

Impact of long-range dependent traffic in IoT local wireless networks on backhaul link performance

Przemysław Włodarski

Department of Signal Processing and Multimedia Engineering,
West Pomeranian University of Technology, Szczecin,
70-313 Szczecin, Poland
`przemyslaw.wlodarski@zut.edu.pl`

Abstract. Performance evaluation in Internet of Things (IoT) networks is becoming more and more important due to the increasing demand for quality of service (QoS). In addition to basic statistical properties based on the distribution of interarrival times of packets, actual network traffic exhibits correlations over a wide range of time scales associated with long-range dependence (LRD). This article focuses on examining the impact of both LRD and number of nodes that transmit packets in a typical IoT wireless local network, on performance of the backhaul link. The analysis of latency and packet loss led to an interesting observation that the aggregation of packet streams, originating from single nodes, lowers the importance of LRD, even causing an underestimation of performance results when compared to the queueing system with Markovian input.

Keywords: IoT wireless network · backhaul link · long-range dependence · traffic analysis · performance evaluation

1 Introduction and related work

Internet of Things (IoT) wireless networks integrate many physical devices and sensors that send data through wire or wireless connections and have different architectures (e.g. fog, edge, cloud). Because of the heterogeneity of IoT systems it is not easy task to design reliable and efficient communication systems [8]. One of the big issue is energy consumption, which depends on: incorrect selection of the microcontroller, energy-inefficient software [3] or communication protocol parameters [12]. Another big issue is the impact of offered load on quality of service (QoS) and quality of experience (QoE), especially for interactive or streaming services. There are two main measures in terms of QoS that relate to end-to-end connection performance: packet loss and latency. This article presents the results of these measures for different scenarios, taking into account long-range dependent (LRD) feature of the traffic.

Wireless communication in IoT networks is usually based on one of the Medium Access Control (MAC) protocols: Time Division Multiple Access (TDMA)

or Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA). In TDMA protocol each node is assigned a time slot for data transmission in a predetermined order. The main disadvantage of this system is that it requires accurate synchronization, which reduces the efficiency of the entire system. The second MAC protocol widely used in IoT systems is well known CSMA/CA (example use: IEEE 802.11 family, IEEE 802.15.4, etc.), that was extensively studied in many articles. Comprehensive study on the throughput, delay and stability performance of CSMA networks was presented in [6]. The problem of stable throughput and bounded mean delay was discussed in [24]. Another interesting contribution in the area of CSMA/CA protocols was a proposition of Handshake Sense Multiple Access with Collision Avoidance (HSMA/CA) protocol [19], which protects a densely deployed network from the classical hidden and exposed terminal problems. This idea was developed using Markov modeling and simulations. Recently, authors of [9] considered maximum effective throughput and suggested that the minimum mean access delay parameter of CSMA/CA system is of great practical interest.

Network traffic affects the reliability and performance of a network. Information on the statistical distribution of interarrival times of packets is not sufficient for evaluation of network performance, since actual network traffic exhibits second-order properties associated with long-range dependence (LRD) [1, 20]. It is very important to consider LRD, because of the impact on queueing performance [16]. Many models were developed from the properties of fractional Gaussian noise [17] or fractional autoregressive moving average process [4]. One of the most popular model that incorporates LRD properties is on/off source, recently analyzed in [25]. There is also a modified version of the Pareto on/off model [13], which is used further in this article.

Network traffic analysis and modeling is crucial for design and implementation of efficient and reliable transmission networks. It helps explain possible problems before they occur. The simulation results can be used to identify anomalies [5, 10, 7] or detect Distributed Denial of Service flood attacks [11, 15]. Furthermore, most of the symptoms that lead to congestion and high level of packet loss rate can be detected in the simulation process of network traffic [2, 23].

Experimenting with physical devices is uneconomical in the first phase of the project, especially if the number of devices is large. Therefore, the simulation approach is optimal in developing new methods and testing new scenarios. For the purposes of this article, all simulations were carried out using OMNeT++ framework [22, 21], which is a powerful open-source discrete event simulation tool. In fact, it is component-based C++ simulation library and framework, but instead of providing components specifically for computer networks, it includes generic component architecture to create any simulation.

The article is organized as follows. Next section characterizes basic properties of long-range dependence and introduces estimation methods used in further sections of this paper. Section 3 presents considered network structure as well as the model of network traffic generated by single node. In addition, it contains all the statistics for the reference queueing system, which are further studied from

the point of view of performance evaluation. In section 4, the results obtained for the analyzed network are then compared with the corresponding statistics of a commonly used queueing model described in section 3.3.

2 Long-range dependence

2.1 Basic properties

In order to explain the concept of long-range dependence (LRD), one needs to take a closer look at the stochastic process for different time scales. Let $Y(t)$ be the stationary stochastic process. The following simple equation describes the relationship for the process that is rescaled in time:

$$Y(at) \stackrel{d}{=} a^H Y(t), \quad a > 0, \quad (1)$$

where a is a stretching factor and H is the Hurst exponent and $\stackrel{d}{=}$ denotes equality in distributions. If $0.5 < H < 1$ then second-order properties associated with correlation structure are preserved regardless of scaling in time and the process becomes LRD. The higher value of H the stronger dependence. The autocorrelation function for the incremental process $X(i) = Y(i) - Y(i-1)$, $i = 1, 2, \dots$, which reflects the similarity between $X(i)$ and $X(i+k)$, has the following form:

$$r_k = \frac{\sigma^2}{2} ((k+1)^{2H} - 2k^{2H} + |k-1|^{2H}), \quad k = 0, 1, \dots \quad (2)$$

and for $H > 0.5$ is not summable:

$$\sum_{k=0}^{\infty} r_k \rightarrow \infty. \quad (3)$$

The value of autocorrelation function decays slowly for LRD processes. In case of no-LRD processes, there is no dependency ($H = 0.5$) and $r_k = 0$ for $k \geq 1$.

2.2 Estimation

In order to evaluate Three methods of estimation of Hurst exponent were used. First one is variance-time method, which is based on the aggregated process of $X(n)$ for discrete times $n = 0, 1, \dots, N$ that corresponds to fixed-length intervals:

$$X^{(m)}(n) = m^{-1} \sum_{t=mn}^{m(n+1)-1} X(t), \quad n = 0, 1, \dots, \lfloor N/m \rfloor - 1, \quad (4)$$

where m denotes the level of aggregation, i.e.:

$$X^{(m)} = m^{-1} \sum_{t=1}^m X(t) = m^{H-1} X \quad (5)$$

The variance of the aggregated random variable in (5) is:

$$Var(X^{(m)}) = \sigma^2 m^{2H-2}, \quad (6)$$

where σ^2 is the variance of X . It can be easily seen that if one performs logarithmic operation on both sides of (6) then Hurst exponent can be estimated from the slope of linear regression $(2H - 2)$ for all aggregated samples according to (4).

The next, similar method of estimation is Index of Dispersion for Counts (IDC). It is defined as a relation of variance-to-mean ratio of the sum of random variable X for period L :

$$IDC(L) = Var\left(\sum_{n=1}^L X(n)\right) / E\left(\sum_{n=1}^L X(n)\right) \approx cL^{2H-1}, \quad (7)$$

where c is a positive value. As with the variance-time method, one can use the linear regression to get the estimated $\tilde{H}$ from the slope $(2H - 1)$.

Another method of estimation is periodogram based on the approximated value of spectral density of LRD processes:

$$f(\lambda, H) \approx \sin(\pi H) \Gamma(2H + 1) |\lambda|^{1-2H}, \quad (8)$$

where λ is the frequency value for analyzed random variable X . Although the estimation operation is done in the frequency domain, the Hurst exponent can also be calculated from linear regression. In this case, the FFT values should be taken as the regression points. LRD refers to the lowest frequencies, which is reflected in the formula (8), where most of the energy concentrates near 0. For that reason, only 10% of the lowest frequencies is considered in the periodogram estimation method.

3 Framework

3.1 Network

In order to analyze network traffic in a typical and commonly used structure for IoT devices shown in fig. 1, a simple and efficient non-persistent CSMA/CA protocol is assumed. This protocol was chosen because it can be easily implemented even in basic and cheap microcontrollers and does not consume much power during wireless operation, which is crucial for IoT wireless sensors. In non-persistent version of CSMA protocol waiting node does not listen to the channel continuously until it becomes idle (like in 1- or p-persistent versions), which reduces energy consumption.

The network consists of $nNodes$ wireless nodes and two stations: $st0$ and $st1$. Each node can transmit packets to the station $st0$ and can sense transmission from another node to avoid collisions. All nodes can hear each other, so there is no hidden node problem [14]. When the channel is busy, because another node is

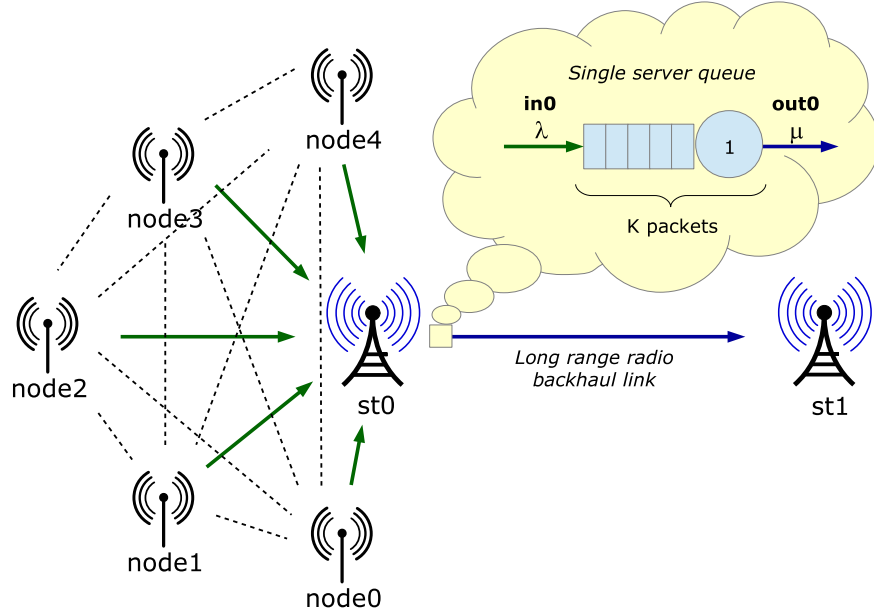


Fig. 1. Schematic diagram of analyzed IoT wireless network, example for 5 nodes.

transmitting, the node that wants to send packets must wait a period of time and then listen (sense) again. If the channel is idle, the node starts its transmission immediately.

Wireless station $st0$ receives packets from all nodes and tries to send them all across the long range radio link to the station $st1$. All packets must pass through the wireless network interface of $st0$ that connects this station to another one ($st1$). The output interface at $st0$ actually incorporates the queueing system that has K places for packets (including the one being transmitted) and the transmission circuit limited by the bandwidth of the radio link. Therefore, because all the cumulative traffic from the nodes goes there, all interesting performance statistics, associated with the impact of LRD traffic, can be found at the radio link network interface of station $st0$. Before the packet leaves $st0$, either it is immediately processed (if the queue is empty) or goes to the queue buffer. If there is not enough buffer space, packet is dropped (buffer overflow). By changing the bandwidth of this radio link channel, one can examine the impact of the local traffic on queueing performance, and thus on the level of packet loss and latency.

3.2 Source node

Every node sends fixed length packets to the np-CSMA channel according to Pareto Modulated Poisson Process (PMPP) [13]. This model was chosen because

it is versatile, efficient and introduces long-range dependence to the traffic while maintaining a constant level of packet rate. In addition, it resembles the behavior of variable-bit-rate services or protocols that transmit packets in batches. The PMPP source consists of two Poisson sources with alternating traffic intensities λ_1 and λ_2 (fig. 2). The sojourn time in each state has Pareto distribution $P\{X \geq x\} = x^{-\alpha}$ with parameter $\alpha > 0$. This distribution has infinite variance for $1 < \alpha < 2$ and is heavy-tailed.

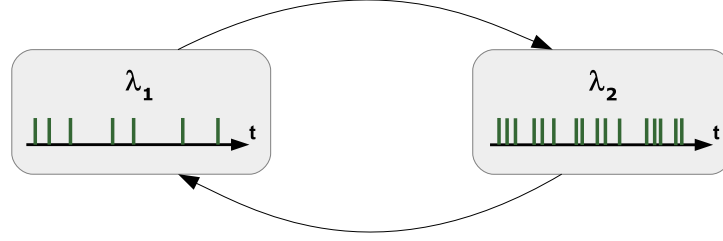


Fig. 2. Two Poisson sources of PMPP packet generator

The approximated value of IDC for PMPP source is:

$$IDC(t) \approx 1 + \frac{(\lambda_1 - \lambda_2)^2}{\lambda_1 + \lambda_2} \left(\frac{\alpha - 1}{\alpha} \right) t^{2-\alpha}, \quad (9)$$

where H can be easily obtained from:

$$H = \frac{3 - \alpha}{2} \quad (10)$$

and is compatible with (7) in terms of the same exponent $(2H - 1)$. The λ_1 and λ_2 values should be selected so that the expected value of number of packets $E(N(t)) = 0.5(\lambda_1 + \lambda_2)t$ corresponds to the desired value of the generated network traffic.

3.3 Performance evaluation

The network performance of backhaul link between $st0$ and $st1$ depends on statistical properties of the inbound traffic as well as the service rate and packet length distribution. Since all packets have fixed size, a deterministic service is assumed. All traffic from local network goes to the input of queueing system inside the output interface of $st0$ (fig. 1). The queueing system consists of one server and has $K - 1$ slots as a buffer space for packets. If the buffer overflows (K packets in the system) then the next incoming packet is dropped. Most common type of queueing system that meets the above assumptions is M/D/1/K, for which explicit formulas of blocking probability, stationary distribution and mean system sojourn time were derived in [18]. Both latency and packet loss can be

expressed in terms of steady state probabilities of number of packets in the system:

$$D = \frac{1}{\lambda(1 - P_{LOSS})} \sum_{k=0}^K k \cdot p_k^{(K)} \quad (11)$$

$$P_{LOSS} = p_K^{(K)}, \quad (12)$$

where:

$$p_k^{(K)} = \begin{cases} (1 + \rho \Lambda_{K-1})^{-1} & \text{for } k = 0 \\ (\Lambda_k - \Lambda_{K-1}) p_0^{(K)} & \text{for } k = 1, \dots, K-1 \\ 1 - \Lambda_{K-1} p_0^{(K)} & \text{for } k = K \end{cases} \quad (13)$$

$$\Lambda_k = \sum_{i=0}^k \frac{(\rho(i-k))^i}{i!} \exp((k-i)\rho). \quad (14)$$

These relationships are the reference for comparing them with the data received from the interface of *st0* for different scenarios, i.e. different levels of LRD as well as different number of nodes.

4 Results

All results were obtained using the OMNeT++ [22] simulation platform. The simulation framework was described in the previous section.

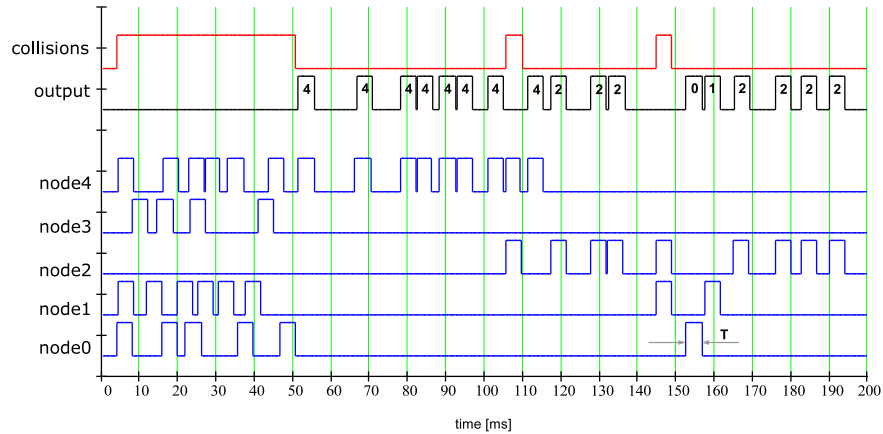


Fig. 3. Sample result for the first 200 ms of highly congested traffic, nNodes: 5, bandwidth: 1 Mbps, packet transmission time (T): 4.096 ms.

The performance results were compared to a typical M/D/1/K queuing system, presented in 3.3, where all performance measures refer to the classical Poisson model of input traffic (without LRD feature).

The purpose of the experiments was to examine the effect of different Hurst exponent values in the range of $0.5 < H < 1$ as well as different number of nodes ($nNodes$) on performance of the backhaul link (fig. 1). The simulation time, bandwidth and packet length for each scenario were 60 min., 1 Mbps and 4096 bits (512 bytes), respectively. The packet length corresponds to the packet transmission time of 4.096 ms shown in fig. 3.

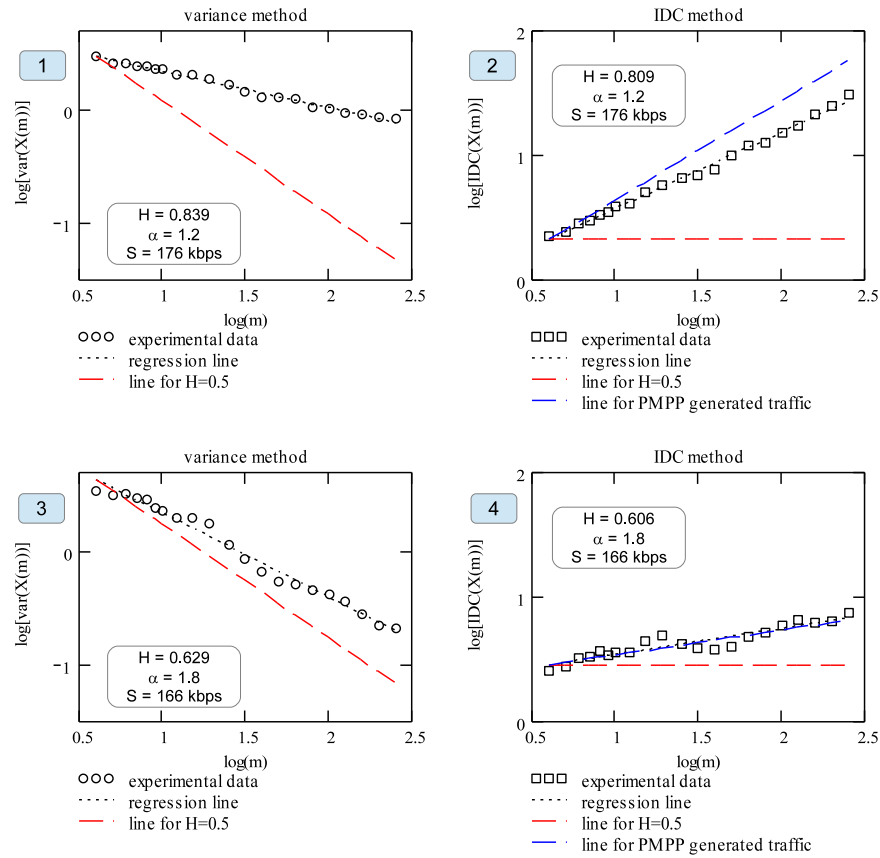


Fig. 4. Variance and IDC plots for two desired Hurst exponent values: $H = 0.9$ ($\alpha = 1.2$) and $H = 0.6$ ($\alpha = 1.8$).

Figure 4 shows the estimation results of Hurst exponent ($\tilde{H}$) for the aggregated outbound traffic from 5 nodes. Two methods of estimation were applied: variance-time and IDC plot (see section 2.2 and equations (5), (7)). The desired

H was 0.9 and 0.6, which corresponds to the $\alpha = 1.2$ and $\alpha = 1.8$ of Pareto distribution in PMPP model.

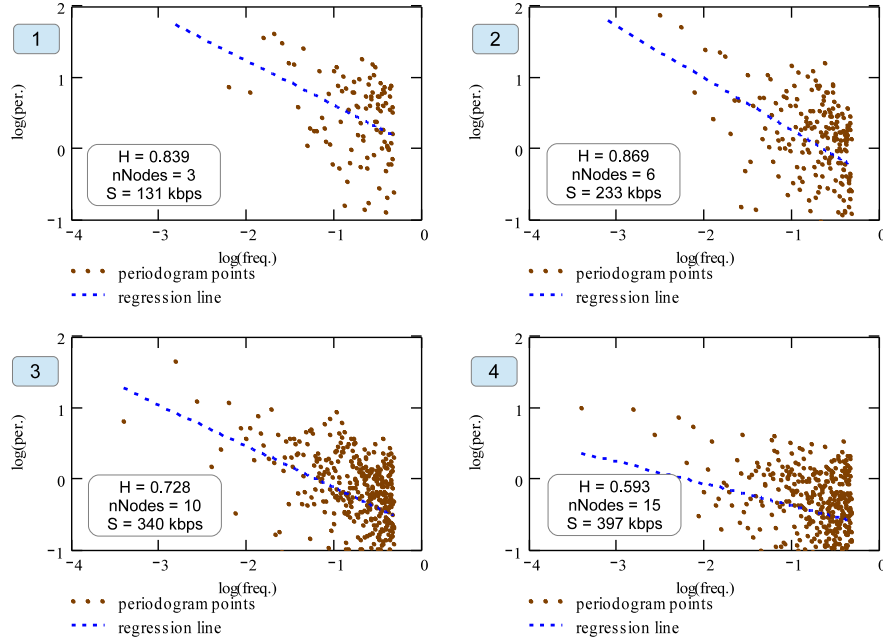


Fig. 5. Periodogram plots for different number of nodes.

Table 1. Main statistics for $H = 0.9$ and different number of nodes.

Number of nodes:	2	3	6	8	10	15
throughput [kbps]	81	131	233	290	397	326
packets at $in0$ [pkts]	70829	112264	192267	239938	291615	345220
collision rate [%]	2	4	12	16	24	39
$\bar{\lambda}$ at $in0$ [pkts/s]	20	31	53	67	81	96
mean $\tilde{H}$ for nodes	0.943	0.881	0.889	0.917	0.859	0.91
mean $\tilde{H}$ at $in0$	0.873	0.813	0.861	0.841	0.758	0.574

In the next figure (fig. 5) there are periodogram estimation plots for different number of nodes for desired $H = 0.9$. All estimation results of Hurst exponent were obtained based on the formula (8). It is clearly seen that the values of $\tilde{H}$ decreases as the number of nodes increases. This is due to the disappearance

of the LRD structure in aggregated stream that consists of many independent network flows coming from single nodes. It can be better observed in table 1, where the mean $\tilde{H}$ estimated for nodes stays the same (approximately), while the mean $\tilde{H}$ at the input of the queue ($in0$) decreases to the low value, suggesting that the LRD properties gradually disappears.

Table 2 shows the main statistics for different desired H values constant number of nodes ($nNodes = 5$). A slight increase of all measured values can be observed, which suggests that increasing H value raises the level of collisions and hence increases values for other statistics.

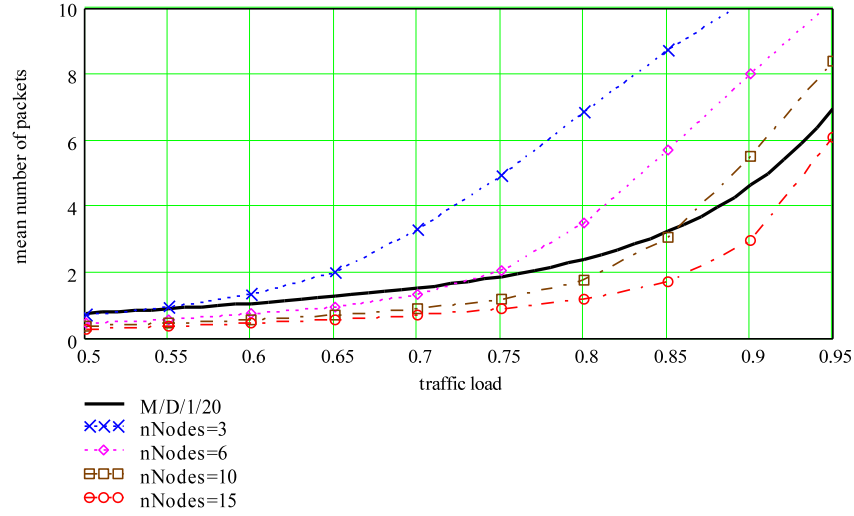


Fig. 6. Mean number of packets in the queueing system for different number of nodes, $\alpha = 1.2$.

Table 2. Main statistics for 5 nodes and different H values.

H value:	0.6	0.7	0.8	0.9
throughput [kbps]	180.5	181.8	183.8	189.5
packets at $in0$ [pkts]	158681	159830	161598	166602
collision rate [%]	7.9	7.8	8.1	8.5
$\bar{\lambda}$ at $in0$ [pkts/s]	40.6	40.8	44.5	48.4

The observations from table 1 are confirmed by the backhaul link performance results. In figure 6 there are curves for mean number of packets in the

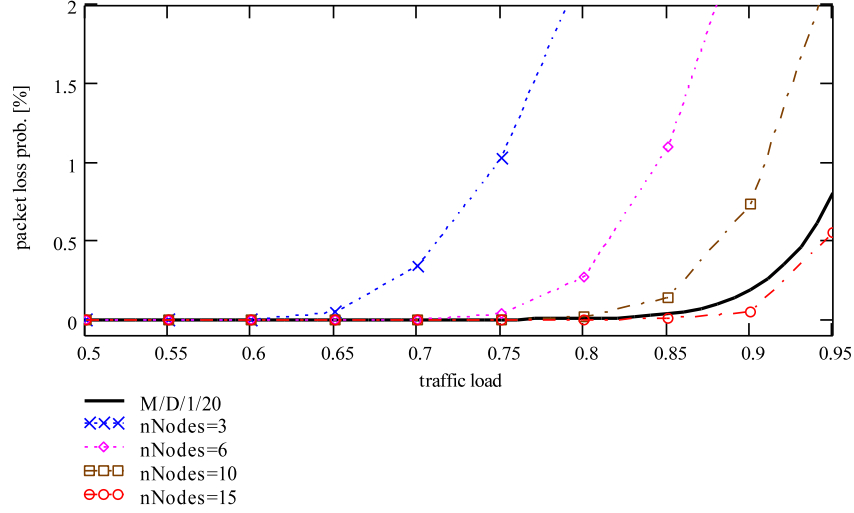


Fig. 7. Packet loss for different number of nodes, $\alpha = 1.2$.

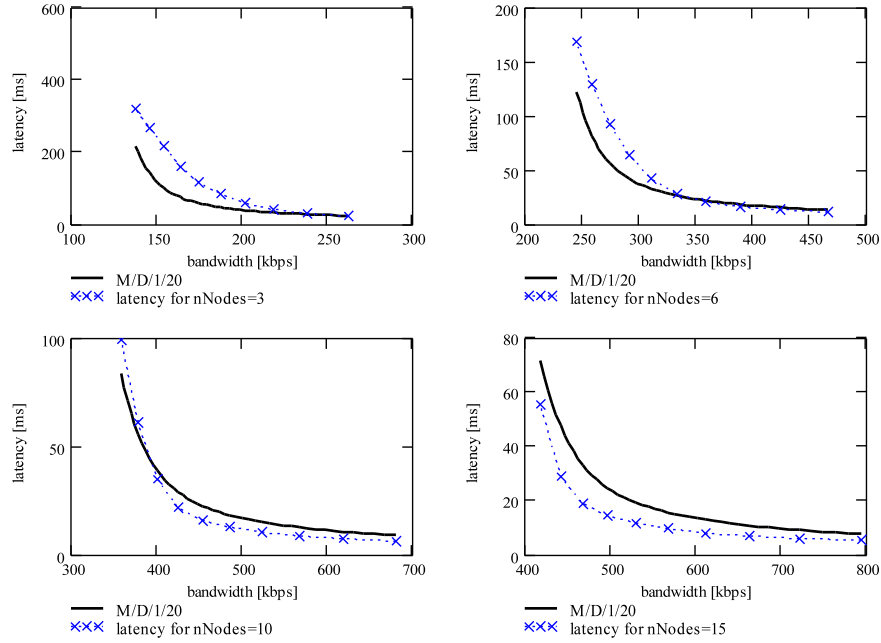


Fig. 8. Latency for different number of nodes compared to M/D/1/20 queueing system.

queueing system for different number of nodes versus offered traffic load. The theoretical curve for M/D/1/20 system, marked with solid black line, is calculated as the mean value of all $p_k^{(K)}$ in (13) for $K = 20$. In the next figure (7) one can observe the same tendency - the packet loss becomes lower as the number of nodes increases. Last figure (8) presents changing latency when number of nodes increases. For $nNodes = 10$ and $nNodes = 15$ the empirical curves goes below the levels of theoretical counterpart calculated from (11), which can be explained by the fact that the aggregation of streams from single nodes causes big change in LRD as well as in distribution.

5 Conclusions

The number of IoT devices and networks is constantly increasing, which means that congestion can occur, especially when the communication channel capacity does not increase. The main performance measures analyzed in this article were latency and packet loss. It is obvious that for higher offered traffic load the values of both measures increases causing poor performance of the connection link. However, situation becomes worse when LRD is considered. There is no doubt that this feature exists in network traffic. The question is, how strong are these long-term relationships and how they influence the performance. The simulation results of analysis of a typical IoT wireless CSMA/CA network with backhaul link provided more insights on the impact of both LRD and number of nodes on latency and packet loss.

All estimation results of Hurst exponent show that $\tilde{H}$ is stable for the same number of nodes. If number of nodes increases, then $\tilde{H}$ becomes smaller. Furthermore, all performance results show that with an increasing number of nodes, performance improves, causing even underestimation of the classical M/D/1/K model of queueing system. It implies that the aggregated stream consisting of many single node streams has changed its structure in terms of LRD feature as well as the distribution. This phenomenon can be used to determine the parameters of the IoT network system in order to reduce the value of latency and packet loss, which in turn has a positive effect on QoS and QoE.

References

1. Abry, P., Flandrin, P., Taqqu, M., Veitch, D.: Theory and applications of long-range dependence. Birkhuser Basel, 1st edn. (2002)
2. Ahmad Jan, M., Jan, S., Alam, M., Akhunzada, A., Rahman, I.: A comprehensive analysis of congestion control protocols in wireless sensor networks. *Mobile Networks and Applications* **23**, 456–468 (2018). <https://doi.org/10.1007/s11036-018-1018-y>
3. Al-Kofahi, M.M., Al-Shorman, M.Y., Al-Kofahi, O.M.: Toward energy efficient microcontrollers and internet-of-things systems. *Computers & Electrical Engineering* **79**, 106457 (2019). <https://doi.org/https://doi.org/10.1016/j.compeleceng.2019.106457>

4. Bai, S., Taqqu, M.: On the validity of resampling methods under long memory. *The Annals of Statistics* **45**, 2365–2399 (12 2017). <https://doi.org/10.1214/16-AOS1524>
5. Bhuyan, M.H., Bhattacharyya, D.K., Kalita, J.K.: Network anomaly detection: Methods, systems and tools. *IEEE Communications Surveys Tutorials* **16**(1), 303–336 (2014). <https://doi.org/10.1109/SURV.2013.052213.00046>
6. Dai, L.: Toward a coherent theory of csma and aloha. *IEEE Transactions on Wireless Communications* **12**, 3428–3444 (2013). <https://doi.org/10.1109/TWC.2013.052813.121605>
7. Dymora, P., Mazurek, M.: Anomaly detection in iot communication network based on spectral analysis and hurst exponent. *Applied Sciences* **9**(24) (2019). <https://doi.org/10.3390/app9245319>
8. Fortino, G., Gravina, R., Russo, W., Savaglio, C.: Modeling and simulating internet-of-things systems: A hybrid agent-oriented approach. *Computing in Science Engineering* **19**(5), 68–76 (2017). <https://doi.org/10.1109/MCSE.2017.3421541>
9. Gao, Y., Dai, L.: Random access: Packet-based or connection-based? *IEEE Transactions on Wireless Communications* **18**(5), 2664–2678 (2019). <https://doi.org/10.1109/TWC.2019.2906596>
10. Iglesias, F., Zseby, T.: Analysis of network traffic features for anomaly detection. *Machine Learning* **101**, 59–84 (2015). <https://doi.org/10.1007/s10994-014-5473-9>
11. Jing, X., Yan, Z., Jiang, X., Pedrycz, W.: Network traffic fusion and analysis against ddos flooding attacks with a novel reversible sketch. *Information Fusion* **51**, 100–113 (2019). <https://doi.org/https://doi.org/10.1016/j.inffus.2018.10.013>
12. Koseoglu, M., Karasan, E.: Energy-optimum throughput and carrier sensing rate in csma-based wireless networks. *IEEE Transactions on Mobile Computing* **13**(6), 12001212 (Jun 2014). <https://doi.org/10.1109/TMC.2013.124>
13. Le-Ngoc, T., Subramanian, S.N.: A pareto-modulated poisson process (pmpp) model for long-range dependent traffic. *Computer Communications* **23**, 123–132 (2000)
14. Ley-Bosch, C., Alonso-Gonzalez, I., Sanchez-Rodriguez, D., Ramirez-Casaas, C.: Evaluation of the effects of hidden node problems in iee 802.15.7 uplink performance. *Sensors* **16**, 216 (2016). <https://doi.org/10.3390/s16020216>
15. Li, G., Yan, Z., Fu, Y., Chen, H.: Data fusion for network intrusion detection: A review. *Security and Communication Networks* **2018**, 1–16 (2018). <https://doi.org/10.1155/2018/8210614>
16. Park, K., Willinger, W.: Self-Similar Network Traffic and Performance Evaluation. John Wiley & Sons, Inc., USA, 1st edn. (2000)
17. Purczynski, J., Wlodarski, P.: On fast generation of fractional Gaussian noise. *Computational Statistics and Data Analysis* **50**, 2537–2551 (2006)
18. Seo, D.W.: Explicit formulae for characteristics of finite-capacity M/D/1 queues. *ETRI Journal* **36**(4), 609–616 (2014). <https://doi.org/10.4218/etrij.14.0113.0812>
19. Shafiq, M., Ahmad, M., Afzal, M., Ali, A., Irshad, A., Choi, J.G.: Handshake sense multiple access control for cognitive radio-based iot networks. *Sensors* **19**(2), 241 (2019). <https://doi.org/10.3390/s19020241>
20. Smith, R.: The dynamics of internet traffic: Self-similarity, self-organization, and complex phenomena. *Advances in Complex Systems* **14**(6), 905–949 (07 2011). <https://doi.org/10.1142/S0219525911003451>
21. Varga, A.: OMNeT++, pp. 35–59. Springer Berlin Heidelberg, Berlin, Heidelberg (2010). https://doi.org/10.1007/978-3-642-12331-3_3
22. Varga, A.: OMNeT++ (accessed December 19, 2019), <https://omnetpp.org/>

23. Wang, J., Yang, X., Liu, Y., Qian, Z.: A contention-based hop-by-hop bidirectional congestion control algorithm for ad-hoc networks. *Sensors* **19**(16) (2019). <https://doi.org/10.3390/s19163484>
24. Wong, P.K., Yin, D., Lee, T.T.: Analysis of non-persistent CSMA protocols with exponential backoff scheduling. *IEEE Trans. Communications* **59**(8), 2206–2214 (2011). <https://doi.org/10.1109/TCOMM.2011.051811.100241>
25. Zhang, Y., Huang, N., Xing, L.: A novel flux-fluctuation law for network with self-similar traffic. *Physica A: Statistical Mechanics and its Applications* **452**(C), 299–310 (2016). <https://doi.org/10.1016/j.physa.2016.02.031>