. Royer. Ad-hoc on-demand distance vector routing. In *Proc. IEEE Workshop on Mobile Computing Systems and Applications*, pages 90 –100, 1999.

[10] N. Pham, D. Pediaditakis, and A. Boulis. From simulation to real deployments in wsn and back. In *IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks*, pages 1 –6, 2007.

[11] J. Tate and I. Bate. Sensornet protocol tuning using principled engineering methods. *The Computer Journal*, 53:991–1019, 2010.

[12] J. Tate, B. Woolford-Lim, I. Bate, and X. Yao. Comparing design of experiments and evolutionary approaches to multi-objective optimisation of sensornet protocols. In *Proceedings of the 10th IEEE Congress on Evolutionary Computation*, pages 1137–1144, May 2009.

[13] A. Vargha and D. Delaney. A critique and improvement of the CL common language effect size statistics of McGraw and Wong. *Journal of Educational and Behavioral Statistics*, 25(2):101–132, 2000.

[14] F. Wilcoxon. Individual comparisons by ranking methods. *Biometrics Bulletin*, (6):80–83, 1945.

[15] Y. Wu, K. Kapitanova, J. Li, J. Stankovic, S. Son, and K. Whitehouse. Run time assurance of application-level requirements in wireless sensor networks. In *Proceedings of the 9th ACM/IEEE International Conference on Information Processing in Sensor Networks*, IPSN '10, pages 197–208, 2010.

[16] F. Zhao. Challenge problems in sensornet research. Keynote at NSF NOSS PI meeting and Distinguished Lectures, 2005.

[17] J. Zheng and M. Lee. *A comprehensive performance study of IEEE 802.15.4*. 2006.