

## **PH-DISTRIBUTED FAULT MODELS FOR MOBILE COMMUNICATION**

Katinka Wolter

School of Computing Newcastle University  
Newcastle upon Tyne, NE1 7RU, UK

Philipp Reinecke

HP Labs Bristol  
Bristol BS34 8QZ, UK

Tilman Krauss

Daniel Happ

Florian Eitel

Freie Universität Berlin

Takustr. 9, 14169 Berlin, GERMANY

### **ABSTRACT**

In this paper we analyze the quality of wireless data transmission. We are primarily interested in the importance of the distance between sender and receiver when measuring data loss rate and the length of lossy and loss-free periods. The ultimate purpose of this type of study is to quantify the effects of mobility. We have sampled data and find that distance certainly is an important indicator but the loss rate of packets is also determined by other factors and does not always monotonically increase with the distance. We further find that while the distribution of the length of lossy periods mostly shows an exponential decay the distribution of the length of loss-free periods does not even always monotonically decrease. Both, the packet loss probability and the distribution of the length of loss-free periods can be well represented using probabilistic models. We fit simple Gilbert-Elliot models as well as phase-type distributions to the data using different fitting tools and provide loss models that can easily be used in simulation and testbed studies.

### **1 INTRODUCTION**

Wireless networks are used today for various applications. Since modern wireless technology has a competitive data rate and allows for much more flexibility in network layout, a wireless last hop becomes increasingly common in homes and enterprises. With the increase of critical applications that use wireless network connections the quantitative evaluation of network delays and data loss rates for those applications becomes relevant.

Performance and dependability evaluation has been performed for networks for the past decades. Different techniques have been used: measurement studies, simulation and analytical modeling. Measurement studies provide insight into the characteristics of the monitored scenario. They can be used to parameterize simulation and analytical models which will then allow the study of counter-measures to performance and dependability deterioration.

The underlying motivation of the work presented in this paper are research questions of the following type: Given a network setup where wireless technologies are used to provide connectivity for applications relying on TCP/IP (e.g. web browsing), how do packet-loss characteristics of the wireless stack affect quality of service for the application? In order to address such questions, the packet-loss process must be modeled correctly. We investigate the loss process in a measurement study and provide stochastic models that describe the measurements well.

Stochastic models describing data loss have been proposed before (Haßlinger and Hohlfeld 2008; Hohlfeld, Geib, and Haßlinger 2008; Zhang, Duffield, Paxson, and Shenker 2001) but these authors have analysed Internet traffic while we study wireless data transmission. Analysis of wireless communication has concentrated mostly on measurement studies with very little analytical representation of data (Eckhardt and Steenkiste 1996; Tang and Baker 2000; Henderson and Kotz 2006). The objective of this paper is to join both efforts where the focus certainly is on modeling the data as the data sets we collect are small compared to previously sampled data. We have taken measurements on two different days and we use the data to investigate loss behavior at the IP packet level. While we do not aim at a root-cause analysis we describe the impact of unknown faults in the lower layers of the protocol stack by behavioral models at the IP layer, the fault impact models.

The paper is organized as follows. The next section describes the experiment setup, in Section 3 we analyze important properties of the sampled data, Section 4 introduces the most common fault-impact models and in Section 5 we fit those models to our data. Section 6 concludes the paper.

## 2 THE EXPERIMENT SETUP

In the experiments a wireless sender transmits data to a receiver. The objective of the experiments was to relate transmission quality, i.e. loss rate and length of lossy and loss-free periods, with the distance between sender and receiver in a wireless network based on IEEE 802.11g. Therefore, an indoor environment was chosen that allows to increase the distance between sender and receiver step-wise up to a distance of 180m. The experiments were carried out in the corridor of the Physics building at Freie Universität Berlin, as shown in Figure 1. This corridor connects two parts of the building across a street and passes along offices, labs, a library and a lecture hall.

The sender is indicated as the black square, while the receiver at the maximum distance is marked by a black dot. The sender was placed stationary at the black square and the intention was to move the receiver step-wise away from the sender. However, it turned out that moving the receiver away from the sender did not yield meaningful results. Since the network was in infrastructure mode rather than in ad-hoc mode, the connection turned out to either be established or broken and the data loss rate was either close to zero or at 100%.

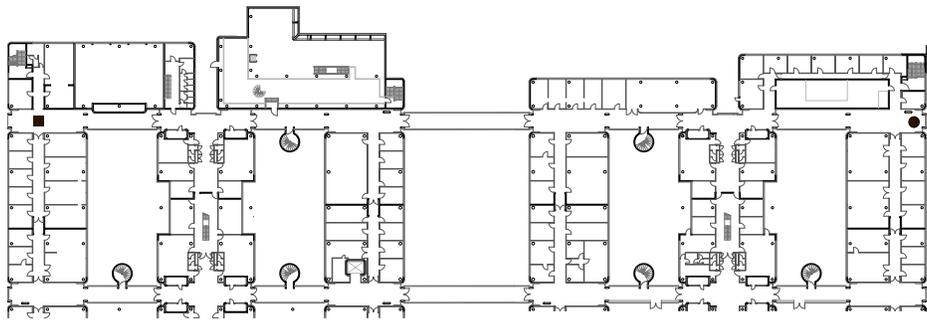


Figure 1: Topology of experimentation environment.

In order to be able to observe deterioration of the transmission quality the setup was slightly modified as shown in Figure 2. The receiver (circled 1) was placed as the black square in Figure 1 and the sender (circled 2) was stationary at a distance of 2m as this showed the best transmission results. Transmission quality depending on distance was captured through a sniffer station (circled 3) in monitor mode. The sniffer station tracked those packets from the data exchange between the stationary mobile node (2) and the base station (1) that were still seen at the respective distance.

In consequence it can be observed that while the stream between sender and receiver remains constant, the sniffer experiences increasing loss as it is moved away from sender and receiver.



Figure 2: Measurement Setup with a data stream between mobile (2) and base station (1) and a mobile station in sniffing configuration (3).

The base station was an ASUS EEE-PC with Atheros chipset configured as a software access point. The sender and sniffer both were Thinkpad laptops with Intel chipsets. The sniffer was moved away from the sender and receiver in subsequent experiments.

To achieve a constant bitrate and avoid IEEE 802.11 rate adaptation, `hostapd` (Malinen 2012) has been used as a software access point on the receiver side. Only a fixed bitrate of 54 MBit/s has been allowed in both the supported rate and basic rate set.

## 2.1 Workload Generation

The workload was created targeting full utilization of the available transmission capacity of 54 MBit/s. The actual observed data rate, however, was only approximately half of the nominal transmission rate. The UDP protocol was used for transmission without any encryption. Retransmissions at the MAC layer were disabled. To use the full Maximum Transmission Unit (MTU) of 1500 Byte the payload of 1454 without the 42 Byte for the IP/UDP header was randomly filled with zeros and ones using `pcap`. The four remaining Bytes are used for a serial number to identify lost packets. The packets were generated with `scapy` and stored in a `pcap`-file which was then transmitted at maximum speed using `tcpreplay` (Tcpreplay 2012).

For monitoring the sniffer laptop we used the tool `airodump-ng` (which is part of the `aircrack-ng` framework (Aircrack-ng 2012)). All traffic was recorded and analyzed as shown and discussed in Section 3. Also note that this sniffing configuration allows us to capture all packets on the same and overlapping channels including beacons and traffic from other networks.

Measurements were performed in stretches of 5 minutes. The first set of experiments was carried out on December 13, 2011 between 3:45pm and 6:03pm and the second set followed on January 3rd, 2012 from 4:50pm until 6:11pm. On the first day of experimentation one measurement series of 5 minutes was carried out at each distance between sender and sniffer ranging from 10 to 180 meters with an increase of 10 meters.

On the second day of experimentation the measurements were repeated 4 times but only selected distances were used. Those are 10, 60, 110, 160, 180, for which in consequence much more data exists. Within one experiment of 5 minutes approximately 524.287 packets were sent.

## 3 DATA ANALYSIS

The collected data sets contain information for every packet. This includes packets sent by our sender but also background traffic from other networks in the building, such as e.g. `eduroam`. The packets we transmitted carry subsequent numbers as part of the header which allow to identify lost packets although data analysis is performed only on the receiver side. The sampled data includes received time, size and content of all seen packets. In addition, a boolean variable is set indicating the successful or unsuccessful transmission for each packet. Hence our data contains among other information a list of binary success indicators for each experiment of 5 minutes duration. This means that for all considered distances between sender and sniffer one such data set exists, for selected distances there are five data sets each with results from 5 minutes data transmission.

From this data the loss probability for one distance setting is computed as the number of lost packets relative to the total number of transmitted packets. Since this yields only one value for each experiment the

results from the experiments have been split into 52 subsequent batches of equal size to create a considerable sample. Please note that each such sample is obtained as the relative number of successful packets of a total of 10.000 packets.

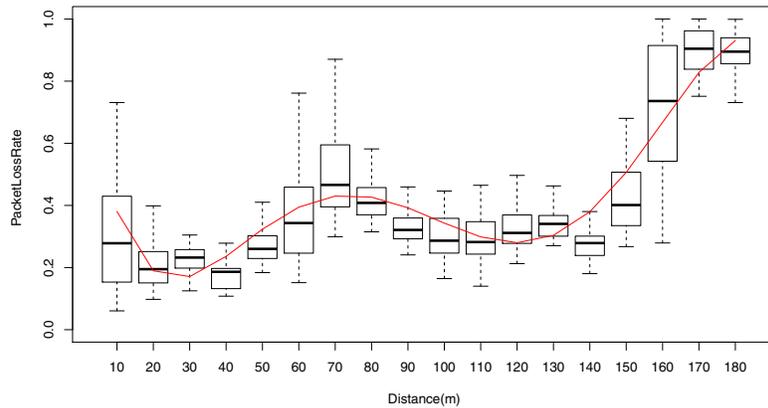


Figure 3: Packet Loss Rate vs. Distance on both days combined including a fit with the polynomial (red).

Figure 3 shows a box plot of the computed data of relative loss frequencies using all data. Obviously, although the distances with repeated experiments rely on a larger sample, the variation in the data for the distances 10, 110, 160 and 180 meters still was quite large. Since data across all distances only exists for the first day of experimentation only this data set will be used for the further analysis.

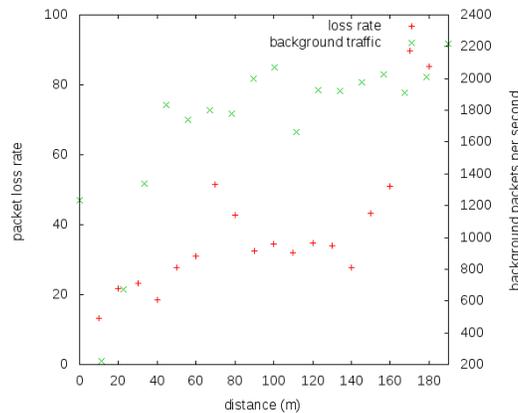


Figure 4: Background packet rate and loss rate for all distances.

As in many wireless communication experiments, not all effects shown in the figures can be directly explained. Clearly, the loss ratio increases significantly as the distance between the sniffer and the sender exceeds 160 meters. However, the high loss ratio at 70 meters distance sampled starting at 4:26pm on December 17, 2011 can not with certainty be related to any circumstances. This could be attributed to a larger number of background packets from other sources as possibly caused by other wireless traffic in the library or the lecture hall located approximately at 70m away from the sender on the ground floor and the first floor respectively.

As an analytical representation of the mean loss rate we fit a polynomial of degree 7. This polynomial can be used to determine the mean loss rate for a given distance  $d$  between sender and receiver in a

simulation study or for fault-injection in a test bed. The polynomial was fitted using the statistics tool R (R Development Core Team 2008).

$$loss(d) = -3.165385 \cdot 10^{-10} \cdot d^5 + 1.595935 \cdot 10^{-7} \cdot d^4 - 2.864736 \cdot 10^{-5} \cdot d^3 \quad (1)$$

$$+ 2.209461 \cdot 10^{-3} \cdot d^2 - 6,760923 \cdot 10^{-2} \cdot d + 8.625900 \cdot 10^{-1} \quad (2)$$

To further investigate the increase in the loss rate at a distance of approximately 70 meters Figure 4 shows the number of background packets observed per second during our experiments together with the average loss rate.

The packet loss rate is shown on the y-axis on the left while the number of background packets are indicated on the y-axis on the right. Clearly, both show a swift trend towards higher values as distance increases. This observation is corroborated by the coefficient of correlation between both variables of 0.5310328 which indicates positive correlation. However, the peak in the loss rate at 70 meters is not accompanied by a peak in the background traffic rate. So the issue remains unresolved. Sources of interference could be other devices emitting noise on 2.4 GHz like experiments in the physics labs or microwave ovens. The time of the day is unlikely to be a relevant parameter. Samples at large distance were taken after lecture and office hours and should have been less disturbed by background traffic.

The effects of other wireless traffic on our measurements is further illustrated in the following three figures.

Figure 5 shows as dots the background traffic that has been observed by our monitoring, the lines show the throughput of our own packets. Note that different metrics are shown for both types of traffic. Our data consisted of packets of 1500 bytes while the background packets were much smaller. The packets of 610 bytes are the ICMP messages the receiver sends to the sender indicating that the destination port has been closed and the 10 byte packets are WLAN beacons of all networks. The packets of 610 bytes could be avoided in future experiments by using a fake service that would listen on the respective port and accept all datagrams. However, their number is negligible in relation with the total traffic.

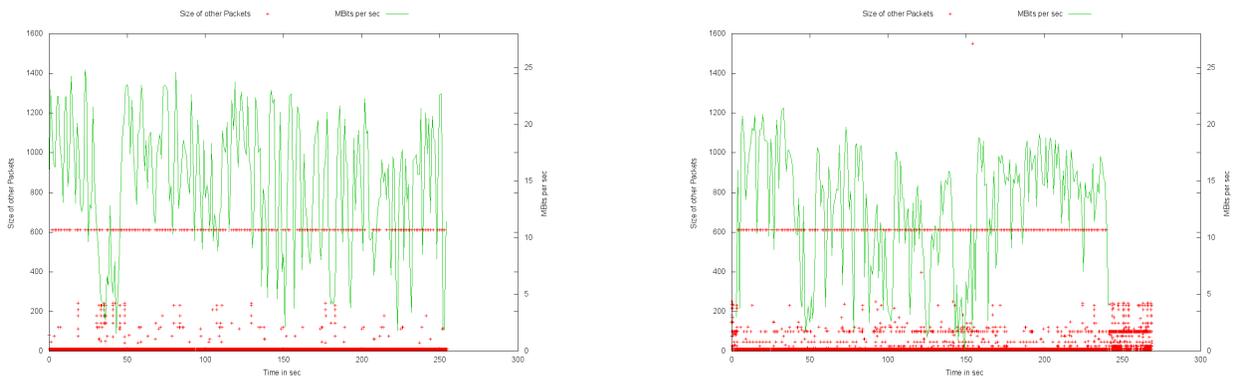


Figure 5: Visible impact of other networks on throughput, 60m (left) and 70m (right).

Both graphs in Figure 5 show a decrease in the observed throughput as more background packets are present. In the plot on the right for data at 70 meters the effect is more pronounced than in the figure on the left for a distance of 60 meters. We can conclude that our experiments did take place in an environment with disturbing wireless traffic and that this might lead to effects in the data that would have been different in a setting without noise. However, it should be noted that such an undisturbed setting would be unrealistic.

### 3.1 Packet Loss Traces and Runs

When turning our attention towards the use of experimentation results we may want to find characteristics of the data that are typical for wireless networks and can be used in simulation studies to reproduce the

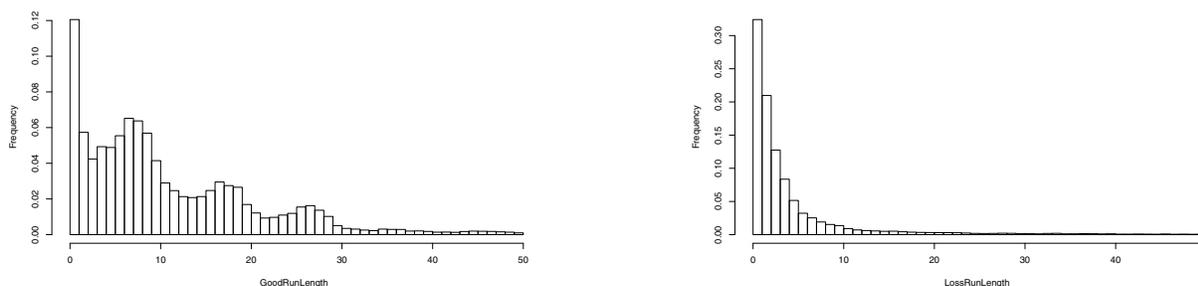


Figure 6: Histogram of success run lengths (10m, left) and loss run length (180m, right).

behavior of the networks. While in the previous subsection we have described some of our data in terms of loss probability and throughput, in this subsection we want to explore the probabilistic characteristics, such as burstiness. To this end we study the binary sequences of successfully transmitted or lost packets. This binary trace or time series (Konrad, Zhao, Joseph, and Ludwig 2003; Yajnik, Moon, Kurose, and Towsley 1999) can be analyzed following the approach of (Zhang, Duffield, Paxson, and Shenker 2001) by splitting the trace into sequences of successful and unsuccessful runs.

In the following, a success run or, equivalently, loss-free period is defined as a sequence of successfully transmitted packets that is situated in the trace between at least one unsuccessfully transmitted packet before and after the success run. Loss runs or lossy periods are defined analogously.

Figure 6 (left) shows a histogram of successful run lengths for the data set where the distance between sniffer and sender equals 10 meters. Most success runs only have length one, i.e. single packets are transmitted successfully, while the packet before and after were lost.

Most interesting in this figure are the waves in the histogram that have decreasing local maxima approximately at multiples of 8. While we do not have a complete explanation for these oscillations, one potential reason might be the buffer sizes in the access point. The run length distribution clearly exhibits high variability, truncated at length 50 in the figure.

The empirical loss run length is shown in the histogram in Figure 6 (right). The loss run lengths do not show local maxima patterns and instead have an exponential decay indicating the suitability of a simple model as we will show in the following section.

Before turning our attention to the loss models a final step in descriptive data analysis shows the evolution of the histograms of loss-free and lossy periods across all distances in Figure 7.

The left graph shows the loss free periods where some distances show the previously mentioned wave patterns while other distances do not. As a rough description one may say that for short distances loss-free periods of length one are much less frequent than for the long distances. For long loss-free periods the opposite holds.

The lossy periods, shown on the right in Figure 7 are mostly very short and at higher distances the longer periods are slightly more frequent.

#### 4 FAULT IMPACT MODELS

Fault-impact models model the impact of underlying faults on the behavior of a system component. These models can then be used to introduce the effects of these faults into a model where other systems depend on this component (cf. (Reinecke, Wolter, and Malek 2010)). In the scenario considered here, we want to derive models for the characteristics of lossy and loss-free periods of packet transmission in a wireless LAN environment.

A very simple kind of fault-impact model is the Bernoulli trial, where each packet is dropped with a constant probability  $h$ . As loss probabilities may vary over the course of time, models with several

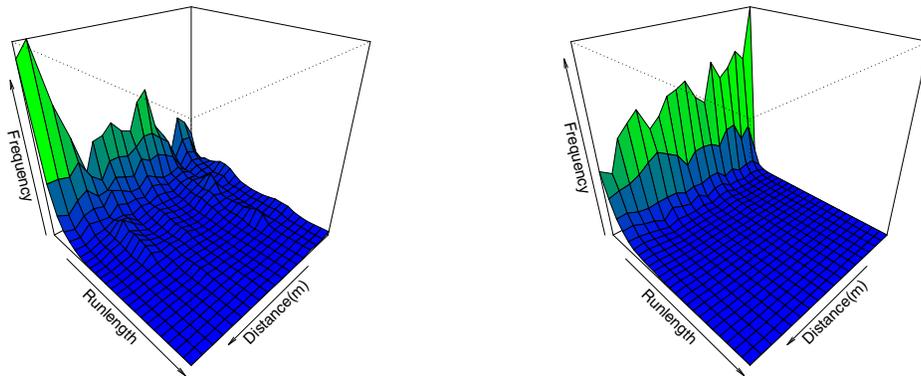


Figure 7: 3d plot of loss-free (left) loss-run lengths (right) over all distances (day 1).

states, each with a different loss probability, have been proposed. In the following, we first describe the two-state Gilbert-Elliot model. In Gilbert-Elliot models, state-sojourn times are characterized by geometric or exponential distributions. Taking a more general view, sojourn times may also be described by other distributions. We discuss the case where phase-type distributions are used to characterize state transitions. Since phase-type distributions have Markovian representations, such models may also be seen as extended Gilbert-Elliot models with more than two states.

Other models that may be considered are e.g. the Markovian Arrival Processes and the Markov-Modulated Poisson Process (see e.g. (Horváth and Telek 2002)).

#### 4.1 Two-State Gilbert-Elliot Models

Gilbert-Elliott loss models have been used to describe packet loss in the Internet (e.g. in (Zhang, Paxson, and Shenker 2000; Zhang, Duffield, Paxson, and Shenker 2001; Hohlfeld, Geib, and Haßlinger 2008)). A Gilbert-Elliot (GE) model is a Markov chain where each state is assigned a different loss probability. Such models can capture periods of elevated loss as well as loss runs better than simple Bernoulli models, where the packet loss is described only by a single loss probability for each packet (Zhang, Paxson, and Shenker 2000; Zhang, Duffield, Paxson, and Shenker 2001).

In Gilbert-Elliot models, both the state transitions and the loss probabilities for each state must be specified. The literature commonly defines GE models as Discrete-Time Markov Chains where state transitions are specified by probabilities. In (Reinecke and Wolter 2011) we argued that Continuous-Time Markov Chains may be more appropriate. We will discuss both the discrete-time and the continuous-time Gilbert model in the following.

##### 4.1.1 Discrete-Time Gilbert Model

With discrete-time Gilbert models, state transitions happen at packet arrival instances, that is, every time a packet arrives, the next state is chosen based on the successor-state probabilities. A two-state Gilbert-Elliot model is thus defined by two tuples of parameters: State transitions are described by the probabilities  $p, q \in [0, 1]$ . Loss probabilities for the good and the bad state are given by  $h_G, h_B \in [0, 1]$ . In a discrete-time Gilbert model, state-sojourn times are measured by packet arrivals, i.e. the good and the bad state model the numbers of successive packet transmissions or packet drops, respectively. Since transitions depend on

Bernoulli trials, state sojourn times follow Geometric distributions with support  $k = 1, 2, \dots$  and parameter  $1 - p$  and  $1 - q$ , respectively. The mean sojourn times are thus  $m_G = 1/(1-p)$  and  $m_B = 1/(1-q)$ .

#### 4.1.2 Continuous-Time Gilbert Model

In the discrete-time Gilbert model, the lengths of lossy and loss-free periods are governed by the packet arrival process. As shown in (Reinecke and Wolter 2011), this type of model is inappropriate for scenarios where the arrival process adapts to the loss process. In particular, with TCP senders throughput decreases dramatically, as TCP interprets packet loss as an indicator for overload and reduces its sending rate, thus effectively increasing the length of the loss period.

In this situation continuous-time models may be more appropriate. In these models, state-sojourn times for the good state and the bad state are described by exponential distributions with rates  $\lambda, \mu > 0$ . Each time the model enters a state, a new sojourn time is drawn from the exponential distribution. After this time has passed, the model transitions to the other state. State transitions are then completely independent of packet arrivals. As with the discrete-time model, packet loss probabilities are given by  $h_G$  and  $h_B$ .

### 4.2 Phase-type Models

Gilbert-Elliot models, as described so far, consist of two states with different loss probabilities. In these models, sojourn times (and thus period lengths) follow geometric or exponential distributions, depending on whether the model is in discrete time or in continuous time. As we have observed in Section 3, these simple distributions may not be sufficient to describe real-world measurements. This issue can be addressed by using a larger class of distributions for the state sojourn times. In the following we consider phase-type distributions for this purpose.

#### 4.2.1 Phase-type Distributions

Phase-type (PH) distributions represent the time to absorption in a Markov chain with one absorbing state (Neuts 1981). A discrete-time phase-type (DPH) distribution is defined by a tuple  $(\alpha, \mathbf{P})$  of a row vector  $\alpha$  containing initial probabilities and a transition probability matrix  $\mathbf{P}$ . Similarly, a continuous-time phase-type (CPH) distribution is defined by an initial probability vector  $\alpha$  and a sub-generator matrix  $\mathbf{Q}$ . Phase-type distributions are particularly attractive for modeling and simulation since they can be fitted easily to empirical distributions and enable efficient algorithms for random-variate generation. Furthermore, due to their interpretation as a Markov chain, PH distributions can be employed in analytical methods to validate and complement simulation studies. Due to space constraints, we refer to e.g. (Neuts 1981; Horváth and Telek 2002) and the references therein for more details on PH distributions.

#### 4.2.2 Extended Gilbert-Elliot Models Using PH Distributions

Extension of Gilbert-Elliot models using phase-type distributions is straightforward: Sojourn times in the good state and the bad state are described by two PH distributions  $F_G, F_B$ , while packet loss probabilities are again given by  $h_G, h_B$ . If  $F_G, F_B$  are discrete-time PH distributions, the length of each state sojourn time is measured in packet arrivals; if  $F_G, F_B$  are continuous-time PH distributions, the state sojourn times are independent of the arrival process and measured in wall-clock time.

## 5 FAULT-IMPACT MODELS FOR MOBILE COMMUNICATION

We fitted Gilbert-Elliot and phase-type models to the measurements described in Section 3. We used both discrete-time and continuous-time models, in order to obtain appropriate models for a wide range of application scenarios. We provide discrete-time Gilbert-Elliott models and discrete-time and continuous-time phase-type models. In this section we first describe the methods we used for fitting the data before giving a detailed discussion and evaluation of the fitted models.

### 5.1 Fitting Methods

We applied two approaches for fitting discrete-time Gilbert-Elliot models to the data. First, we assigned loss probabilities  $h_G = 0$  and  $h_B = 1$  and computed  $p_{12}$  and  $p_{21}$  such that mean sojourn times,  $m_G$  and  $m_B$ , fit the mean length of the loss-free and the loss runs, respectively. Second, we used the Baum-Welch algorithm (Baum, Petrie, Soules, and Weiss 1970), which also adjusted the loss probabilities based on the traces.

In order to fit discrete-time phase-type models, we used the PhFit tool (Horváth and Telek 2002) to fit the run lengths using a discrete-time phase-type distribution of size  $n = 25$ . The continuous-time phase-type models were obtained by fitting Hyper-Erlang distributions to the run length samples using the HyperStar tool (Reinecke, Krauß, and Wolter 2012). Note that for the continuous-time model we assume that one packet was sent per unit of time.

### 5.2 Results

In the following we present our models. We discuss the models for distances 40m and 100m in detail, as these traces can serve as examples for the two classes of traces we observed. With distance 40m we observed waves in the distribution of the loss-free run lengths, while with distance 100m no such waves occurred.

Table 1: Model Parameters for the Gilbert-Elliot models.

Distance (m)	Fitting Method	$h_G$	$h_B$	$p_{12}$	$p_{21}$
40	Matching mean	0	1	0.09	0.40
40	Baum-Welch	0.12	$1 - 5.6 \times 10^{-6}$	$4.3 \times 10^{-4}$	$6.5 \times 10^{-4}$
100	Matching mean	0	1	0.19	0.35
100	Baum-Welch	0.12	0.72	0.08	0.23

Table 1 shows the parameters for the Gilbert-Elliot models, while Table 2 gives the Hyper-Erlang models returned by Hyper-\*. Due to space constraints, we omit the parameters for the DPH models here. All models will be made available at <http://cst.mi.fu-berlin.de/staff/wolter.html>.

Table 2: Hyper-Erlang Model Parameters. The parameters are specified as tuples of the branch probability  $\beta$ , branch length  $b$ , and rate  $\lambda$  for each Erlang branch.

Distance (m)	run	Hyper-Erlang Distribution $[(\beta, b, \lambda)]$
40	good	$(0.08, 99, 3.87), (0.31, 3, 1.13), (0.35, 26, 2.97), (0.23, 69, 4.27), (0.02, 55, 1.34)$
40	bad	$(0.09, 2, 0.45), (0.91, 9, 6.72)$
100	good	$(0.61, 4, 2.00), (0.39, 5, 0.47)$
100	bad	$(0.23, 2, 0.42), (0.77, 9, 7.14)$

### 5.3 Evaluation

We evaluate the models as follows. For the discrete-time models we performed simulations of the models, by simply playing the model. We compare the resulting traces to our measurement traces. We evaluate the continuous-time model by comparing the theoretical density of the fitted distribution to the empirical density. In Figure 8 we plot the histograms of the measured run lengths for loss-free and lossy periods, respectively, at a distance of 40 meters, and the densities of the run-lengths generated of the fitted models. Observe that, while the simple Gilbert models fit the density of the loss-run lengths well, they cannot capture the oscillations present in the histogram of the loss-free runs. These can be fitted by both PH

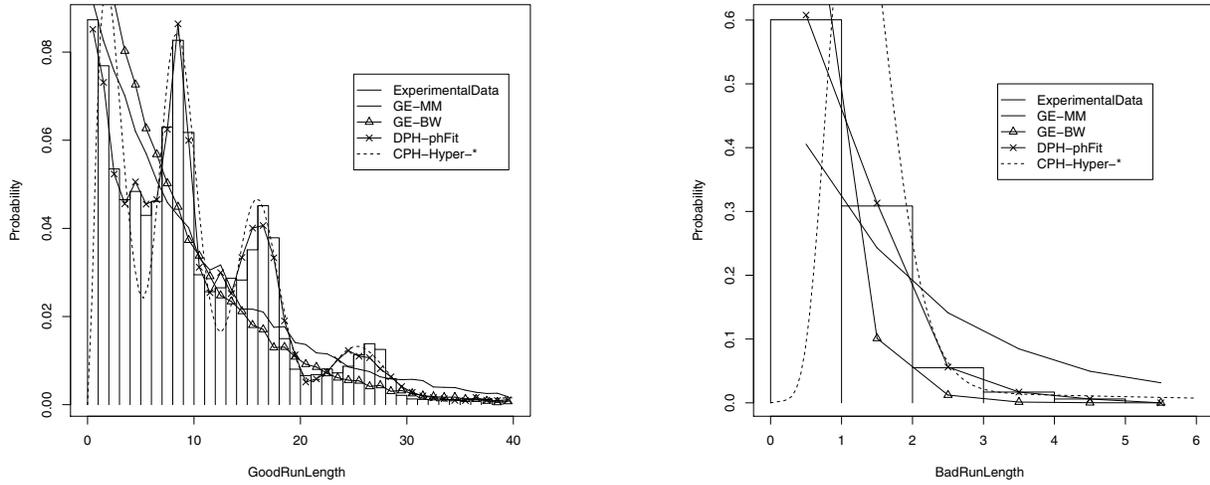


Figure 8: Histograms of measured loss-free period lengths (left) and loss period lengths (right), and model densities (distance 40 meters).

models. Although both PH models fit the peaks of the oscillations well, the DPH model fits the valleys better.

Concerning the run-lengths at a distance of 100 meters (Figure 9), we observe that all models fit the length of loss periods well. Furthermore, in Figure 9 we see that the shape of the density is well-captured by the Gilbert models. However, although this data set does not exhibit pronounced oscillations, there is a slight peak around 10, which the PH models capture better than the Gilbert models.

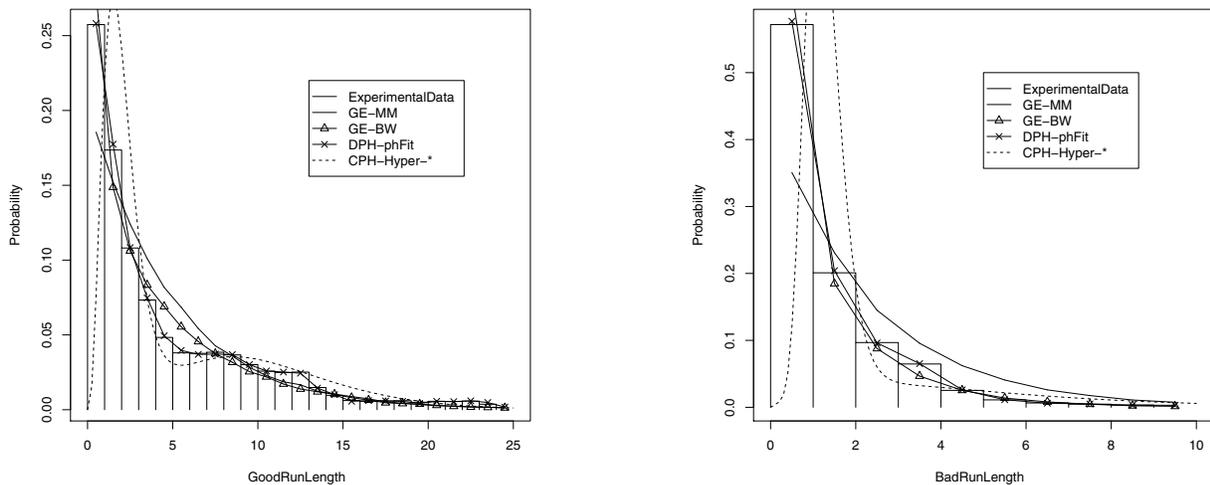


Figure 9: Histograms of measured loss-free period lengths (left) and loss period lengths (right), and model densities (distance 100 meters).

Our evaluation shows that the models can reproduce the characteristics of the loss-free and lossy periods well. If the density of loss-free periods exhibits oscillations, as observed in some of our data sets, phase-type models may capture the behavior better than the simple Gilbert models, but at the cost of having a larger

model. Discrete PH models appear to have a slight advantage over continuous-time models, which may be due to the fact that the measurements are discrete, since they model period lengths in packets.

## 6 CONCLUSIONS

Wireless communication is known to be loss-prone and even more so as the communicating parties move or move away from each other. To approach the analysis of loss characteristics as nodes move we have conducted a series of experiments of data transmission with increasing distance between the transmitting parties and a third node. This allowed us to observe degradation in the communication which could not be seen when moving sender and receiver. In the latter case the connection can either be established or not and little variation in the loss rate can be seen.

We have shown experiments for distances between 10 and 180 meters resulting in data loss rate between 13% and 85%. The high loss rates can be partially attributed to the fact that MAC layer retransmissions were disabled and the UDP protocol was used for packet transmission. Nonetheless, in future experiments the distance should probably be reduced and short distances should be investigated at finer granularity.

From our experiments it can be seen that the loss characteristics in wireless communication in a normal academic environment changes with the distance between the parties. Transmission quality further is impaired by numerous effects that remain unknown but will always exist.

We have divided the data into sequences of lost packets and sequences of successfully received packets and have fitted several models to those runs. In this paper we provide a set of models for both classes of runs that can easily be used for fault-injection in a test bed or a simulator. We find that very simple Gilbert-Eliot models fit the loss runs very well, while the loss-free runs exhibit more diverse behavior and are therefore better expressed with more complex phase-type models. In practice, the choice of model should be guided both by the quality of the models and the evaluation scenario. For evaluation scenarios where the arrival process is independent of the loss process, or where dependence between the arrival process and the loss process is desired, we recommend the use of Gilbert models for the loss periods and DPH models for the loss-free periods. These models fit our measurements very well, but imply that the loss process depends on the arrival process. Our continuous-time models avoid this dependence and may therefore be better suited for evaluations where the loss process should be independent of the arrival process. However, these models are larger than the corresponding DPH models.

## 7 ACKNOWLEDGMENTS

This work was supported by the German research council under DFG grants Wo 898/3-1 and Wo 898/5-1.

## References

- Aircrack-ng 2012, March. "Aircrack-ng". <http://www.aircrack-ng.org/>.
- Baum, L. E., T. Petrie, G. Soules, and N. Weiss. 1970. "A Maximization Technique Occurring in the Statistical Analysis of Probabilistic Functions of Markov Chains". *The Annals of Mathematical Statistics* 41 (1): 164–171.
- Eckhardt, D., and P. Steenkiste. 1996, August. "Measurement and Analysis of the Error Characteristics of an In-Building Wireless Network". *SIGCOMM Comput. Commun. Rev.* 26 (4): 243–254.
- Haßlinger, G., and O. Hohlfeld. 2008. "The Gilbert-Elliott Model for Packet Loss in Real Time Services on the Internet". In *MMB 2008*, 269–286.
- Henderson, T., and D. Kotz. 2006. *Mobile, Wireless, and Sensor Networks: Technology, Applications, and Future Directions*, Chapter Measuring Wireless LANs, 5–29. John Wiley & Sons, Inc.
- Hohlfeld, O., R. Geib, and G. Haßlinger. 2008, June. "Packet Loss in Real-Time Services: Markovian Models Generating QoE Impairments". In *Proc. of the 16th International Workshop on Quality of Service (IWQoS)*, 239–248.

- Horváth, A., and M. Telek. 2002. "Markovian Modeling of Real Data Traffic: Heuristic Phase Type and MAP Fitting of Heavy Tailed and Fractal Like Samples". In *Performance Evaluation of Complex Systems: Techniques and Tools*, edited by M. Calzarossa and S. Tucci, Volume 2459 of *Lecture Notes in Computer Science*, 267–282. Springer Berlin / Heidelberg.
- Horváth, A., and M. Telek. 2002. "PhFit: A General Phase-Type Fitting Tool". In *TOOLS '02: Proceedings of the 12th International Conference on Computer Performance Evaluation, Modelling Techniques and Tools*, 82–91. London, UK: Springer-Verlag.
- Konrad, A., B. Y. Zhao, A. D. Joseph, and R. Ludwig. 2003, May. "A Markov-based channel model algorithm for wireless networks". *Wirel. Netw.* 9 (3): 189–199.
- Jouni Malinen 2012, March. "hostapd: IEEE 802.11 AP, IEEE 802.1X/WPA/WPA2/EAP/RADIUS Authenticator". <http://hostap.epitest.fi/hostapd/>.
- Neuts, M. F. 1981. *Matrix-Geometric Solutions in Stochastic Models. An Algorithmic Approach*. New York: Dover Publications, Inc.
- R Development Core Team 2008. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0.
- Reinecke, P., T. Krauß, and K. Wolter. 2012. "Cluster-Based Fitting of Phase-Type Distributions to Empirical Data". *Computers & Mathematics with Applications* (0).
- Reinecke, P., and K. Wolter. 2011, October. "On Stochastic Fault-Injection for IP-Packet Loss Emulation". In *Computer Performance Engineering. Proceedings of the 8th European Performance Engineering Workshop, EPEW 2011*, edited by N. Thomas, Number 6977 in LNCS: Springer.
- Reinecke, P., K. Wolter, and M. Malek. 2010, February. "A Survey on Fault-Models for QoS Studies of Service-Oriented Systems". Technical Report B-2010-02, Freie Universität Berlin.
- Tang, D., and M. Baker. 2000. "Analysis of a local-area wireless network". In *Proceedings of the 6th annual international conference on Mobile computing and networking*, MobiCom '00, 1–10. New York, NY, USA: ACM.
- Tcpreplay 2012, March. "Tcpreplay". <http://tcpreplay.synfin.net/>.
- Yajnik, M., S. B. Moon, J. F. Kurose, and D. F. Towsley. 1999. "Measurement and Modeling of the Temporal Dependence in Packet Loss". In *Eighteenth Annual Joint Conference of the IEEE Computer and Communications Societies*, Volume 1 of *INFOCOM '99*, 345–352. New York, NY, USA: IEEE Computer Society.
- Zhang, Y., N. Duffield, V. Paxson, and S. Shenker. 2001. "On the Constancy of Internet Path Properties". In *IMW '01: Proceedings of the 1st ACM SIGCOMM Workshop on Internet Measurement*, 197–211. New York, NY, USA: ACM.
- Zhang, Y., V. Paxson, and S. Shenker. 2000. "The Stationarity of Internet Path Properties: Routing, Loss, and Throughput". *ACIRI Technical Report*.

## AUTHOR BIOGRAPHIES

**KATINKA WOLTER** received her PhD from Technical University Berlin and her habilitation degree from Humboldt-University Berlin. Before joining Newcastle University she has been senior researcher and guest professor at Freie Universität Berlin from 2009 until 2012. Her interests are in model-based dependability analysis of complex computing systems and networks. Her email address is [katinka.wolter@ncl.ac.uk](mailto:katinka.wolter@ncl.ac.uk).

**PHILIPP REINECKE** joined Hewlett-Packard Labs, Bristol in May 2012 after working as a researcher and PhD student at Freie Universität Berlin from 2009 until 2012. He graduated from Humboldt-University Berlin in 2007 and worked there as researcher until 2009. His email address is [philipp.reinecke@hp.com](mailto:philipp.reinecke@hp.com).

**TILMAN KRAUSS, DANIEL HAPP, FLORIAN EITEL** are Masters' students at Freie Universität Berlin. Their email addresses are [tilman.krauss@fu-berlin.de](mailto:tilman.krauss@fu-berlin.de) | [daniel.happ@fu-berlin.de](mailto:daniel.happ@fu-berlin.de) | [florian.eitel@fu-berlin.de](mailto:florian.eitel@fu-berlin.de).