On the Topological Repeatability of Experiments with Wireless Multihop Networks

Wolfgang Kiess Andreas Tarp Martin Mauve
Computer Networks Research Group
University of Düsseldorf, Germany
{kiess, tarp, mauve}@cs.uni-duesseldorf.de

ABSTRACT

Repeatability has most often been neglected in experiments with wireless multihop networks. In this paper, we propose to consider repeatability in a coarse-grained fashion on a topological level. For this, a metric for comparing the similarity of topologies in static and mobile setups is presented and used to examine the level of repeatability achievable in such experiments. This metric is able to classify experiments according to the presence or absence of interference and to changes in the behavior of mobile nodes. It can be used to identify experimental runs where application layer performance is influenced by topological effects.

Categories and Subject Descriptors

C.2.1 [**Network Architecture and Design**]: Wireless communication

General Terms

Experimentation, Measurement, Performance

Keywords

Experiments, experiment similarity, comparability

1. INTRODUCTION

Repeatability, the "closeness of the agreement between the results of successive measurements of the same measurand carried out under the same conditions of measurement" [8] is one of the cornerstones of scientific evaluations. In realistic experiments with wireless multihop networks (WMNs) like mobile ad-hoc networks, mesh networks or sensor networks, up to now repeatability was addressed by letting all devices perform similar actions in multiple runs. This approach is founded on the implicit assumption that repeated actions lead to repeated experimental runs that can be compared or averaged [2, 4]. In this paper, this assumption is examined.

To this end, we study the similarity of repetitions of very simple, identically executed, strictly controlled experiments. Thus, the controllable factors are all similar between the repetitions and the uncontrollable factors are isolated. However, monitoring all these factors, on the physical layer for example, can be extremely expensive and it is not clear if and how all relevant factors can be recorded. Instead, we propose to use the network topologies as indicator for their influence. This has the advantage that all physical layer factors, even the unknown ones, are represented therein. Furthermore, instead of requiring special hardware or modifications to the used software, the corresponding values can be recorded with standard tools.

In the initial step, a simple metric to quantify the topological similarity of experimental runs is defined. We show that this metric is sensitive to interference and distinct node behavior. In a further step, the occurring topology variations in real-world environments both for static and mobile setups are studied. This examination shows that even runs with mobile nodes can be very similar, however we repeatedly discovered groups of runs that strongly differ from the rest. Thus, the existing concept, executing runs under strict control, must be complemented with an additional step in the experiments' post-processing phase. Here, the reasons for the occurring differences and their impact on the interpretation of the results have to be explored

This paper is structured as follows. In the following Section 2, we review related work. In Section 3, our metric is presented and the experiments can be found in Section 4. Section 5 concludes the paper.

MSWiM'08, October 27–31, 2008, Vancouver, BC, Canada.

2. RELATED WORK

In the context of the APE project [6], the differences of repetitions of experimental runs with mobile ad-hoc networks are considered. For this, signal strength measured with a custom driver is used to calculate the "virtual mobility" as perceived on the radio layer. This calculation results in a two-dimensional time-mobility graph for each run that averages link quality changes over all nodes. By comparing graphs for different runs, the similarity of the average quality change in the runs can be visually assessed. In contrast, our metric only requires layer two information. Similarity is considered on a per-link basis, thus it is also suited for experiments where nodes follow individual movement paths. Furthermore, our metric *quantifies* the similarity between run pairs and thus allows for an automatic comparison.

In [3], it is shown that similar devices can report very different signal-strength values under similar conditions. Thus, they have to be calibrated to reproduce measurements. The need for calibration is furthermore explored by examining a

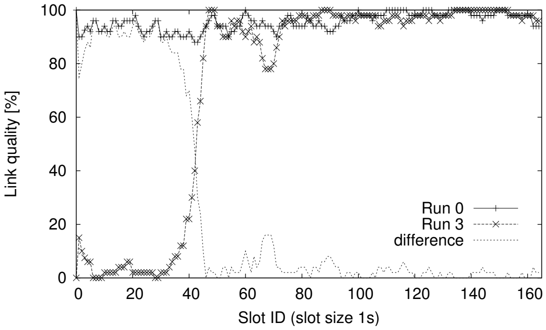


Figure 1: Example for difference calculation.

link in several similar runs conducted over one month. The resulting variations of throughput and signal strength are studied with overlayed time/value plots of these parameters. In [5], we have presented the EXC toolkit and validated it with a study that assessed link similarities also with overlayed plots. In contrast to such descriptive approaches for specific experiments, we now provide a generic method that considers the influence of interference and mobility.

For network anomaly detection, it has been proposed to view a network as graph and model changing topologies as a series of graphs [7]. Techniques like graph edit distance are used to discover the one outstandingly different graph (i.e. topology) in a coarse-grained time series of graphs. By contrast, the examination of topological repeatability requires that a fine-grained series of *differences* between graphs considering the time dimension as well as the average difference be assessed on a per-link basis.

3. A SIMILARITY METRIC

We assume that an experiment is divided in R runs. Two runs have a similar *type* if the controllable factors are similarly set, i.e. the used hardware and software, the actions, and the movements of the devices are similar. To quantify the topological similarity of two runs, we follow a bottom-up approach that combines estimates for the similarity of individual links to an estimate for the topologies.

A wireless link is not simply "up" or "down" but exhibits a quality that varies over time [1, 6]. With link layer information, this quality can be approximated as the fraction of packets successfully delivered in an interval i.[1] To be able to easily compare the quality of the same link in different runs, we divide each run in S similarly sized intervals called *slots*. As the link quality estimate will be rather coarse for a low number of sent packets, the quality is averaged over a sliding window. This has the advantage to be sensitive to movement-induced quality changes. The examination of time variation of loss rates in [1] indicates that short-time fluctuations occur for at least an interval of one second. Thus, the quality calculation is parameterized with a slot size of one second and a window length of five seconds.

Based on this quality, the similarity for the same link in different runs can be assessed as visualized in Figure 1. The graph shows a link in two runs of the same experiment together with the slot-wise absolute value of their difference. To quantify the quality differences, the *average* μ and the *standard deviation* σ of these slot-wise differences are used. In the remainder of this paper, the notation (μ, σ) will be used to characterize such a comparison of the same link in two runs and it will be called *AD metric*. Obviously, the lower the values in these figures, the higher the similarity of the runs. To combine these values to a single value, we take advantage of the fact that a lot of random physical layer factors influence the links and thus the topology. Therefore, we can assume that the difference between the links follows a normal distribution. For a normal distribution, 68.3% of all values lie within an interval of $\mu \pm \sigma$ and 84.2% of all values are below the upper boundary of this interval. By using this upper boundary $\mu + \sigma$ as measure for the link differences, the overall behavior can be quantified while outliers are left out. For the link in Figure 1, the similarity estimate is $(25.0, 37.4) = 62.4\%$, indicating a rather low similarity.

The similarity of two complete runs, i.e. topologies can then be computed as the average AD value of the single links. The lower this average, the smaller the overall difference and the higher the similarity of the runs. Although manually comparing the estimates for all run pairs can become quite laborious, this task cannot be fully automated. The resulting estimates are a relative measure, it is difficult to give absolute numbers for "similar" or "non-similar". Therefore, we use a graphical representation of the similarity estimates called *AD plot* as shown in Figure 4. It displays the estimates between all $R \cdot (R-1)/2$ run pairs in an experiment in ascending order, the x-axis shows the IDs of the compared runs, the y-axis displays the corresponding estimate. Similar run-pairs, for which the curve is flat and does not exhibit major jumps, can be found on the left hand side of the figure. A steep curve and abrupt jumps like in the right part indicate topologies with more differences.

4. EXPERIMENTS

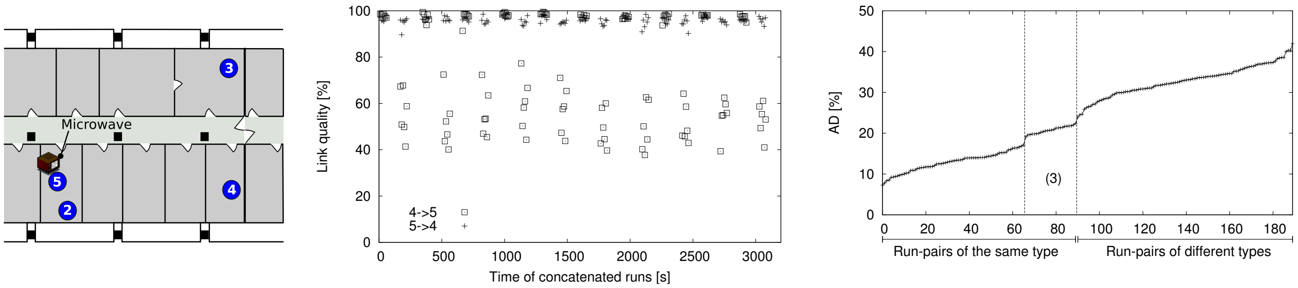
All experiments have been conducted using EXC [5], a toolkit to steer experiments with wireless multihop networks. EXC automates the devices' actions, node movement paths can be exactly controlled and specified. All devices were Linux systems and used the integrated 802.11b cards in adhoc mode as network interface. The laptops used were IBM Thinkpads and the personal digital assistants (PDAs) are ARM-based Sharp Zaurus devices. For every run type, a total of ten repetitions have been performed.

4.1 Interference

The first experiment has been conducted with deliberately provoked interference. To this end, the laptops 2-5 where distributed on our office floor as shown in Figure 2(a). The microwave displayed in the graph operates in the same frequency band as the 802.11b hardware and therefore produces some interference when running. A run in this experiment has a duration of 60 seconds. In runs of type A, the interferer was switched off while it was switched on in runs of type B. Each node broadcasted 40 packets/s with a size of 100 bytes and recorded all incoming packets.

The strongest reaction on the interference can be observed between the nodes 4 and 5, see Figure 2(b). In each second run (runs of type B with interference), the quality for 4$\longrightarrow$5 is 40 to 70% while it exceeds 95% in type-A runs. In contrast, the opposite direction 5$\longrightarrow$4 is barely affected. The AD plot for the 190 possible comparisons between runs of type A and type B can be found in Figure 2(c)[2]. As the links in the two

[1] If no packets are sent, we use linear interpolation.



(a) Positions. (b) Quality of the links 4→5 and 5→4. (c) AD plot for all runs pairs.[2]

Figure 2: The experiment with deliberately provoked interference.

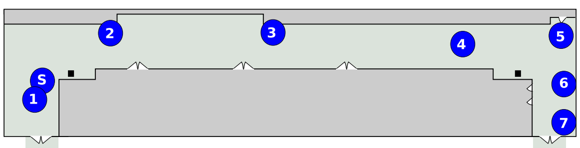


Figure 3: Setup in the basement experiment.

run types behave differently, the 90 comparisons with the lowest AD metric differences are those between runs of the same type (A-A or B-B comparisons). The 100 runs with the highest differences are those of differing type (A-B or B-A). Thus, the metric is able to distinguish runs with and without interference solely based on layer two information.

4.2 Static Setup with a Single Sender

To examine single, isolated links, the next experiment has been conducted in the basement of our university. For this area, we verified with a spectrum analyzer that the 2.4 GHz interference was low and that no other 802.11 network could be received. As shown in Figure 3, we set up seven PDAs and the laptop with ID S as a sender. With only one sender, the other nodes do not produce packet collisions. The source broadcasts sequences of packets with fixed, different sizes of 50, 100, 500, 1000, and 1400 bytes at a rate of 80 packets/s.

The AD plot for this experiment is shown in Figure 4. The values of about half of the possible run comparisons are indeed low and close together, indicating that the compared runs are similar. However, there are also run-pairs starting with the comparison 8-2 that exhibit higher dissimilarity (towards the right end of the plot). A more detailed examination of the involved runs reveals that all comparisons to the right of this gap contain one of the runs 0, 1, or 2. This indicates that there was some interference during these first three runs influencing the topology. An examination of the AD metric of the single links reveals that a number of links contribute virtually no error while some links are responsible for most of the deviation. The links with the highest error contribution are those to the nodes 5 and 6. The reason for this is the bad link quality that varies between 0 and 60% during the experiment. This bad quality is most likely the result of a low signal-to-noise ratio that is affected already by the slightest change in environmental conditions.

[2]The higher differences marked in Figure 2(c) with (3) all occur between runs of type B, i.e. with interference and are an artifact of the microwave's interference pattern.

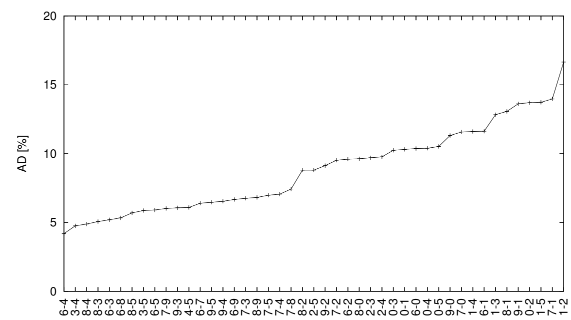


Figure 4: AD plot for the basement experiment.

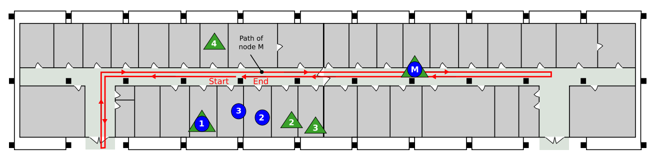


Figure 5: Mobile experiments. Circles mark nodes in the first mobile experiment, triangles mark nodes for the experiment with delayed mobility.

Besides such *unstable links*, three other classes can be identified.[3] The first class is the *zero link* to node 7 that produces no variations as it receives nearly no packets at all. Furthermore, there are *perfect links* to the nodes 1 and 2 with an average difference of 0.7% and a constant quality of almost 100% and *intermediate links* to the nodes 3 and 4 with differences of 3.7% and 6.0% and a quality of above 90% for most of the time.

From these experiments, it can be concluded that the similar behavior of static links in different runs is directly related to link quality: if this quality is high (> 90%) or close to zero, the link is stable throughout the repetitions, an intermediate quality leads to strong variations and instability.

4.3 Mobile Setup in an Office Environment

In the next step we move on to an experiment with mobility and multiple senders. As shown in Figure 5, three static nodes are set up within each others radio range and the mobile node M follows the path indicated by the displayed line.

[3]In a follow-up experiment on our office floor, where more interference is present, similar link characteristics were found.

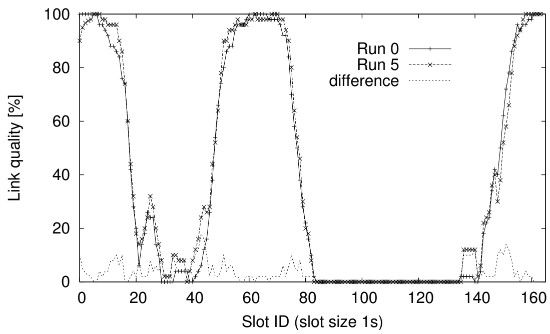


Figure 6: The most similar static→mobile link comparison from 1→M in the office setup.

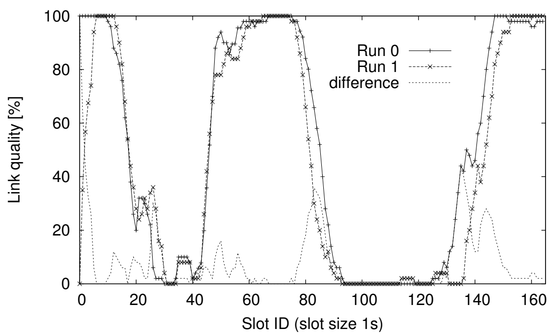


Figure 7: The least similar static→mobile link comparison from 3→M in the office setup.

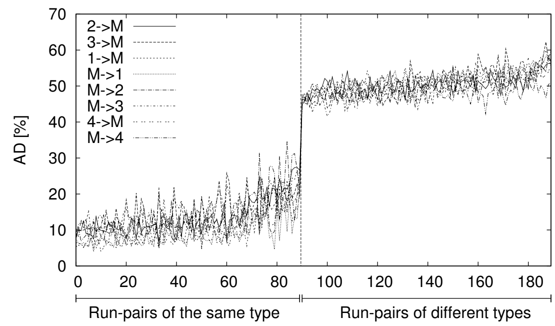


Figure 8: AD values for all mobile↔static links in the experiment with differently-typed runs.

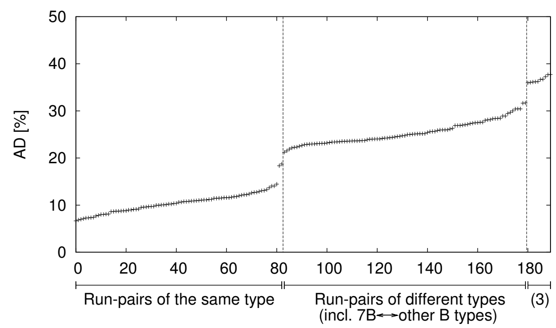


Figure 9: AD plot for the experiment with differently-typed runs. Part (3) represents comparisons of type-A with run 7b.

The track is designed such that the mobile node looses connection to the static nodes in the outermost areas. A run has a duration of 160 s, each node permanently broadcasts 100 byte packets at a rate of 10 packets/s.

The interesting point here is the similarity of links between mobile and static nodes. The best link for the direction static→mobile (s→m) is shown in Figure 6. The average difference in the link qualities is 3.2% with a standard deviation of 3.9%, resulting in an AD value of 7.1%. The similarity for s→m links is in general rather high: even the s→m link with the biggest difference has an AD value of (8.4, 13.3) = 21.7%, see Figure 7. In both graphs, the pattern of connected as well as unconnected phases induced by mobility is clearly visible and exhibits a high repeatability. The highest differences occur in the transition phases from connected to non-connected states. As the links here work in the same quality spectrum as instable links, this is not surprising. Nevertheless, these experiments show that mobile links can be very stable with a high repeatability.

4.4 Mobile Setup with Runs of Different Types

To examine the effect of slight topology changes on repeatability, we have performed an experiment with two types of runs. For both configurations, the setup consists of four static and one mobile node that follows the same path as in the previous experiment, see Figure 5. In runs of type A, the mobile node M directly starts to move when the run begins, whereas this is delayed for ten seconds in type B runs. In both run types, each node broadcasts 20 packets of 100 bytes per second and all other settings are equally similar. Thus, the only difference between the two types is the 10 second movement delay. We have performed ten runs of both types, each with a duration of 170 s.

The effect of this small change in the mobility pattern on the similarity of mobile links is severe. Similar to the delayed movement, the link-quality curves in the corresponding comparisons are shifted by 10 seconds between the different run types, resulting in large AD values and thus a large difference. This effect is visible for all links between mobile and static nodes, see Figure 8. It shows the AD values for all these links, sorted according to their average, for all 190 possible run combinations. The combinations on the left of the gap are those for runs of the same type while those to the right with an AD value of well above 40% are combinations of differing type. Thus, already small changes in the movement can have a severe impact on topological repeatability and can be detected with layer two information.

The AD plot in Figure 9 shows that both run types achieve an equal level of repeatability. For the first 81 entries of the graph (read from the left), the similarity values are close together and only exhibit small variations. All of these entries correspond to comparisons of runs of the same type, i.e., either comparisons of type A with type A or B-B comparisons. With 81 comparisons of similar-type runs, nine combinations of the same type are missing. These are all combinations of run 7b with the other type-B runs. Two of these are prominently positioned as the two crosses in the middle of the gap. The others are intermixed with the A-B/B-A comparisons to the right of the gap. For this di-

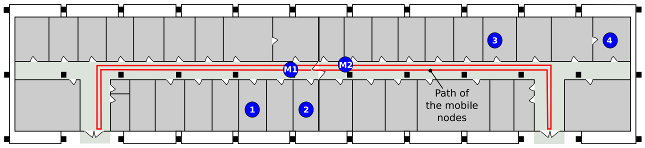


Figure 10: Setup with two mobile nodes that move along the corridor and start in different directions.

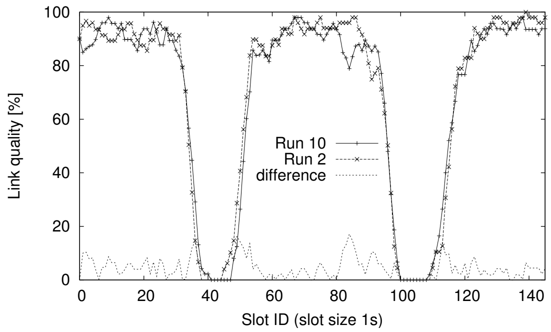


Figure 11: The most similar mobile→mobile link, here from M2→M1. The AD value is (4.4, 4.0).

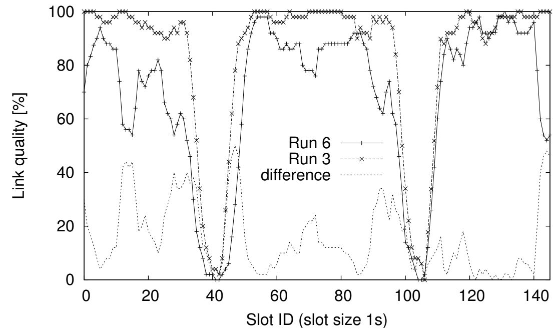


Figure 12: The least similar mobile→mobile link, here from M1→M2. The AD value is (15.7, 13.2).

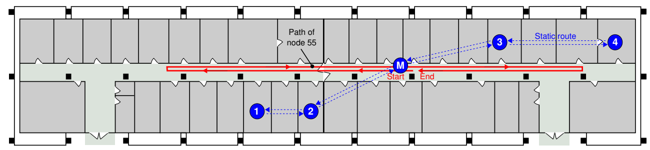


Figure 13: Setup for the unicast experiment. Packets are sent from 51 to 54 and vice versa.

version the two links 1↔3 and 2↔4 between static nodes are responsible. A detailed investigation of the type-B runs reveals that 1↔3 increases in quality in the last three runs and that the quality of 2↔4 drops in run 7B and reaches values equal to the first six runs thereafter. Similar to the basement experiment, these significantly different topologies in the consecutive runs 7b, 8b, and 9b are a strong indicator for changes in environmental conditions.

4.5 Two Mobile Nodes

In the next step, a setup with a second mobile node has been examined, see Figure 10. Here, the two mobile nodes follow the same path but in opposite directions. Each node transmits 100 packets/s, a run had a duration of 160 s. In addition to the broadcast packets, the nodes 1 and 4 sent out multihop unicast packets to each other, however their transmission did not work correctly due to a non-working MAC address resolution for one of the mobile nodes. Therefore, unicasted packets are examined in the next experiment and we concentrate here on the repeatability of the link between the two mobile nodes based on the broadcast packets.

This similarity can be high, the most similar comparison is shown in Figure 11 for the direction M2→M1. It has an AD value of $(4.4, 4.0) = 8.4\%$. The link comparison with the lowest similarity has an AD value of $(15.7, 13.2) = 28.9\%$, see Figure 12. This is mainly a result of the phases of disconnection that are slightly shifted here and the quality variations in the connected phases. Nevertheless, this demonstrates that it is possible to conduct experiments where links between mobile nodes exhibit a highly identical behavior in different repetitions.

4.6 Unicast and an Application Layer Metric

Now we examine unicasted packets in the scenario with four hops displayed in Figure 13. The four laptops (1-4) are set up similar to the previous experiment and we verified by means of pings that the links 1↔2 and 3↔4 have a high quality and the nodes 2 and 3 were separated until no ping could be transmitted any more. This results in a disconnected chain. The mobile PDA with ID M moves from one end of the chain to the other and acts as forwarder between 2 and 3. To allow for a communication over four hops, the static route 1↔2↔M↔3↔4 is set up. The nodes 1 and 4 exchange 100 byte UDP packets via this route at a frequency of 20 packets/s, a run has a duration of 170 s. Also in this experiment, two types of runs are performed where the mobile node delays its movement until second 10 for runs of type B. In the first half of the experiment, five repetitions of type-A runs are executed and then five runs of type B. In the second half of the experiment, this order is reversed. All nodes trace packets in promiscuous mode, thus they also record unicasted packets that are not directed towards themselves.

Due to the distances between static nodes, there is no link between 4 and the nodes 1 and 2, and also for 3→1, the quality is close to zero most of the time. However, the link in the opposite direction, 1→3 changes from a bad to a good quality around the middle of the experiment, and the link 3→2 fluctuates during the whole experiment. This behavior is more pronounced for the reverse direction 2→3. Thus, although the nodes were positioned such that no ping traffic could be transmitted over 3↔2, an instable link exists. As we are interested in the interplay between single links and multihop behavior on the application layer, only the links 1↔2, 2↔M, M↔3, 3↔4 that transport packets along the chain will be considered in the analysis.

As metric for the application layer, the delivery rate over time is used, i.e. the percentage of packets sent in a certain second that arrive at the other end of the chain. An example plot of this metric can be found in Figure 14 for the runs 0a and 3a that achieve one of the highest AD values in the experiment. The movement of the mobile node is clearly reflected on the application layer: starting in a connected state, the chain is interrupted twice as the mobile

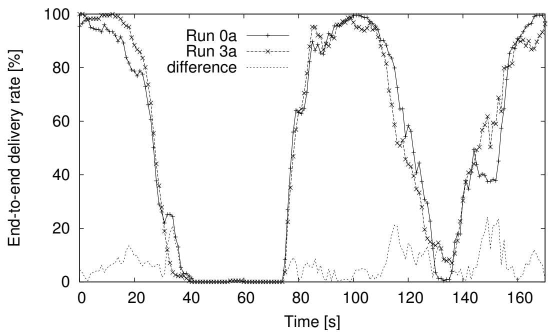


Figure 14: Overall end-to-end delivery rate for both directions for the runs 0a and 3a.

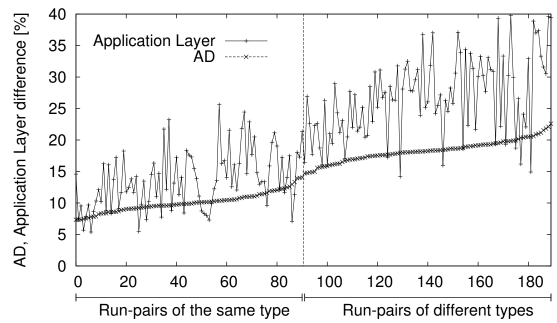


Figure 15: The run comparisons in the unicast experiment sorted according to the AD metric.

node approaches its turning points. It is impressive that the delivery rate over four hops is so similar here, the difference for the displayed runs is 5.7% on average. For the following analysis, this averaged difference will be used to assess the application level similarity.

The AD plot for the whole experiment together with the similarity of the end-to-end delivery rate in these run pairs is shown in Figure 15. The AD similarity estimate of run combinations of equal type (A-A, B-B) is higher than that of all the combinations where the types differ. Thus, here again the AD metric correctly classifies the runs according to the mobility-induced topology differences. However, the transition is not as extreme as in the experiment from Section 4.4. Especially the quality fluctuations in the mobile link 2↔M in runs 7a and 9a have a strong influence and their topological similarity with other type-A runs is therefore rather low: the twelve comparisons between type-A runs with the lowest AD values all comprise either run 7a or run 9a. Thus, again some runs show significant differences.

The application layer metric shows a strong correlation to the topology. In runs of type B in which the movement is delayed, the delivery rate curves are also shifted in time. Nevertheless, as multiple links are involved, their differences amplify, resulting in high variations between the different runs, as shown in Figure 15. In comparison to the similarity estimate of the topology, the curve for the application layer displays high fluctuations. In spite of these variations, a general trend is visible. The delivery rate comparison shows small differences for similar topologies (the values on the left of the plot) and tends towards greater differences if also the topologies are more different. This visual impression is reflected in a correlation of 0.82 between the two displayed metrics. Considering topological similarity here thus allows to identify those run pairs in which the application layer metric variations are the effect of topology changes.

5. CONCLUSIONS

We have considered the repeatability of experiments with wireless single- and multihop networks. For this, we transformed the problem of examining repeatability to the problem of comparing network topologies. We have shown how a simple metric can be created that quantifies differences of static and mobile network topologies solely based on layer two information. Already this simple metric produces good similarity estimates and is sensitive to both interference and mobility-induced topology variations.

In a number of well-controlled experiments, we have examined the similarity of links and topologies. It has been discovered that links can be classified according to their potential similarity and that there is a relation between similarity and link quality. Furthermore, these experiments show that high repeatability can be even achieved for links between mobile nodes. Also for whole topologies, it is possible to perform repetitions with similar outcome. However, we often encountered groups of consecutive runs that show higher differences, indicating changes in environmental conditions. As a consequence, before comparing or averaging runs, it is necessary to verify that they have been performed under sufficiently similar conditions. One central task of future work will be to determine what "sufficiently" means, furthermore it has to be explored how to cope with differing runs.

6. REFERENCES

[1] D. Aguayo, J. Bicket, S. Biswas, G. Judd, and R. Morris. Link-level measurements from an 802.11b mesh network. *ACM SIGCOMM Computer Communication Review*, 34(4), 2004.

[2] E. Borgia, M. Conti, F. Delmastro, and E. Gregori. Experimental comparison of routing and middleware solutions for mobile ad hoc networks: Legacy vs cross-layer approach. In *E-WIND '05*, Aug. 2005.

[3] S. Ganu, H. Kremo, R. Howard, and I. Seskar. Addressing repeatability in wireless experiments using ORBIT testbed. In *TRIDENTCOM '05*, Feb. 2005.

[4] R. S. Gray, D. Kotz, C. Newport, N. Dubrovsky, A. Fiske, J. Liu, C. Masone, S. McGrath, and Y. Yuan. Outdoor experimental comparison of four ad hoc routing algorithms. In *MSWiM '04*, Oct. 2004.

[5] W. Kiess, T. Ogilvie, and M. Mauve. The EXC Toolkit for Real-World Experiments with Wireless Multihop Networks. In *EXPONWIRELESS '08*, June 2008.

[6] H. Lundgren, D. Lundberg, J. Nielsen, E. Nordström, and C. Tschudin. A large-scale testbed for reproducible ad hoc protocol evaluations. In *WCNC '02*, Mar. 2002.

[7] P. Showbridge, M. Kraetzl, and D. Ray. Detection of abnormal change in dynamic networks. In *IDC '99*, Feb. 1999.

[8] B. N. Taylor and C. E. Kuyatt. Guidelines for evaluating and expressing the uncertainty of NIST measurement results. Technical Note 1297, National Institute of Standards and Technology (NIST), 1994.