A Black Box Measurement Method for Reliability of Wireless Communication

Matthias Herlich, Christian Maier, Peter Dorfinger
<firstname>.<lastname>@salzburgresearch.at

ABSTRACT

In the future, connected vehicles and automated factories will use wireless communication. To implement these applications, wireless networks must be reliable. Because the cost of communication failures in such applications is high, reliability should not only be estimated in advance, but also measured directly in practice. To date, no method exists that is able to determine, if a wireless network is 99.999 % reliable.

To address this need, we propose a black box test that uses only the success/fail status of transmissions to determine the reliability of wireless communication. The test is based on two assumptions: (1) reliability is time-invariant and (2) transmissions are statistically independent. Because it is crucial for any measurement method that its assumptions hold, we test both assumptions as part of the method.

The proposed method is especially suited when access to lower layer information is limited to the information returned by off-the-shelf hardware. The method measures reliability of wireless networks without support and knowledge from the network operator and administrator. In the future, methods such as the proposed one are needed to ensure reliable operation of wireless networks in critical scenarios.

CCS CONCEPTS

• **Networks** → **Network reliability**; **Network measurement**; *Wireless access networks*;

KEYWORDS

Reliability, Measurement, Wireless, Black box

ACM Reference Format:
Matthias Herlich, Christian Maier, Peter Dorfinger. 2018. A Black Box Measurement Method for Reliability of Wireless Communication. In *EM-5G '18: Workshop on Experimentation and Measurements in 5G, December 4, 2018, Heraklion, Greece.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3286680.3286681

1 INTRODUCTION

Wireless communication has become ubiquitous in personal communication. Trends such as Industry 4.0 and connected vehicles increasingly also use wireless communication. In contrast to personal communication these professional applications demand high reliability.

In our earlier work [6] we have determined two reference scenarios, which are highly dependent on reliable wireless communication: (1) vehicles transmitting collision warnings at urban intersections and (2) wireless emergency-stop buttons in factories. These often require that the network has a reliability of at least 99.999 % [6]. Because lost transmissions in these cases are expensive and can potentially result in deaths, reliability of the wireless communication in these scenarios should be tested.

The methods to determine reliability can be grouped into white box and black box approaches. Black box approaches have knowledge only about the result of the transmissions (success or failure), but not about the lower layers (e.g., signal strength). In contrast, white box approaches have access to this internal knowledge.

White box approaches can generate better predictions with the same number of measurements or provide the same precision with fewer measurements, because they can exploit the additionally available information. However, white box approaches need additional assumptions to interpret the additional information (e.g., the distribution function of the signal strength). When such an assumption does not hold, the results of a white box approach can be worse than those of a black box approach. For important applications both black and white box approaches should be used.

This paper describes a method to measure reliability of wireless communication. The method consists of a black box test, which needs two assumptions, as well as white box tests of these assumptions. It would be possible to also test the assumptions using black box approaches, but the number of necessary measurements would be large. Hence, we use white box tests to test the assumptions of the black box test. Figure 1 provides a simplified overview of the method.

2 RELATED WORK

Many channel models have been created that describe averages and quantiles (usually 90 % and 99 %) of wireless channel quality. However, using these models alone is not enough to determine, if a network provides reliability of 99.999 %. First, it is unclear how accurate these models are at the edges of their reliability range. Second, these models only consider the physical channel, but not higher layer effects, such as interaction with medium access control (MAC), schedulers and retransmission schemes.

Direct measurements of wireless reliability have been an area of active research (e.g., [1]) and continue to be of interest (e.g., [8]). Vehicular communication has also been analyzed passively (e.g., [10]), but not its reliability. Approaches, which measure reliability directly usually measure relatively unreliable networks (< 99 %), because currently few wireless networks provide higher reliability.

EM-5G '18, December 4, 2018, Heraklion, Greece

ACM ISBN 978-1-4503-6083-8/18/12…$15.00
https://doi.org/10.1145/3286680.3286681

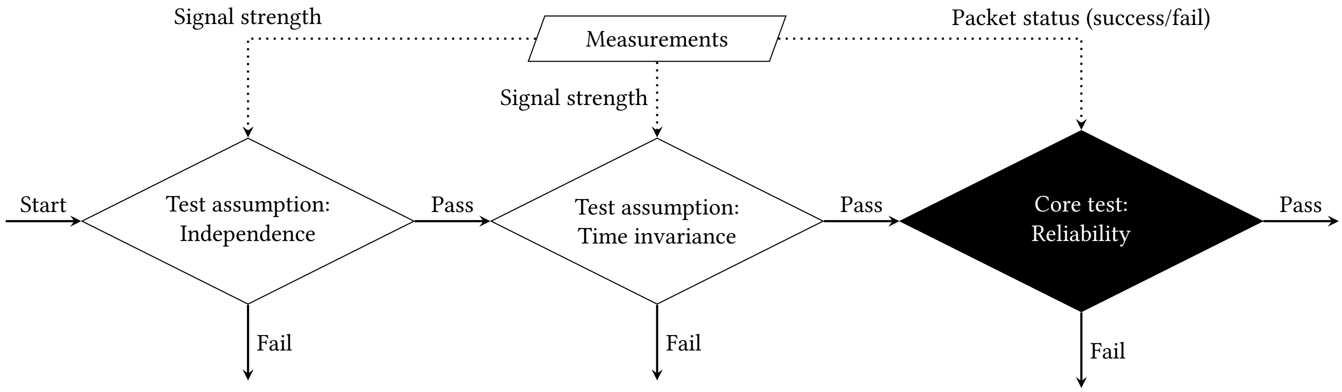

Figure 1: The simplified overview of the proposed method consists of three tests (two white box, one black box).

The proposed methods do not consider effects such as false positives and false negatives and do not test their assumptions.

Bai and Krishnan [2] analyze reliability of wireless networks for automotive applications. Most of the connections they test have lower reliability (80 % and 99 %) than we want to measure (≥ 99 %). Others, such as Ma et al. [7], use analysis and simulation to analyze safety-critical communication.

Woo and Kim [12] describe that the signal strength (RSSI or SINR) does not provide a high quality estimate of reliability. Salyers et al. [9] describe that individual hardware plays a large role in quality of a channel. Measurements for individual hardware exist (e.g., CC1000 [11]), but usually lack a detailed statistical evaluation.

In summary, a detailed description of the necessary statistical methods to determine and predict reliability in wireless networks is missing. We will provide it in this paper.

3 SYSTEM MODEL

To provide a method that determines the reliability of a wireless channel, we first provide a mathematical model of reliability. To confirm the reliability of a wireless system, our model should not be based on the same assumptions that were used during the systems design. Thus, we do not build a physical model of the channel or transmission system. Our methods should not be used to replace, but to supplement, engineering based on Safety integrity level (SIL) or similar.

We assume a random process determines whether a packet is correctly transmitted or not. The status and additional data (e.g., signal strength) of a single packet transmission constitutes a measurement. We define reliability as the success probability of the random process, with the following assumptions:

(1) the success probability is constant over time and
(2) the outcome of a transmission (success or failure) does not affect the outcome of future transmissions.

Note that the assumptions do not require that the wireless channel has to be constant over time. Our assumptions describe a Bernoulli process, where we denote:

- r as the real, but unknown reliability of the system and
- t as the target reliability against which is tested.

The question whether reliability is higher than the given target threshold thus becomes to determine if the inequality $r \geq t$ holds. However, as it is impossible to measure r directly, methods are necessary to estimate it. In this paper, we will estimate reliability r from the boolean received/not received data from a successive transmission of n packets and determine if it is higher than the target reliability t.

Both assumptions will be false in some environments. To determine if they hold in the environment under measurement, we propose to test them as part of the measurement method (see Section 4.3). Usually the outcome of transmissions are dependent on short time scales, but we are interested in longer term behavior, where they are independent.

In contrast to a definition based on the packet delivery rate, our definition of reliability relies on the two assumptions. However, our definition also carries information for future transmissions and not only about the measured interval.

4 MEASUREMENT METHOD

In this section we describe a method to determine if a wireless network provides the given target reliability based on the results (success or failure) of repeated packet transmission. The method will also test the necessary assumptions. We denote the number of transmitted packets with n. For the evaluation of the results we used the software environment for statistical computing R in version 3.3.1.

We first describe a simple test that will only be useful when packet losses are extremely rare. Then we describe an extended test, which generalizes the idea of the simple test and is applicable in a wider range of scenarios.

4.1 Simple Test

We propose to test whether the real reliability r of the system is larger than or equal to the given target reliability t (positive test result) or not (negative test result). There are two possible kinds of errors:

- a positive result although the real reliability r is less than the target reliability t (false positive) and

Table 1: The number of measurements for a given target reliability t and significance level α using the simple test; measurements to ensure property $\mathcal{P}$ in bold.

		Target reliability t				
		0.9	0.99	0.999	0.9999	0.999 99
	0.1	**22**	230	2 302	23 025	230 258
	0.01	44	**459**	4 603	46 050	460 515
α	0.001	66	688	**6 905**	69 075	690 773
	0.000 1	88	917	9 206	**92 099**	921 030
	0.000 01	110	1 146	11 508	115 124	**1 151 287**

- a negative result although the real reliability r is larger than the target reliability t (false negative).

Since our primary goal is to avoid false positives (that is, the test says the network has the target reliability t, although it does not), it is necessary to keep the worst case probability of a false positive low. That is, we suggest selecting a significance level $\alpha = 1 - t$. Thus, the test should satisfy:

(*Property* $\mathcal{P}$) If the real reliability r is less than the target reliability t, the test will provide a positive result with a probability less than $1 - t$.

For example, if $t = 0.999\,99$, out of 100 000 tests of networks that do not provide 99.999 % reliability, the test returns a positive result in 1 or fewer cases on average. We wrote property $\mathcal{P}$ in such a way that it does not depend on understanding confidence intervals or p-values, which may lead to misinterpretations [5].

In the simple test, we consider the test positive if and only if all packets are transmitted successfully. To ensure the bound for the false positive rate, the probability of no failures has to be less than α if the real reliability is lower than the target reliability. With the assumptions that repeated packet transmission is a Bernoulli process with probability r for a successful transmission made in Section 3 this yields the inequality $\text{dbinom}(n, n, r) < \alpha$ for $r < t$. Here $\text{dbinom}(\,\cdot\,, n, r)$ denotes the probability mass function of the binomial distribution. Since $\text{dbinom}(n, n, r) = r^n$ this simplifies to $r^n < \alpha$. Hence a necessary and sufficient criterion to bound the false positive rate as stated is $t^n \leq \alpha$. Thus, the smallest number of necessary measurements is

$$n = \left\lceil \frac{\log(\alpha)}{\log(t)} \right\rceil,$$

where $\lceil\,\cdot\,\rceil$ is the ceiling function. Table 1 shows the number of needed measurements for the target reliability t and the significance level α.

When a system provides the number of independent measurements of successfully transmitted packets that are stated in Table 1, it provides the necessary certainty that it allowed reliable communication in the measured time interval. However, when a single failure is detected the test has to be marked as failed (to keep the false positive rate low). This leads to a high false negative rate, when the real reliability is only slightly higher than the target reliability (see Figure 2).

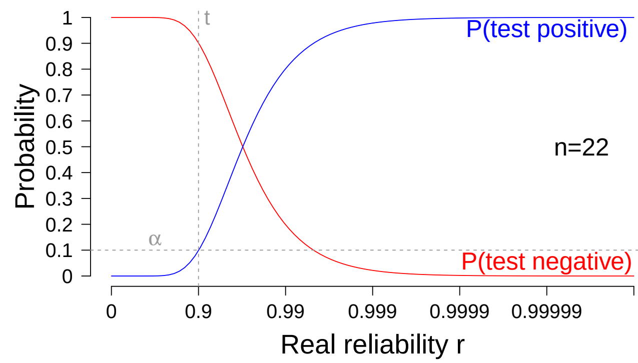

Figure 2: The probability for a positive/negative result of the test depends on the unknown real reliability r. For a real reliability lower than t the probability for a positive result is lower than α (Property $\mathcal{P}$).

In the next section we provide an extension of the simple test that puts a bound on the false negative rate as well.

4.2 Extended Test

To improve the simple test, we also limit the false negative rate in the extended test. That is, the extended test should (in addition to property $\mathcal{P}$) satisfy:

(*Property* $\mathcal{Q}$) If the real reliability of the system is higher than or equal to s (where $s > t$ is fixed), the test will provide a negative result with a probability less than $1 - t$.

To reduce the false negative rate the extended test allows a constant number of failures a before the test fails. The number of measurements n has to be increased accordingly so that the false positive rate does not increase. Directly setting the number of allowed errors is, however, not as intuitive as limiting the false negative rate. Thus, we propose to select the smallest number of measurements n and allowed failures a that fulfill both properties $\mathcal{P}$ and $\mathcal{Q}$. If we denote the cumulative distribution function of the binomial distribution with $\text{pbinom}(\,\cdot\,, n, r)$, then property $\mathcal{P}$ is equivalent to

$$1 - \text{pbinom}(n - a - 1, n, t) \leq 1 - t \tag{1}$$

and property $\mathcal{Q}$ is equivalent to

$$\text{pbinom}(n - a - 1, n, s) < 1 - t. \tag{2}$$

Algorithm 1 returns the smallest numbers n and a such that both inequalities hold. For our implementation we used the fact that for a fixed number of measurements n the left hand side of inequality 1 is strictly increasing in a and the left hand side of inequality 2 is strictly decreasing in a. It is necessary to check both inequalities for all n, since for example there exists a test for $t = 0.999$ and $s = 0.9999$ with $n = 19\,710$ measurements and $a = 7$ allowed errors, but no value of a fulfills both bounds with $n = 19\,711$ measurements. Table 2 shows some example values. We propose to arbitrarily select $s = 0.1t + 0.9$.

Table 2: The number of measurements using the extended test; in parentheses is the number of allowed errors.

		Target reliability t				
		0.9	0.99	0.999	0.999 9	0.999 99
	0.99	38 (1)	- (-)	- (-)	- (-)	- (-)
	0.999	22 (0)	1 157 (4)	- (-)	- (-)	- (-)
s	0.999 9	22 (0)	662 (1)	19 620 (7)	- (-)	- (-)
	0.999 99	22 (0)	459 (0)	11 225 (2)	293 056 (11)	- (-)
	0.999 999	22 (0)	459 (0)	9 230 (1)	159 132 (3)	3 751 160 (14)

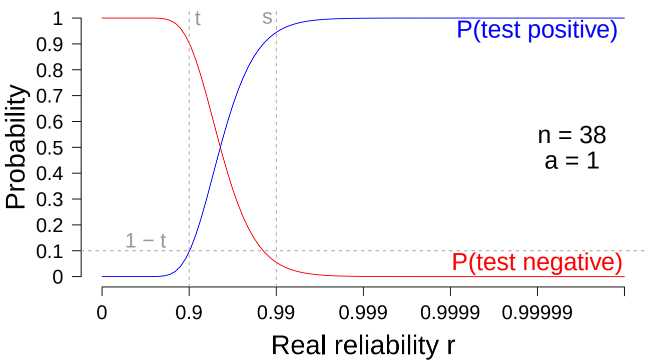

Figure 3: In contrast to the simple test, the extended test has bounds for both false positives and false negatives. For a real reliability higher than s the probability for a negative test is lower than $1 - t$ (property Q).

Algorithm 1 Determine number of measurements n and allowed errors a from target reliability t and s (extended test)

$n = 0$
fpr $= 1$ ▷ false positive rate
while fpr $> 1 - t$ **do**
  $n = n + 1$
  $a = -1$
  fnr $= 1$ ▷ false negative rate
  while fnr $\geq 1 - t$ **do**
    $a = a + 1$
    fnr $=$ pbinom$(n - a - 1, n, s)$
  end while
  fpr $= 1 -$ pbinom$(n - a - 1, n, t)$
end while

In summary the extended test has

- a high probability to give a negative result, if the real reliability r is below t and
- a high probability to give a positive result, if the real reliability r is above s.

For a real reliability between t and s the bound on the false negative rate does not hold. Figure 3 illustrates this. The interval between t and s can in principle be arbitrarily small. However, the smaller the interval, the more measurements are necessary.

4.3 Testing the Assumptions

In Section 3 we made two assumptions. Stated in precise terms, the sequence of n successive packet transmissions $X_1, \ldots, X_n$, which represent the result of each transmission (success or failure), must be independent and identically distributed (i.i.d.). These assumptions depend on the time between successive measurements, but can also be influenced by changes in the environment. Hence, it is necessary to test if both assumptions hold. Next, we describe two such tests.

As the systems under test will have few losses, it will not be possible to check the assumptions by considering only the single boolean received/not-received information of each packet. Therefore, we propose a white box approach: Let $Y_1, \ldots, Y_n$ denote the measured signal strengths captured during the transmission of n packets. If the sequence $X_1, \ldots, X_n$ is i.i.d., then the sequence $Y_1, \ldots, Y_n$ will also be i.i.d. The converse is not always true, since factors not covered by signal stregth may impact packet transmissions. However, because the i.i.d. property of $Y_1, \ldots, Y_n$ should be a strong indicator for the i.i.d. property of $X_1, \ldots, X_n$, we will use Y_i as a proxy for X_i. Since advanced i.i.d. tests usually rely on sophisticated techniques (such as [4]), we use a simple test based on computing the (auto)-correlation. Lost packets do not provide a signal strength, but as only few packets will usually be lost anyway, we still consider the sampling of signal strengths as representative.

To specify our test, let $\overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$ denote the mean of the measured values. For any positive integer k with $k < n$ we compute the autocorrelation coefficient of lag k [3]:

$$r_k = \frac{\sum_{i=1}^{n-k}\left(Y_i - \overline{Y}\right)\left(Y_{i+k} - \overline{Y}\right)}{\sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2}, \tag{3}$$

which is a real number with $|r_k| \leq 1$. These coefficients measure the correlation of the sequence $Y_1, \ldots, Y_{n-k}$ with the shifted sequence $Y_{k+1}, \ldots, Y_n$. For the evaluation of the autocorrelation coefficients we compute the bound

$$B = z_{1-\frac{\alpha}{2}} \cdot \frac{1}{\sqrt{n}},$$

where α denotes the significance level of the test (we arbitrarily propose to use $\alpha = 0.05$) and $z_{1-\frac{\alpha}{2}}$ denotes the $(1 - \frac{\alpha}{2})$-quantile of the normal distribution, which can be calculated with the function qnorm($\cdot$) in R. An approximate calculation [3] shows that if the sequence $Y_1, \ldots, Y_n$ of random variables is i.i.d., then $[-B, B]$ is a confidence interval of confidence level $(1 - \alpha)$ for each correlation coefficient r_k.

4.3.1 Independence of Successive Transmissions. To test assumption 2, we consider only the autocorrelation coefficient of lag 1. Hence, if $|r_1| > B$, the hypothesis that successive values Y_i and Y_{i+1} are independent is rejected with an approximative significance level α. If $|r_1| \leq B$, the hypothesis is accepted. Strictly, the test tests if successive signal strength values are uncorrelated, which is in general not equivalent to being independent. Nevertheless, we consider a test for correlation enough, because we deem other types of dependence unlikely.

4.3.2 Time-invariance of Reliability. To test assumption 1, in theory one has to consider the autocorrelation coefficients r_k for all lags k, but it is general practice [3] to consider only r_k with $k \leq \frac{n}{4}$. Let N be the number of coefficients that lie outside of the interval $[-B, B]$. If $N > \frac{\alpha n}{4}$, the hypothesis that the sequence $Y_1, \ldots, Y_n$ is stationary is rejected. Otherwise, the hypothesis is accepted. Strictly speaking the test only determines if the mean of the distribution changes and not if the variance changes. Table 3 summarizes which assumptions are needed and which are tested.

4.4 Method Summary

We propose to assemble the complete method to test reliability of a wireless network from the individual tests as illustrated in Figure 4. Table 4 summarizes the advantages and disadvantages of the method.

When using the method to test multiple locations it will be necessary to include a correction for multiple tests (e.g., the Bonferroni correction). Also, note that the method measures reliability only in the considered time period and using it to predict reliability in the future should be handled with care.

5 EXAMPLE

This example demonstrates how our method can be applied to determine the reliability of a wireless channel. It is no validation of the method. To validate the method it would be necessary to test the accuracy of the determined reliability against some other link quality metric.

To demonstrate how our measurement method works in practice, we created a WiFi setup in our lab and tested against the target reliability of $t = 99\,\%$ and chose $s = 0.999$ as proposed. While the proposed method is able to measure higher reliability, the wireless network and the environment we tested do not provide higher reliability. We determined this by running our test for $t = 99.9\,\%$, which it repeatedly failed. The test for $t = 99\,\%$ was repeatedly passed. However, the results depend on placement of the hardware and activity in the building.

This example is not intended to prove that the method can be applied in all scenarios, but to illustrate how it can be applied. As the setup of our WiFi is irrelevant to the main contribution of our paper, we do not describe it in more detail. However, the potentially long time to execute the measurements might lead to the test aborting due to non-stationary reliability in many networks.

According to Table 2, we sent $n = 1157$ packets with a constant time interval of 10 seconds between successive packets. To determine a time interval which passes the independence test, we ran a pre-test. However, we did not try to minimize this time interval. For longer measurements it will be necessary to determine a time

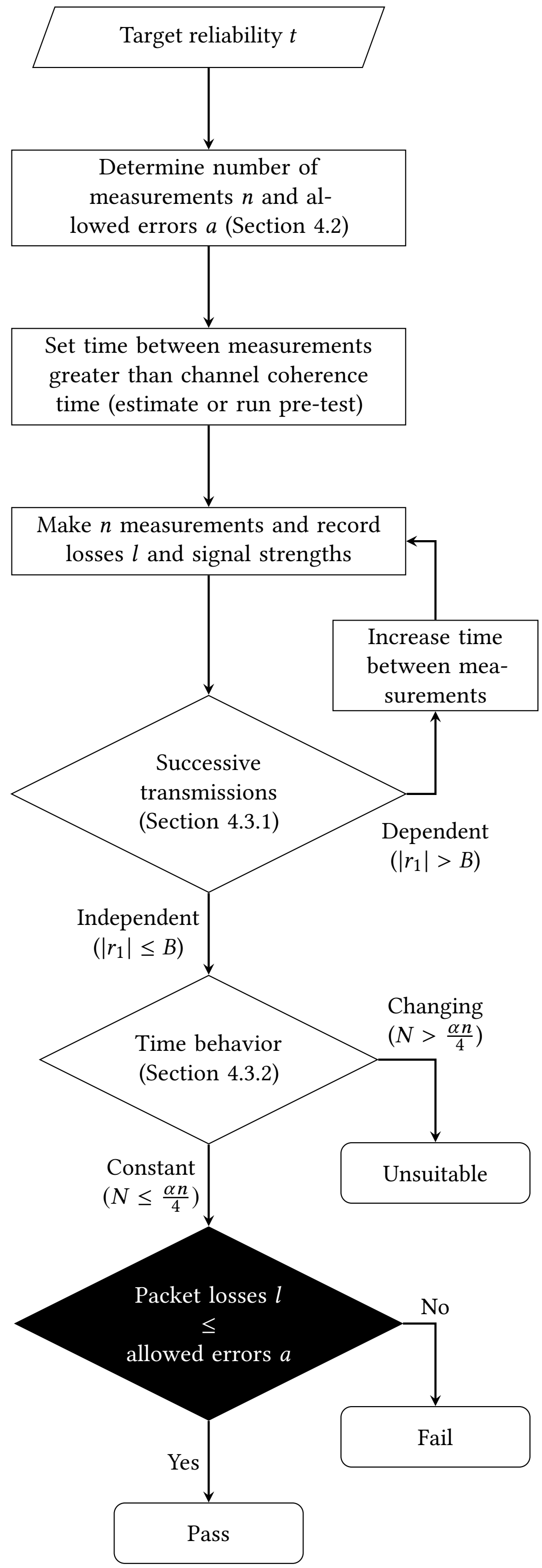

Figure 4: The measurement method is a combination of the individual tests.

Table 3: Assumptions of the black box test and what the white box tests in Figure 4 actually test based on transmission status (success/fail) X_i and signal strength Y_i.

	Assumption of test	Assumption implies	Method tests
1.	Identical distribution of $X_1, \ldots, X_n$	Identical distribution of $Y_1, \ldots, Y_n$	Constant mean of $Y_1, \ldots, Y_n$
2.	Independence of $X_1, \ldots, X_n$	Independence of $Y_1, \ldots, Y_n$	Non-correlation of Y_i and Y_{i+1}

Table 4: Advantages and disadvantages of the method shown in Figure 4

Advantages	Disadvantages
• Works with off-the-shelf hardware	• Many measurements needed
• No need to know technical details of transmission	• Hard to confirm reliability slightly higher than target
• Checking of statistical assumptions	• Statistical checks not exhaustive
• Low false positive and false negative rate	

interval that is as low as possible to reduce the overall length of the measurement. This, however, is not in the scope of this paper. Whether a longer measurement will be able to pass the constant time behavior will have to be tested in practical environments. The same holds for effects of schedulers of cellular networks.

A significance level $\alpha = 0.05$ for the i.i.d. tests yields the bound $B \approx 0.058$. The computed autocorrelation coefficient $r_1 \approx 0.049$ satisfies the inequality $|r_1| \leq B$, hence the measurement passed the independence test. Since $\frac{n}{4} = 289.25$, the relevant correlation coefficients for the stationarity test are $r_1, \ldots, r_{289}$. From these coefficients, $N = 8$ were not contained in the interval $[-B, B]$. Since $N < \frac{\alpha n}{4} \approx 14.46$, we accepted the i.i.d. hypothesis on the measured sequence of signal strength values. Finally, if 4 or fewer packets have been lost, the proposed method verifies that the wireless network provides reliability of 99 % at the measured location. From 100 locations that are not reliable, the test will return such a result on average in fewer than one location.

6 CONCLUSION

The proposed method measures reliability of wireless transmissions. The black box test does not use lower layer information, but only the success/failure status of individual packet transmissions. Additionally, two tests for the assumptions, which use lower layer information (signal strength), are part of the complete method. The proposed method is suited for an independent organization to measure reliability or when detailed network knowledge is not accessible. Its main advantage over other approaches is that it explicitly describes false positive and false negative rates as well as tests its assumptions.

In the future we will make measurements in real scenarios to determine if the use of more complex methods to test the i.i.d. properties are necessary or removal of the assumptions is possible. We will also expand the method to use different models (e.g., the Gilbert–Elliott model), that result in time-variant reliability. Additionally, we will consider how the timing of the sending device influences reliability.

Acknowledgments. This research is partly funded by the Austrian Federal Ministry of Transport, Innovation and Technology (BMVIT) and the Austrian state Salzburg.

REFERENCES

[1] D. Aguayo, J. Bicket, S. Biswas, G. Judd, and R. Morris. 2004. Link-level measurements from an 802.11 b mesh network. In *ACM SIGCOMM Computer Communication Review*, Vol. 34. ACM, 121–132.
[2] F. Bai and H. Krishnan. 2006. Reliability analysis of DSRC wireless communication for vehicle safety applications. In *Intelligent Transportation Systems Conference*. IEEE, 355–362.
[3] C. Chatfield. 2016. *The Analysis of Time Series: An Introduction, Sixth Edition*. CRC Press.
[4] J. S. Cho and H. White. 2011. Generalized runs tests for the IID hypothesis. *Journal of econometrics* 162, 2 (2011), 326–344.
[5] S. Greenland, S. J. Senn, K. J. Rothman, J. B. Carlin, C. Poole, S. N. Goodman, and D. G. Altman. 2016. Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. *European journal of epidemiology* 31, 4 (2016), 337–350.
[6] M. Herlich, T. Pfeiffenberger, J. Du, and P. Dorfinger. 2018. Scenarios for Measurement of Reliable Wireless Communication in 5G. In *Computer Safety, Reliability, and Security (SAFECOMP)*.
[7] X. Ma, J. Zhang, and T. Wu. 2011. Reliability analysis of one-hop safety-critical broadcast services in VANETs. *IEEE transactions on vehicular technology* 60, 8 (2011), 3933–3946.
[8] D. Murray, T. Koziniec, M. Dixon, and K. Lee. 2015. Measuring the reliability of 802.11 WiFi networks. In *Internet Technologies and Applications (ITA)*. IEEE, 233–238.
[9] D. C. Salyers, A. D. Striegel, and C. Poellabauer. 2008. Wireless reliability: Rethinking 802.11 packet loss. In *World of Wireless, Mobile and Multimedia Networks (WoWMoM)*. IEEE, 1–4.
[10] J. Schmid, P. Heß, A. Höß, and B. W. Schuller. 2018. Passive monitoring and geo-based prediction of mobile network vehicle-to-server communication. In *2018 14th International Wireless Communications & Mobile Computing Conference (IWCMC)*. IEEE, 1483–1488.
[11] D. Son, B. Krishnamachari, and J. Heidemann. 2006. Experimental analysis of concurrent packet transmissions in wireless sensor networks. *ACM SenSys, Boulder, USA* (2006).
[12] S. Woo and H. Kim. 2010. Estimating link reliability in wireless networks: An empirical study and interference modeling. In *Proceedings of the IEEE INFOCOM*. IEEE, 1–5.