Telecommunication Testbed Repeatability Assessment

Jeanne Quimby[(1)], Alec Weiss[(1)], Jacob Rezac[(1)], Mary Gregg[(1)], Michael Frey[(1)],
Jason Coder[(1)], and Anna Otterstetter[(1)]
(1) National Institute of Standards and Technology, USA
jeanne.quimby@nist.gov

***Abstract*—Telecommunication testbeds are a fundamental tool in communication research, enabling prototyping and validating new ideas. Unfortunately, these testbeds are often highly complex, costly, and challenging to operate, requiring simultaneous open system interconnection multilayer measurements at different monitoring locations. In this paper, we illustrate our assessment methodology using measurands from a recent pilot study to assess the repeatability of the testbed.**

Keywords—communication device, measurands, multilayer measurements, repeatability, and telecommunication infrastructure

I. Introduction

Telecommunication infrastructure testbeds are an integral part of telecommunication research. Researchers can design, prototype, and verify their ideas. Also, measurement data collected from this testbed can be implemented in machine learning applications to further explore a researcher's idea. Performing measurements using these testbeds is challenging particularly when they include commercial-grade telecommunication hardware. Here, we present the methodology to assess our testbed's repeatability through a rigorous experimental design. Our experimental design enables us to capture testbed's repeatability [1] through collections of measurands over multi-days. Characterizing the repeatability is vital in assessing a researcher's hypothesis. Understanding the cause of systematic variations in the specific measurands is critical for producing high-quality scientific research interpretations and conclusions.

II. Telecommunication Infrastructure Testbed

Our telecommunication infrastructure testbed [2,3] as shown in Fig. 1 is a complete radio access network infrastructure with cell phones (i.e., User Equipment (UE)s), a carrier-grade base station (eNB/gNB), and a virtual Evolved Packet Core (EPC). Here, we used a full-stack traffic generator with the base station to create different operational scenarios. This testbed can operate in a controlled operational environment such as an anechoic chamber or a conducted test circuit. The controlled operational environment reduces external environmental uncertainties such as other radiating sources mixing with our desired base station radiated emissions. Our test equipment consists of a vector signal analyzer (VSA), user equipment test mobile sensor, and a traffic sensor. In our pilot study, network traffic between the base station and the core were used to confirm the proper establishment of a communication link between the UE and the base station. Using the VSA, we can monitor the radiated signal at the Physical (PHY) layer. The Medium Access Control (MAC) layer, Radio Link Control (RLC), Packet Data Convergence Protocol (PDCP), and network layer are monitored using the UE test mobile sensor. A few measurands we collected are in-phase and quadrature signal components, hybrid automatic repeat request (HARQ), reference signal received quality at the PHY layer, mean transport block (TB) size, mean resource block allocation, block error ratio (BLER) at the MAC layer, and PDCP downlink throughput at the PDCP layer. We collected data packet size, padding size, messages transferred between the core and the UE, and other measurands through our traffic sensor. In all, we have over 1000 possible measurands. Through engineering judgment, we narrowed down our collected measurands to around 200 with discrete, continuous, binary, and categorical data types for the pilot study. Fig. 2 presents measurands of different types.

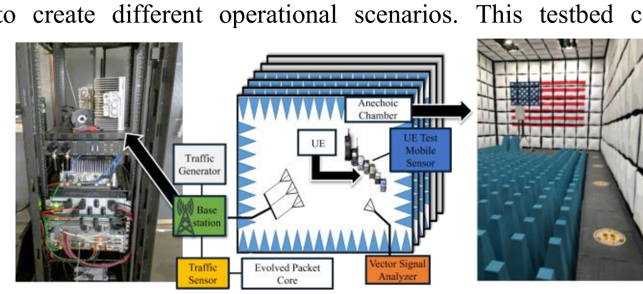
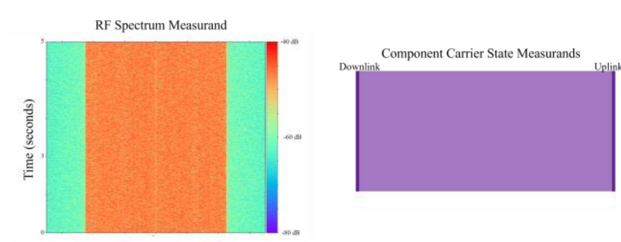


Figure 1. Commercial-grade Telecommunication Infrastructure Testbed with monitoring at different locations.

III. Assessment Methodology

A well-designed test, or measurement campaign, is based on capturing measurands that validate or negate a researcher's hypothesis. Hypothesis verification requires an understanding of the capabilities and limitations of the testbed's hardware and post-processing methods. Performing testbed measurements using sound metrological foundations (e.g., a controlled environment, test matrix, and design of experiment) ensures that the measurands are accurate. Clear documentation,



(a) (b)

Figure 2. Multilayer measurements capture measurands such as radio frequency (RF) spectrum and component carrier state.

including testbed design, test procedures, and measurement notes, provides helpful information when diagnosing unexpected features within the collected data. Finally, monitoring the testbed to ensure proper function is critical.

Researchers should identify the measurands of interest through a combination of test cases, engineering judgment, and prior research. Next, a well-articulated experimental design, as shown in Fig. 3, yields a structured block methodology for measurand collection and analysis. The order within each block is known by its "position number." The "session number" indicates the day each block is measured. The measurement campaign structure is determined by a pre-defined number of positions per block, blocks per session, and sessions per measurement campaign and should reflect the test cases. During the measurements, this block schedule is repeated many times with factors-of-interest changing for each block. From these repeat measurements, we perform analyses based on well-founded statistical principles [4] and begin preliminary training of machine learning-based algorithms [5]. This structure provides and understanding and quantifiable data on how these measurement factors impact the measurands.

Next it is important to include the testbed configuration to the experimental design specifications. Initial pilot studies should include diagnostic test procedures to verify their performance. The optimization of these pilot studies frequently requires software and hardware automation to realize the test design for complex states. We quantified our testbed instability through statistical and visualization tools such as an analysis of variance (ANOVA) and stabilograms [6] during the initial pilot studies. Additional pilot studies can leverage the formal design of experiment to be used during a measurement campaign.

A. Repeatabilty of Measurements

Supervised classical and machine learning algorithms, used to analyze multilayer telecommunications measurements, depend on labeled training examples of measurands. One additional benefit of a structured experimental design is its simplification of the measurement assessment. We can determine if they are repeatable/stable even while uncontrollable factors vary (e.g., time of day or internal device-under-measurement parameters). Analysis of these measurands can highlight trouble within the testbed, such as testbed instability.

Figure 4 (top) presents examples of measurands when the testbed is stable (over time or run number); Fig. 4 (bottom) shows examples when the testbed is operating unstably. The

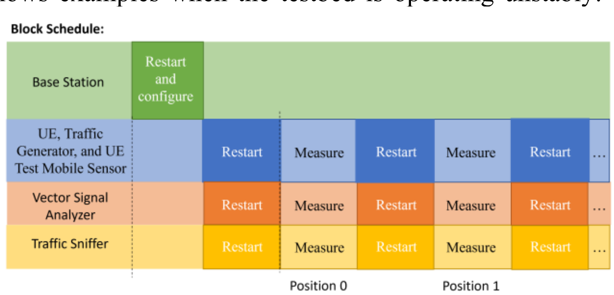


Figure 3. Block schedule of design of experiment.

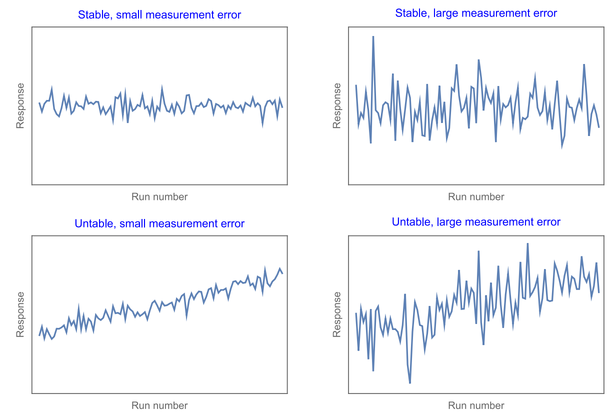


Figure 4. Stable (top) and unstable (bottom) testbed measurands with small (left) and large (right) measurement error.

measurands in Fig. 4 (bottom) are drifting (increasing) on average over the course of the experiment, indicating the presence of some unidentified variable(s) in the testbed or its environment affecting testbed measurements. Fig. 4 (left, right) shows measurands with, respectively, small and large measurement errors. The measurands collected in Fig. 4 (right) come from an experiment design that calls for test equipment restarts every 90 seconds. This design choice resets the internal parameters of the base station and UE uncontrollably and unrealistically. Training data for machine learning applications from this measurand might be fruitless because this large instability may adversely affect any machine learning results. Fig. 4 (left) provides examples of measurands in which the test equipment restarts every 90 minutes, a scenario more consistent with real-world applications. The differences in measurand variability as a function of restart frequency are an artifact of the experiment design; high restart frequency in Fig. 4 (right) poorly reflects real-world operation of the base station and UE. Through this assessment, we determined our best choice of experimental design for our larger measurement campaign to ensure stability in the testbed.

IV. Conclusion

Telecommunication research requires testbed measurements that follow a solid measurement methodology, founded on well-established experiment principles. Care and attention to testbed set-up are vital to achieving stable measurands that meet the measurement campaign's experiment design and scenario test cases.

References

[1] International Vocabulary of Metrology – Basic and General Concepts and Associated Terms, 2012

[2] J. Coder, et al., "Characterizing LTE User Equipment Emissions: Factor Screening," NIST Technical Note 2069, September 2019.

[3] A. Sanders, et al., "Laboratory Methods for Recording AWS-3 LTE Waveforms", Technical Note (NIST TN), National Institute of Standards and Technology, Gaithersburg, MD

[4] D. C. Montgomery, Design and Analysis of Experiments, 9th ed. Hoboken, NJ: John Wiley & Sons, 2017.

[5] I. Goodfellow, et al., Deep learning. MIT Press, 2016.

[6] M. Gregg, et al., "Testbed Instability in a Pilot Experiment for Cellular Communication Anomaly Detection," *in progress*