Self-Similarity of Traffic Within a 5G Standalone Network

Bryan Martin, Jacob Snyder, and Chad A. Bollmann
Dept. of Electrical & Computer Engineering
Naval Postgraduate School
Email: {bmartin, jacob.snyder, cabollma}@nps.edu

***Abstract*—Research has shown network traffic to be self-similar for various connection protocols, and 5G traffic has been modeled as self-similar based on the assumption that it adheres to these previously established traffic behaviors. However, to support the continued use of known self-similar traffic models in 5G technology, it is necessary to demonstrate self-similarity within 5G network traffic in a physical environment and investigate the factors that affect it. This research uses an AMARI Callbox Mini to develop a 5G Standalone (SA) network testbed to generate and analyze 5G network traffic. Using the Rescaled Range estimation method, the Hurst parameter of this traffic is measured to determine its degree of self-similarity. Our analysis demonstrates that traffic within a 5G standalone network is statistically self-similar, and the degree to which it displays this property increases with traffic load. Furthermore, this research shows that self-similarity varies dependent on the medium in which the traffic is collected and analyzed.**

***Index Terms*—5G, Self-Similarity, Network Traffic, Long-Range Dependence**

I. Introduction

The seminal work by Leland, Willinger, et al. [1] observed that network traffic captured within an Ethernet network is statistically self-similar, i.e., network traffic appears to retain its general shape when measured in bytes/second or packets/second, independent of the time-scale under which it was analyzed. Their work, which they expanded upon in [2], utilized the ideas presented by Mandelbrot [3] and described network traffic as possessing highly variable on and off periods (Noah Effect) while also exhibiting long-range dependence (Joseph Effect). Researchers have furthered these findings by demonstrating that network traffic originating from WiFi [4], [5], Cellular Digital Packet Data (CDPD) [6], and other cellular traffic [7], such as that from LTE/LTE-A devices [8], display self-similar characteristics.

However, it remains to be shown that the characteristics of self-similarity are present within 5G network traffic. The landscape of the Internet has fundamentally changed since [1], and 5G devices are the second fastest growing source of Internet traffic after Machine-to-Machine (M2M) / Internet-of-Things (IoT) [9]. A study into the self-similarity of 5G network traffic is warranted due to the variations in the 5G control and user plane protocol stack, changes to the signaling protocol within the network core, as well as other layer 1 and layer 2 improvements within 5G New Radio [10]. Given that previous research has modeled 5G network traffic under the assumption that it is self-similar [11], it is required to demonstrate empirically within a physical environment the existence of self-similarity in 5G network traffic. Furthermore, as the composition of the Internet evolves and how we deliver data changes, it is necessary to understand the factors contributing to the self-similarity of 5G traffic.

Therefore, the objective of this research is to analyze the behavior of 5G network traffic under varying conditions to determine its degree of self-similarity and the significant factors affecting this characteristic. The contributions of this research are to:

1) Demonstrate that network traffic originating from a 5G connected source is self-similar.
2) Show that the self-similarity of 5G network traffic differs under varying traffic load conditions.
3) Demonstrate that the self-similarity of 5G network traffic differs when measured on the wireless source link compared to when it is measured at the wired endpoint link.

Confirming that 5G traffic is self-similar supports the adaption of these models for 5G traffic and would lead to improvements in network performance and overall reliability for the end-user. Examples of this research, which can then be implemented in 5G networks, include utilizing self-similarity in resource allocation models for congestion control and improved quality of service [12], [13], queuing delay models [6], and as a metric for network anomaly detection [14].

This paper is organized as follows. Section II presents the definitions of self-similarity and how it is measured, along with a summary of previous related research. Section III discusses the testbed developed for this research and the methods used to collect and analyze the data for this experiment. Section IV shows the results of our experiments and explores how the network traffic behaves under varying conditions. Finally, Section V summarizes our results, providing our conclusions and proposed follow-on work.

II. Background

This section provides an overview of self-similarity and how it is measured. This is followed by a review of prior work in measuring the self-similarity of various network traffic models and how this traffic was generated and analyzed.

A. Self-Similarity

Self-Similarity, as described by [3] and originally observed in Ethernet network traffic in [1], is the concept that network traffic will retain its general shape or characteristics over varying timescales. Specifically, the autocorrelation of a time series $X(t)$ will remain equal (or proportional) to the original series when aggregated over varying time scales of size m.

Given a wide-sense stationary random process $X_t(t)$ with an autocorrelation function $r(k)$, for $k = 0, 1, 2, ...$, then for each $m = 1, 2, 3, ...$, the aggregated series is given by

$$X_k^{(m)}(t) = (1/m)(X_{km-m+1} + ... + X_{km}). \quad (1)$$

The process $X_t(t)$ is exactly second-order self-similar if the autocorrelation and variance of $X_k^{(m)}(t)$ is equivalent to that of the original process $X_t(t)$:

$$r^{(m)}(k) = r(k), k \geq 0. \quad (2)$$

The process $X_t(t)$ is asymptotically second-order self-similar when

$$r^{(m)}(k) \sim r(k) \quad (3)$$

as $m \to \infty$ [2].

When discussing network traffic analysis, the term asymptotically second-order self-similar is often used interchangeably with self-similar. Also, within the literature, when a time-series is asymptotically second-order self-similar, it is referred to as long-range dependent [15].

B. Hurst Parameter Estimation

The degree of self-similarity of a times series is described by the Hurst parameter H, which is a measurement of autocorrelation, and a series is said to be self-similar if $1/2 \leq H \leq 1$ [2].

A method for estimating the Hurst parameter the Rescaled Range (R/S) analysis [2]. We chose this method as it is commonly used within the literature. Also, as shown within the analysis conducted in [16], this estimator provides similar results when compared to other estimation methods such as the aggregated variance analysis. When considering the size of the data set to be analyzed, it was noted in [15] that Internet traffic is self-similar for the given timescale in which it is observed and data sets on the order of hours or minutes can be analyzed and shown to be self-similar. Therefore, we use packet captures of approximately 30 minutes in length for our results.

1) Rescaled Range: To estimate the Hurst parameter of a given random process or time series, consider the set of observation $X_k(n); k = 1, 2, ..., n$ with sample mean $\bar{X}(n)$ and sample variance $S^2(n)$. Next consider

$$R(n) = [max(0, W_1, W_2, ..., W_n) - min(0, W_1, W_2, ..., W_n)] \quad (4)$$

with

$$W_k = (X_1 + X_2 + ... + X_k) - k\bar{X}(n) \quad (5)$$

for $k = 1, 2, ..., n$. Then the expectation is

$$E[R(n)/S(n)] \sim cn^H \quad (6)$$

as $n \to \infty$ and c is a finite positive constant independent of n. Plotting these values on a log-log scale, the slope of the best fit line for these points equates to the Hurst parameter and will have a value of $H \geq 0.5$ for a self-similar process [2].

C. Prior Work

In [1], multiple sets of Ethernet traffic collected over a 4-year period were analyzed to determine their degree of self-similarity. The authors demonstrated that the Ethernet traffic was self-similar across all data sets and that the aggregated traffic across the link maintained this "bursty" nature. Their continued research in [2] and [17] supported their initial findings and demonstrated how self-similarity can be modeled to relate the Hurst parameter to the parameters of the source model distribution.

Similar to this research, the authors of [4] developed a wireless testbed to analyze the long-range dependence and self-similarity of network traffic originating from wireless devices under the IEEE 802.11b standard. In their experiment, the generated traffic consisted of HTTP, FTP, and video streaming data that was collected over a period of days. The authors showed that the wireless traffic was self-similar, and the degree of self-similarity varied under a given load, with the higher load of traffic having a greater degree of self-similarity. Also, the authors empirically demonstrated that the self-similarity of TCP connections is greater than that of UDP connections. This research was supported by the work in [5], where the authors recreated a similar scenario using the simulation tool OPNET. Their work corroborated that wireless traffic is statistically self-similar while also demonstrating the aggregated traffic from multiple self-similar sources retains its self-similarity.

In [6], researchers analyzed CDPD from a network provider, finding that the activity within this packet trace was self-similar and long-range dependent. Using their findings, the authors then simulated a CDPD network in OPNET and demonstrated how the queuing delays of self-similar traffic differ from short-range dependent models such as Poisson arrivals.

The authors of [8] obtained a data set from an active LTE/LTE-A network from a US-based cellular provider that was captured over a 24-hour period at the eNodeB level. The logs, which varied in size up to 30 minutes, were analyzed using the variance analysis method to determine the Hurst parameter of the given traffic. Their work confirmed that network traffic utilizing the LTE/LTE-A infrastructure is self-similar. They also showed that the degree of self-similarity varied throughout the day in conjunction with normal daily human use and the traffic load at the eNodeB.

Our work picks up from previous research and expands the field by including 5G enabled devices as the source of traffic within the network. This work demonstrates that network traffic originating from a 5G connected device is self-similar and that self-similarity differs under varying load. These findings allow for the continuing use of long-held theoretical models on how traffic behaves in a network.

III. Methodology

In this section, we present the design implementation of the 5G portion of the network testbed used in this research, along with the method of generating traffic. Finally, we discuss the various trials implemented within this experiment.

A. Testbed Setup

For this research, a 5G Standalone (SA) Network testbed was developed to analyze the statistical characteristics of network traffic, which is depicted in Figure 1.

The 5G Core (5GC) and Radio Access Network (RAN) for this testbed is an AMARI Callbox Mini, which serves as a 3GPP compliant standalone network [18]. The AMARI Callbox Mini was chosen for implementation within this testbed due to its capability to act as a "network-in-a-box", providing both the gNB and 5GC capabilities to connect and manage the User Equipment (UE) within the testbed.

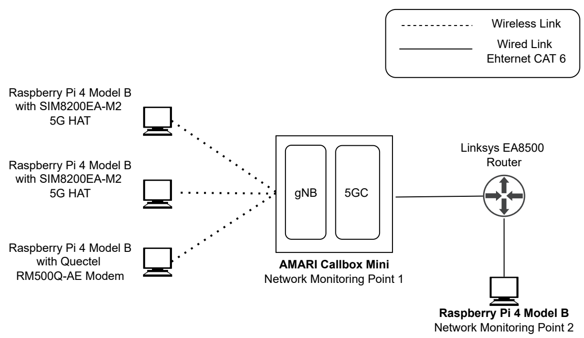

Figure 1. 5G Standalone Network Testbed Architecture

The hosts for this network consisted of four Raspberry Pi 4 Model Bs operating with the latest Raspberry Pi OS (version 5.15). Three of these hosts were 5G capable, serving as the UEs connected to the AMARI Callbox Mini, while the fourth is connected to the Linksys EA8500 router and acts as the endpoint destination for the traffic generators.

The three UEs were placed approximately six feet from the AMARI Callbox while in operation. Two of the UEs utilized the SIM8200EA-M2 5G Hardware Attached on Top (HAT), which supports 3G/4G/5G connectivity as well as multi-mode, multi-band capability and 5G SA and NSA (non-standalone) operability. The third UE used the Quectel RM500Q-AE module to connect to the network. This module is capable of both 4G and 5G implementations as well as 5G SA and NSA operability. This device configuration was chosen due to equipment availability and not for any technical differences between the SIM8200EA-M2 5G Hat and Quectel RM500Q-AE modules. The Quectel RM500Q-AE enabled host was only used for the trials which consisted of generating traffic from all three devices.

B. Traffic Generation

The traffic generated within the network was created using the Multi-Generator (MGEN) Network Test Tool developed by the U.S. Naval Research Laboratory. This tool generates real-time UDP or TCP network traffic under variable, user-defined traffic loads. The tool generates traffic which emulates either periodic, Poisson, or bursty source models [19].

The source model for this experiment was defined to use the burst option, transmitting packets at random intervals that were exponentially-distributed with a mean value of 5 s. The duration of a given burst transmission was exponentially-distributed with a mean value of 2 s and a fixed packet size of 1024 bytes. To test the effects of traffic load on the degree of self-similarity within the network, the packet send rate was a Poisson distribution varied with a mean value of 1, 10, or 100 packets per second, as shown in Table I. These input values for the traffic generator were chosen to provide bursty traffic while providing the capability to vary traffic load (packets per second). The generated traffic was captured using the Tshark packet capture tool, which allowed for packet captures at the granularity of one microsecond [20]. An example of the MGEN configuration script used for generating UDP traffic at a mean transmission rate of 100 packets per second is shown in Figure 2.

Table I
Trial Specifications

Trial	Active UEs	Transport Protocol	Load [pps]
1-3	1	TCP	1, 10, 100
4-6	1	UDP	1, 10, 100
7-9	2	TCP	1, 10, 100
10-12	2	UDP	1, 10, 100
13-15	3	TCP	1, 10, 100
16-18	3	UDP	1, 10, 100

```
0.0 ON 1 UDP DST 192.168.1.10/5001
      BURST [RANDOM 5.0  Poisson [100.0 1024] EXP 2.0]
```

Figure 2. Example of MGEN UDP Traffic Script (100 packets per second)

C. Experiment Procedures

Using the MGEN traffic generation tool, a 30 minute sample of traffic was generated at the UEs for 18 trials under the conditions listed in Table I. Only one UE was active in the first set of trials (Raspberry Pi 4 Model B with SIM8200EA-M2 5G HAT). For each trial, the traffic load was varied from a mean of 1, 10, or 100 packets sent per second under both UDP and TCP transport protocols. The experiment was conducted for a second set of trials with the same set of load and transport protocol variations, adding an additional Raspberry Pi 4 Model B with SIM8200EA-M2 5G HAT, before a third and final set of trials was conducted using the full network configuration as shown in Figure 1. For each trial listed in Table I, the traffic was collected at the AMARI Callbox, which is annotated as Point 1 on Figure 1, and the self-similarity was calculated as discussed in Section II.

To investigate how the self-similarity of a single device may vary based on the connection medium and where it was measured, trials 1-6 were conducted a second time. For this set of trials, the traffic load was varied from a mean of 1, 10, 50, or 100 packets sent per second under both UDP and TCP transport protocols. The traffic from these trial sets was collected at the AMARI Callbox, labeled as Point 1 in Figure 1, and the end device, which is annotated as Point 2 in Figure 1. The results were compared and presented in Section IV.

IV. Results

In this section we discuss and summarize the results of each trial described in Section III, as well as address any areas for improvement discovered within these methods.

Before analyzing the data to determine the degree of self-similarity based on the estimation of the Hurst parameter, an Augmented Dickey-Fuller Test (ADF Test) was conducted in MATLAB to determine if the collected data represented a stationary or non-stationary time series [21]. Testing against the null hypothesis that the time series represented by our data has a unit root (and is thus non-stationary), the null hypothesis was rejected for all trials, suggesting that the data collected is stationary.

A. TCP Traffic Results

Over each of the nine trials conducted generating only TCP traffic, the network traffic originating from a 5G source exhibited the property of self-similarity. The results of each of these TCP trials are shown in Table II. As an example of the analysis conducted, the Rescaled Range analysis for the trial consisting of 3 active UEs and 100 packets/second load is shown in Figure 3. Within the figure, m is the number of aggregated samples, R is the Range of the series, S is the standard deviation, and R/S is the calculated Rescaled Range value. When graphed on a log-log plot, the slope of the best-fit line for this series equates to the estimated Hurst parameter, and any estimation greater than 0.5 and less than 1.0 is deemed self-similar. For this trial, the Hurst parameter was 0.8460.

The degree of self-similarity did increase with the increase in traffic load (packets/second) but did not change significantly with the increase in user volume (number of active UEs). This is demonstrated in Figure 4, and could be a result of the low number of UEs active on the network being insufficient to cause a change in self-similarity. Further experimentation with a larger number of active users is required to demonstrate whether self-similarity varies with volume as it does with the load. Yet, data suggests that an increase in the volume of users would change the self-similarity due to the subsequent increase in load from the additional users.

B. UDP Traffic Results

For the nine trials consisting of UDP generated traffic, similar results were seen as with the TCP generated traffic. These results, shown in Table III, demonstrate that 5G UDP network traffic is self-similar. As an example of these results, the Rescaled Range analysis for the trial consisting of 3 active UEs and a 100 packets/second load is shown in Figure 5. The estimated Hurst parameter, represented by the slope of the line fitted to the data, is 0.8325. Similarly, the comparison of self-similarity based on volume and load for all trials is shown in Figure 6. As with the TCP traffic, though the degree of self-similarity increases with an increase in load, the change in number of UEs is presumed insufficient to effect the outcome.

Table II
R/S Analysis Results of TCP Traffic

Acitve UEs	Load [pps]	$\text{Hurst}_{R/S}$
1	1	0.5342
	10	0.7323
	100	0.8950
2	1	0.5565
	10	0.7486
	100	0.8597
3	1	0.5622
	10	0.7573
	100	0.8749

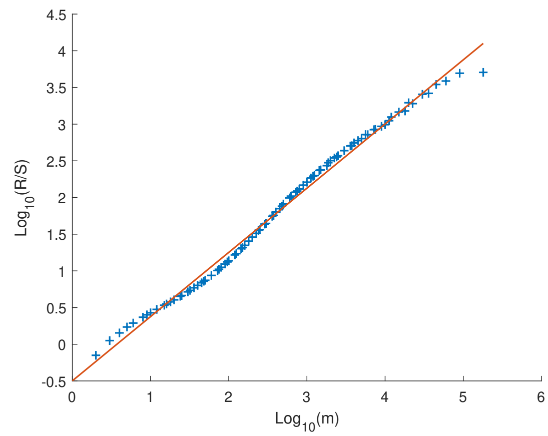

Figure 3. R/S plot for TCP traffic

When comparing the results in Tables II and III there appears to be little to no variation between the two protocols, as would be expected based on previous research [4]. This may be due to the operation of the MGEN traffic generator. The traffic generator creates a single, open connection for the specified time-interval when creating TCP traffic. The use of the traffic generator would need to be modified in future experiments to simulate a more realistic series of open and closed connections. For this reason, we consider the results comparing the self-similarity of UDP versus TCP traffic in this experiment to be inconclusive.

C. Analysis based on measurement location

The results for the final experiment, where traffic from one UE was measured at two locations while varying the load, are shown in Figure 7 and listed in Tables IV and V. For both traffic scenarios, the degree of self-similarity was noticeably different whether the traffic was measured at the wireless link between the UE and the AMARI Callbox or at the wired link

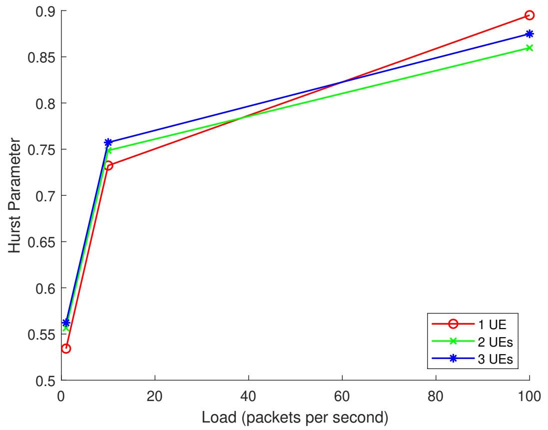

Figure 4. Hurst comparison based on TCP traffic load

Table III
R/S ANALYSIS RESULTS OF UDP TRAFFIC

Active UEs	Load [pps]	$\text{Hurst}_{R/S}$
1	1	0.5125
	10	0.7183
	100	0.8641
2	1	0.4952
	10	0.7024
	100	0.8607
3	1	0.5516
	10	0.7131
	100	0.8574

between the router and the end-device. These results suggest that, for a given load, there is a smoothing effect on bursty data as it transitions from a wireless to a wired medium. Further investigation is required to determine to what degree the self-similarity changes based on the connection medium, if and at what rate it converges, and the underlying causes behind these observations.

Table IV
TCP TRIAL RESULTS

Load [pps]	$\text{Hurst}_{R/S}$ AMARI Callbox	$\text{Hurst}_{R/S}$ End Device
1	0.6139	0.6098
10	0.7610	0.6363
50	0.8412	0.7790
100	0.8542	0.8330

V. CONCLUSIONS AND FUTURE WORK

It is a long-held notion that network traffic is self-similar, a characteristic that has been observed within Ethernet, WiFi, and the previous generations of cellular technology. As technology advances, however, it is necessary to reaffirm and validate previously established concepts within the field for continued use of known models.

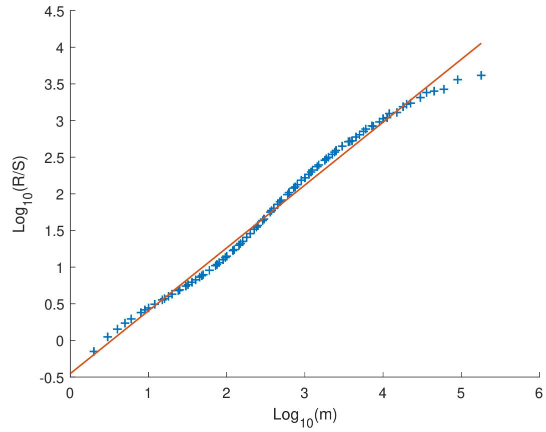

Figure 5. R/S plot for UDP traffic

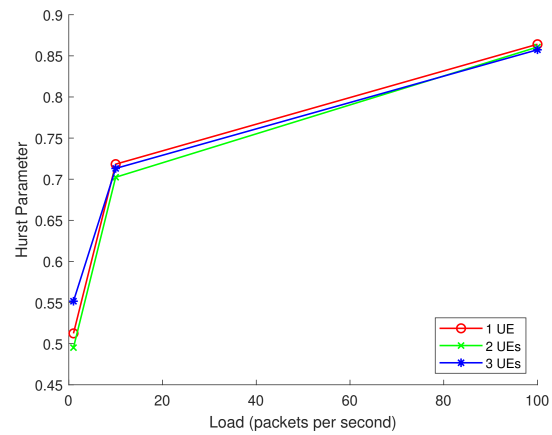

Figure 6. Hurst comparison based on UDP traffic load

To this end, the research in this work demonstrates the self-similarity of network traffic originating from devices in a 5G SA network. We also confirm that the self-similarity of 5G network traffic increases with increased traffic load. This research was unable to establish any changes in self-similarity based on the volume of users, which requires further study with a larger testbed or dataset.

An interesting result is how the self-similarity of network traffic changes as it transitions through the network. Our results suggest that the connection medium and protocol have a measurable effect on the self-similarity of network traffic, along with the effects of the transport and application layer protocols.

For future work, our team hopes to acquire an anonymized dataset from an active gNB for analysis of real-world user traffic. The traffic generated in this experiment was strictly UDP or TCP traffic, with identified deviations from true traffic within the traffic generating tool. To fully demonstrate the difference in self-similarity of UDP and TCP 5G traffic, a traffic generating profile that more closely emulates human and/or machine behavior is required.

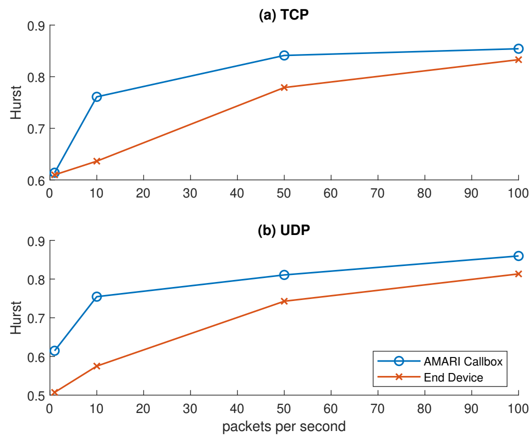

Figure 7. Hurst comparison based on point of measurement

Table V
UDP Trial Results

	$Hurst_{R/S}$	
Load [pps]	**AMARI Callbox**	**End Device**
1	0.6147	0.5072
10	0.7545	0.5751
50	0.8107	0.7427
100	0.8597	0.8133

Other plans for future experimentation are to compare the degree of self-similarity of network traffic originating from devices using various connection protocols. Expanding the testbed used in this research to include Ethernet- and WiFi-connected devices, we plan to analyze any changes in self-similarity that may occur as the traffic aggregates. This will present the opportunity to expand upon the results seen here, where self-similarity differed based on where the traffic was measured.

References

[1] W. E. Leland, W. Willinger, M. S. Taqqu, and D. V. Wilson, "On the self-similar nature of ethernet traffic," *ACM SIGCOMM computer communication review*, vol. 25, no. 1, pp. 202–213, 1995.

[2] W. Willinger, M. S. Taqqu, W. E. Leland, and D. V. Wilson, "Self-similarity in high-speed packet traffic: analysis and modeling of ethernet traffic measurements," *Statistical science*, pp. 67–85, 1995.

[3] B. Mandelbrot, "Self-similar error clusters in communication systems and the concept of conditional stationarity," *IEEE Transactions on Communication Technology*, vol. 13, no. 1, pp. 71–90, 1965.

[4] C. Oliveira, J. B. Kim, and T. Suda, "Long-range dependence in ieee 802.11b wireless lan traffic: an empirical study," in *2002 14th International Conference on Ion Implantation Technology Proceedings (IEEE Cat. No.02EX505)*, 2003, pp. 17–23.

[5] Q. Yu, Y. Mao, T. Wang, and F. Wu, "Hurst parameter estimation and characteristics analysis of aggregate wireless lan traffic," in *Proceedings. 2005 International Conference on Communications, Circuits and Systems, 2005*, vol. 1. IEEE, 2005, pp. 339–345.

[6] M. Jiang, M. Nikolic, S. Hardy, and L. Trajkovic, "Impact of self-similarity on wireless data network performance," in *ICC 2001. IEEE International Conference on Communications. Conference Record (Cat. No. 01CH37240)*, vol. 2. IEEE, 2001, pp. 477–481.

[7] C. Qi, Z. Zhao, R. Li, and H. Zhang, "Characterizing and modeling social mobile data traffic in cellular networks," in *2016 IEEE 83rd Vehicular Technology Conference (VTC Spring)*. IEEE, 2016, pp. 1–5.

[8] R. K. Polaganga and Q. Liang, "Self-similarity and modeling of lte/lte-a data traffic," *Measurement*, vol. 75, pp. 218–229, 2015.

[9] "Cisco annual internet report - cisco annual internet report (2018–2023) white paper," Mar 2020. [Online]. Available: https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html

[10] C. Cox, *An Introduction to 5G*. Wiley, 2020.

[11] A. M. Alba and W. Kellerer, "Large-and small-scale modeling of user traffic in 5g networks," in *2019 15th International Conference on Network and Service Management (CNSM)*. IEEE, 2019, pp. 1–5.

[12] G. He, Y. Gao, J. C. Hou, and K. Park, "A case for exploiting self-similarity of network traffic in tcp congestion control," in *10th IEEE International Conference on Network Protocols, 2002. Proceedings.* IEEE, 2002, pp. 34–43.

[13] K. Park, G. Kim, and M. E. Crovella, "Effect of traffic self-similarity on network performance," in *Performance and Control of Network Systems*, vol. 3231. SPIE, 1997, pp. 296–310.

[14] S. M. Popa and G. M. Manea, "Using traffic self-similarity for network anomalies detection," in *2015 20th International Conference on Control Systems and Computer Science*. IEEE, 2015, pp. 639–644.

[15] W.-B. Gong, Y. Liu, V. Misra, and D. Towsley, "Self-similarity and long range dependence on the internet: a second look at the evidence, origins and implications," *Computer Networks*, vol. 48, no. 3, pp. 377–399, 2005.

[16] W. Rea, L. Oxley, M. Reale, and J. Brown, "Estimators for long range dependence: An empirical study," *arXiv preprint arXiv:0901.0762*, 2009.

[17] W. Willinger, M. S. Taqqu, R. Sherman, and D. V. Wilson, "Self-similarity through high-variability: statistical analysis of ethernet lan traffic at the source level," *IEEE/ACM Transactions on networking*, vol. 5, no. 1, pp. 71–86, 1997.

[18] Amarisoft, "Amari callbox mini," [Online], https://www.amarisoft.com/products/test-measurements/amari-lte-callbox/.

[19] MGEN, "Multi-generator (mgen) network test tool," [Online], https://www.nrl.navy.mil/Our-Work/Areas-of-Research/Information-Technology/NCS/MGEN/.

[20] Wireshark, "tshark - dump and analyze network traffic," [Online], https://www.wireshark.org/docs/man-pages/tshark.html.

[21] "Augmented dickey-fuller test." [Online]. Available: https://www.mathworks.com/help/econ/adftest.html