RESEARCH ARTICLE

Validity of Exponential Distribution for Modelling Inter-failure Arrival Times of Windows based Industrial Process Control Data Exchange

[*] [iD]Ongun Yücesan, [1] [iD]Altan Özkil, [2] [iD]Mehmet Efe Özbek

[*]Atilim University, School of Civil Aviation, Ankara, Turkey.
[*] ongun.yucesan@atilim.edu.tr Orcid:0000-0003-2263-6803,
[1] altan.ozkil@atilim.edu.tr Orcid:0000-0001-8136-6087,
[2]Atılım University, School of Nat. & App. Sci., EEE, Ankara, Turkey.
efe.ozbek@atilim.edu.tr Orcid:0000-0001-5216-7062

HIGHLIGHTS

- *Investigates failure arrivals in a Windows Industrial Workshop Communication environment*
- *Among the rare observations for failure arrival times.*
- *Exponential can be used with reasonable error that paves way for Markovian Analysis.*

GRAPHICAL ABSTRACT

Using a reliability function with a known classical distribution, at an instance in time, the chances of an activity success can be predicted. Based on this curve, a proper reset period can be appointed. This would help to keep the risk under control. For this purpose, a test bed mimicking the machine to machine (M2M) communication of a Digital Oil Rig is employed. It contains an Embedded Computer Server and an ordinary computer acting as a client querying data from the server. On this testbed, as the communication protocol Open Platform Communications - Unified Architecture (OPC UA) is run. In "most occasions, the arrival processes are assumed as a Poisson Arrival Process. The Question of a concern: "Is this assumption valid in case of an Embedded Computer responding to queries?". Our results indicate that the failure arrivals are independent identically distributed (IID) and can be modeled with an Exponential Distribution by accepting a reasonable error.
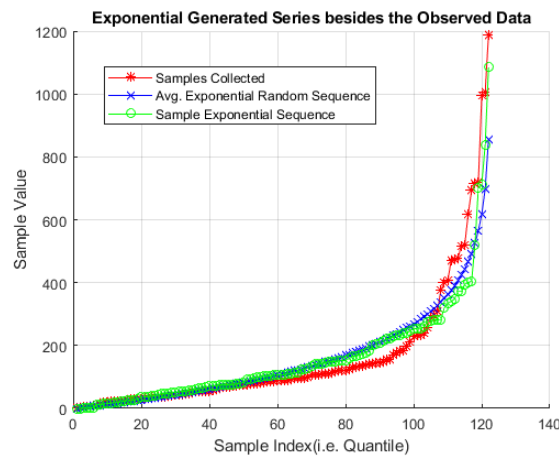
Figure A. or Table A The name figure or table about given info and results

Aim of Article: *To illustrate that the failure arrival times can be according to a known distribution.*

Theory and Methodology: *Observations on physical peer to peer system is performed for Industrial Workshop communication protocol running on Windows. The number of repeated actions is observed and recorded as failure arrival times. These arrival series are investigated with Goodness of Fit tests for their stochastic character.*

Findings and Results: *It is quite possible with reasonable error to model inter-failure arrival times as Exponentially distributed, whereas, for average ranges Gaussian is useful as well.*

Conclusion: *The failures of an windows industrial workshop software communications can be modelled with Markovian Arrival Processes.*

RESEARCH ARTICLE

Validity of Exponential Distribution for Modelling Inter-failure Arrival Times of Windows based Industrial Process Control Data Exchange

*[iD] Ongun Yücesan, [1] [iD] Altan Özkil, [2] [iD] Mehmet Efe Özbek

*Atilim University, School of Civil Aviation, Ankara, Turkey.
* ongun.yucesan@atilim.edu.tr Orcid:0000-0003-2263-6803,
[1] altan.ozkil@atilim.edu.tr Orcid:0000-0001-8136-6087,
[2] Atılım University, School of Nat. & App. Sci., EEE, Ankara, Turkey.
efe.ozbek@atilim.edu.tr Orcid:0000-0001-5216-7062

H I G H L I G H T S

- *Investigates failure arrivals in a Windows Industrial Workshop Communication environment*
- *Among the rare observations for failure arrival times.*
- *Exponential can be used with reasonable error that paves way for Markovian Analysis.*

Article Info	ABSTRACT
  *Corresponding Author: Dr. Ongun Yücesan ongun.yucesan@atilim.edu.tr	Using a reliability function with a known classical distribution, at an instance in time, the chances of an activity success can be predicted. Based on this curve, a proper reset period can be appointed. This would help to keep the risk under control. For this purpose, a test bed mimicking the machine to machine (M2M) communication of a Digital Oil Rig is employed. It contains an Embedded Computer Server and an ordinary computer acting as a client querying data from the server. On this testbed, as the communication protocol Open Platform Communications - Unified Architecture (OPC-UA) is run. In "most occasions, the arrival processes are assumed as a Poisson Arrival Process. The Question of a concern: "Is this assumption valid in case of an Embedded Computer responding to queries?". Our results indicate that the failure arrivals are independent identically distributed (IID) and can be modelled with an Exponential Distribution by accepting a reasonable error.

I. INTRODUCTION

Industrial Internet of Things (IIoT), and Machine to Machine (M2M) communications concepts include research for devices to be able to collaborate with other entities. When there are humans in the cycle, the reliable operations are related to their safety. The verification and validation activities in form of functional tests can be a method for tackling the issue.

The reliability, when there is a single core CPU, can be dependent on the utilizations. When maximum utilization is reached, resources to address any other functionality is not there. Besides, the reliability can be dependent on "how often and how many repetitions of the same function takes place?". The executions of the same command may not result with success always. By observing inter-failure durations, a reliability function can be identified. Under realistic scenarios, a company would be able to establish

guarantee periods and maintenance schedules based on this data. For this instance, a series of experiments targeting to collect the inter-failure durations are conducted.

The tools employed in the study are Commercial of the Shelf (COTS) Window based OPC UA server and a Prosys Client Modified to support an HMI. These are running on a 1.4 GHz Celeron Server and an i7 Desktop computer in respective ordering. The performance aspects of OPC UA Server-Client architectures are widely investigated. As well as, the latencies are considered in many studies.

Among the previous studies about performance bottleneck identifications, Burger et al. [1]. Cavalieri et al. [2] studies the end-to-end delay performance of OPC UA suites. Morato et al. [8] indicates there may be further time delays in cases of Real Time operating systems just to regulate the process period.

Eckhardt et al. [3] studies the round-trip time of OPC UA Server and client architecture. Their Xilinx systems use a dedicated ARM Real Time Processor for performing Ethernet communication tasks. Cenedese et al. [4] compares the RTT performances of open source OPC UA servers and clients, where C based open62541 is identified as more efficient. Kim et al. [5] indicates that there is an optimum point for the highest transferable data rate with respect to the query interval. Unix, Linux and Windows seem to be equivalently open to the Denial-of-Service attacks (DoS) which have been studied by Neu et al. [6] for OPC UA.

If a Markovian Arrival Process is considered, the number of arrivals within a time duration are accepted as Poisson distributed. In this circumstance, the time duration that is needed after an arrival for a new arrival to take place is Exponentially distributed. That is a subcase of Erlangian distribution for k=1. Event is conditioned on a previous arrival had taken place or experiment had started, following a Markovian Character. Garcia-Mora et al. [11] reviews a situation when there are additional significant conditions affecting the duration other than just simple arrival of the previous sample, as an answer to question whether inter-failure times are stationary. Yucesan et al. [9] presents an asymptotic expectation on test duration for a given test suite. This could give an idea about how long a test with given count of test cases could take place. This analysis can be broadened to include variety of failure arrivals. Also, Yucesan et al. [10] present a preceding study investigating effects of increasing load and query periods on reliability.

The rest of the paper is organized as follows: in Section II. Experimental results taken from the OPC-UA testbed are provided, as well as, discussions over them in Section III. Finally, the paper is concluded in Section IV.

II. EXPERIMENTAL RESULTS

A continuous block of data transfer, which is called Historic Data Access (HDA), is employed for this effort. For instance, we monitor failure arrivals on Commercial of the Shelf (COTS) Windows OPC UA platform. Two observation series will be investigated for how all samples statistically could have come from the same distribution. First comes from Mean time Between Failure (MTBF) observations. The second is obtained by gathering all observations.

A. MTBF Observations and Candidate Random Distributions

MTBF data placed into a sorted series. To compare the MTBF series, same size pseudo-random exponential sequences are presented in an averaged and an example format. These sequences are sorted and rounded to the nearest integer. We can see them in Figure 1. The purpose is to find out if the underlying distribution of observed series are suiting to an Exponential or another random variable, see Trivedi et al. [7].
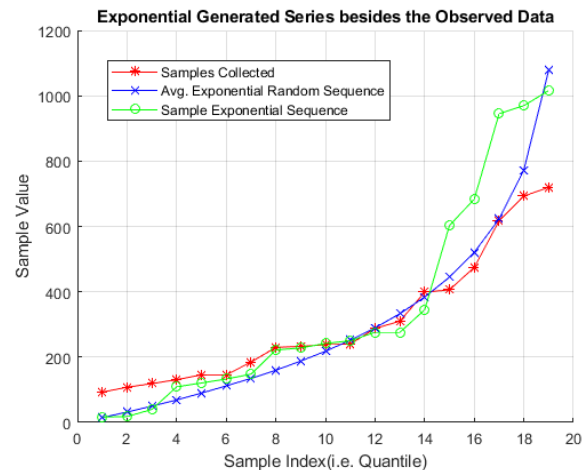

Figure 1 Collected samples sorted manner along with Sorted Series of generated Random Numbers according to an Exponential Random Variable. An example and averaged series.

From the figure, it is seen that the samples collected do not start from zero (0) level. They have a threshold above this value. Since the data are collected by

continuing the experiment till to a point, where the transmitted window remains empty, a threshold form. If we look at them side by side, the lower quantiles or indices of exponential random specimens have lower sample values with regard to the collected data. The higher quantiles or indices for exponential have higher values. Mid ranges have better fit among the series. For better understanding the underlying character, further examinations will be performed in the following sections.

B. Distribution Comparison with QQ plots

In this section, the results of comparisons which have been made using available QQplot functionality of MATLAB is presented. The consistently collected MTBF results are compared to the various series obtained by using the random number generator of MATLAB through variety of candidate stochastic distributions. Once again sorted series are compared.
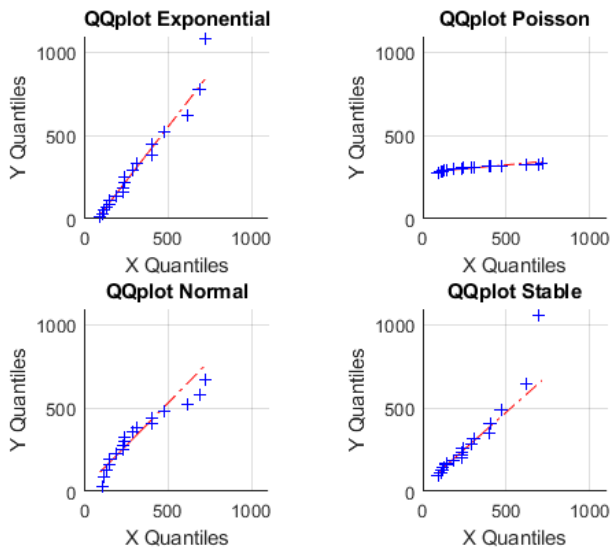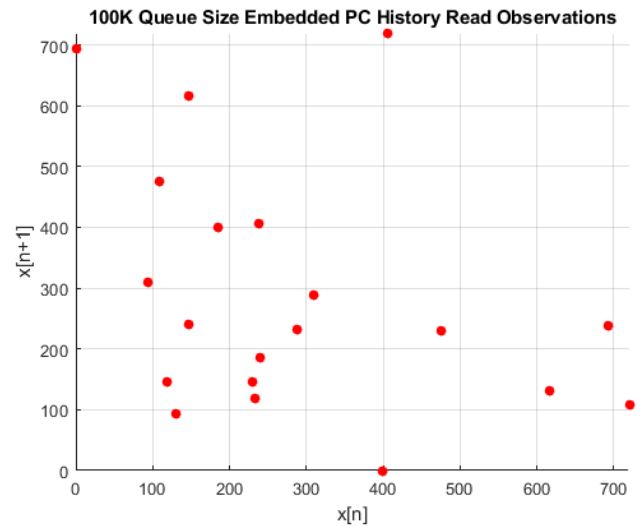

Figure 2 Plots for comparisons with Exponential, Poisson, Normal and Stable Distributions.

For this instance, the sample mean μ= 304.0526 and the standard deviation σ = 197.4078 from the consistent MTBF series was calculated. The series used for comparison are the samples from the Exponential, Poisson and Normal pseudorandom generated series regarding to the obtained mean and standard deviation. The Quantile to Quantile (QQ) plots are presented in Figure 2. X-axis corresponds to the samples of the observed experimental results. The y-axis corresponds to the quantiles of distributions generated with pseudo random number generator. Figure is drawn as (x,y) ordered pairs. The stable

distribution characteristic equation is as defined in Equation (1). Others are classical known distributions. The parameters for the Stable distribution are α= 1.02024, β= 1.0, δ= 72.4507 and γ= 190.45.

$$E(e^{(itX)})$$
$$= \begin{cases} exp\left(-\gamma^{\alpha}|t|^{\alpha}\left[1 + i\beta sign(t)tan\frac{\pi}{2}((\gamma|t|)^{1-\alpha})\right] + i\delta t\right), \alpha \neq 1, \\ exp\left(-\gamma|t|\left[1 + i\beta sign(t)tan\frac{\pi}{2}(\gamma|t|)\right] + i\delta t\right), \alpha = 1 \end{cases} \quad (1)$$

Figure 3 XY Scatterplots drawn with x[n] versus x[n+1]
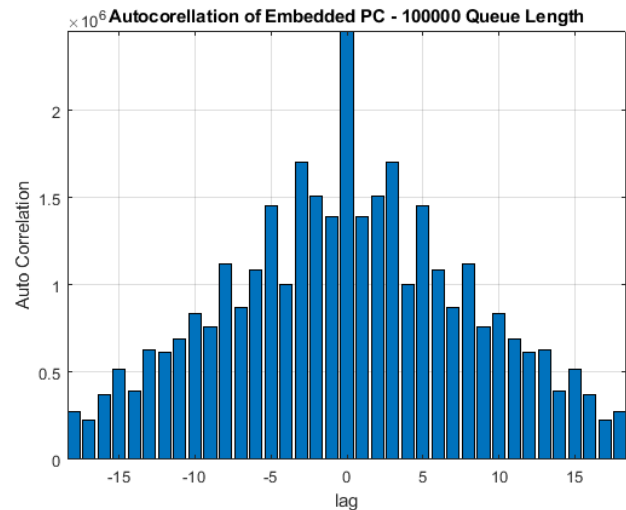
Figure 4 The auto-correlation plot of the Independent Identical Distribution (IID) property

In the QQ plots, the Exponential distribution presents a linear manner fit with an acceptable error. There is a shift towards right, representing the threshold of the consistent MTBF series. The Poisson is generally used for modelling the probabilities of the

accumulated chances of arrivals, which is not the case in the examination perspective. The Gaussian distribution presents a good match around the mid-range values. Early quantiles from the Gaussian series have negative sample values. The tail also presents fluctuations. The Stable distribution is coming from the MATLAB Distribution Fitter. It is more suitably crafted for the specifics of the MTBF series observations, rather than general characteristics of the experiment.

C. Is the sequence Independent Identically Distributed?

In the scatter graph, obtained in respective order of the observation from the MTBF series, as observed from Figure 3 there are no evident patterns, other than an insignificant alignment between 200 to 300 values. When the autocorrelation results in Figure 4 are considered for consistent MTBF series, with the first shift there is a significant fall. This emphasizes the independence among samples. For a shorter series, the values obviously have limited difference. The level keeps on falling slowly. This would be due to reasonable amounts of errors. The oscillations and slow behavior represent an identical structure.

D. All Embedded PC Observations

The data considered for this section is an all-inclusive observation set from the experiments. It even includes early phases' results and outcomes, when the test scenario was not mature, included all sorts of different working conditions.

In Figure 5, we can observe the red asterisk pointed line as all samples obtained during experimentation. The highest duration observed is 1200 sample valued experiment result with index 122. There is also a 1005 sample valued time with 120th indexed outcome. This value was observed during a later experiment with 12 variables. The observation is way above the average in the consistent MTBF series that could have appeared within this series as well.

The green line with circle points is an example exponential series, where blue crossed pointed lines are averaged exponential series. The Figure 5 also indicates that an exponential generated random number sequence tail resembling to the observations could have had happened in reality, as green line is very close to red observations series.

The QQ plots available in Figure 6 tells that the early samples of exponential present very good match with the observation series. When middle ranges

reached, a second line is observed for exponential. Omitting Poisson, the Normal has similar character.
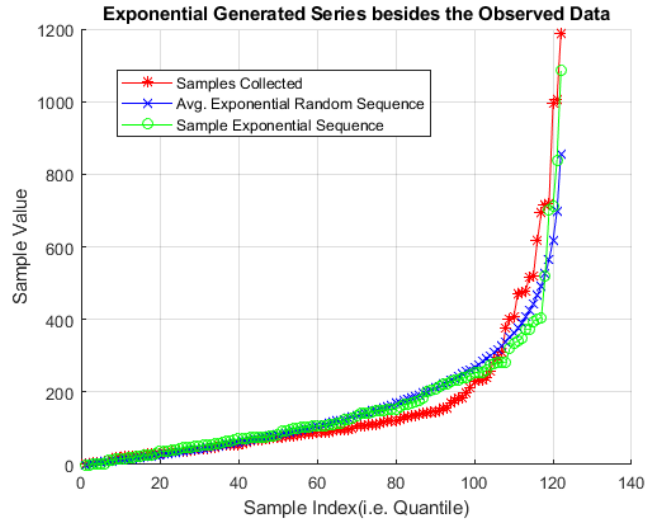

Figure 5 Sorted Series of observed Mean Time Between Failure (MTBF) Results for all samples
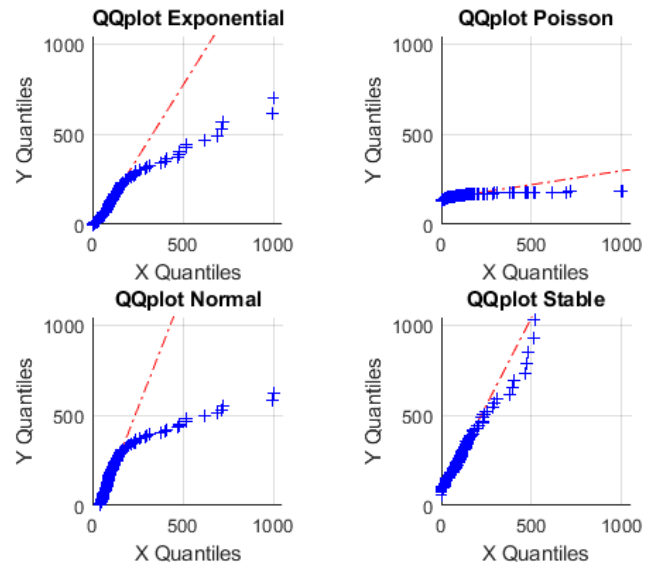
Figure 6 Plots for comparisons with Exponential, Poisson, Normal and Stable Distributions over all sample

However, the shift towards right is due to the early negative sample valued quantiles of the Gaussian pseudo random series. Even though *Stable* sequence looks like matching almost perfectly to the all-inclusive set, as can be observed from the Figure 7, it has excessively high values for the tail part as represented with red lines. This figure includes averaged sequence over repeated generation of the series of random numbers for more than 120 experimentations with pseudorandom generator. The blue line in this figure is for the Gaussian, the green

is the Exponential, where almost flat black dotted line for the Poisson random number generators. It reflects specifically crafted character of the Stable again regarding just the available set of samples for MTBF series. In the Scatter graph presented in the Figure 8, the scatter graph has no evident patterns. The Autocorrelations available in Figure 9 shows that there is a great fall with the first shift. There are also oscillations and eventually correlation slightly fade. This reflects the identical behavior and also some imperfections. Since level is below a threshold, it is reasonable to indicate that the series has IID property acceptable within error tolerances.
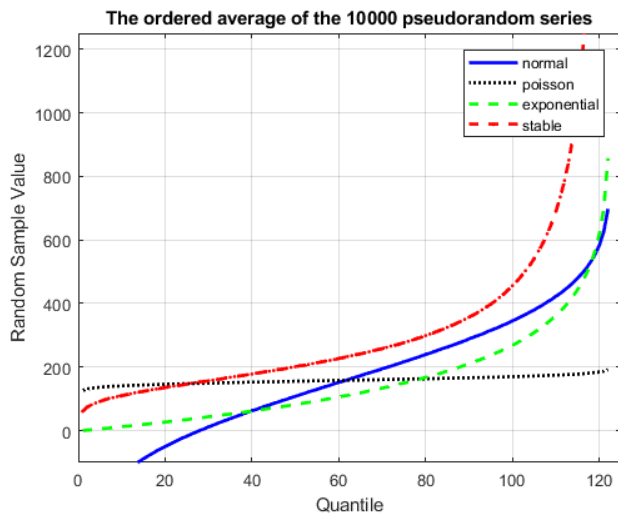

Figure 7 Ordered Samples from pseudo random series generated according to Normal, Poisson, Exponential and Stable distributions over all samples from Embedded PC

### E. Distribution Comparisons with Goodness of Fit Tests

There are two set of tables available for this section. First of them is on the results obtained from the consistently obtained set. Second is the set of all included set of observations. Since all embedded PC samples have IID property and include the consistent set, the results are comparable to each other by accepting certain errors and statistical differences.

These are referred to as Manual K-S and MATLAB K-S respectively to the ordering in Table I-A. The Manual K-S in the table indicates 10% significance for not being able to reject the Exponential distribution and 20% significance for the Gaussian random variable. However, it rejects the Stable distribution, because the manual implementation does not exclude the tail samples that are off. Omitting some samples is common in statistical techniques.

This is reflected in the results for the MATLAB K-S tests, they mention Stable distribution generated one have the best fit with 35% significance. The Gaussian and the Exponential are at 10% and 30% significance. The MATLAB A-D indicates similar results, 15%, 35% and 40% significance for the Exponential, Gaussian and the Stable respectively. Where it is 20%, 25% and 40% for χ^2 test with same respective ordering.
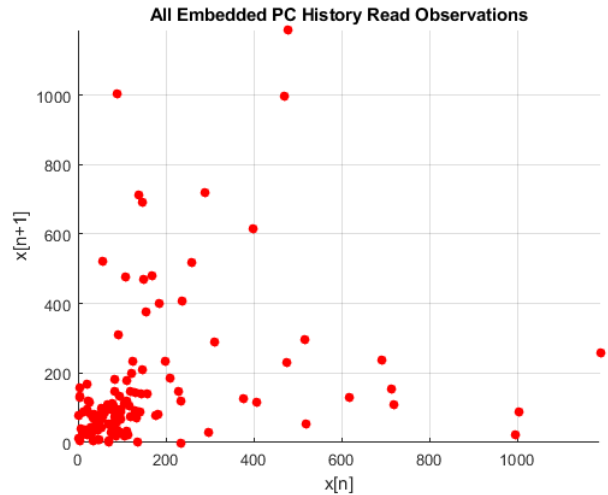
Figure 8 XY Scatterplots drawn with x[n] versus x[n+1]
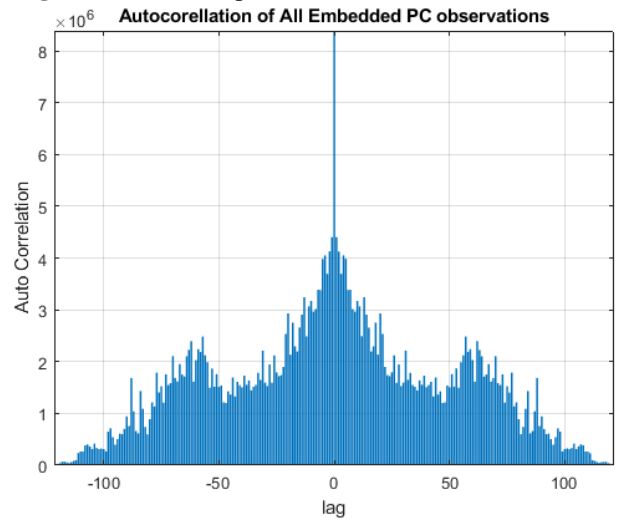
Figure 9 The auto-correlation plot of the Independent Identical Distribution (IID) property for all observed samples

When considering the Table I-B, we observe that for samples collected from all experiments, even the significance levels are low, one cannot deny the existence of exponential character. The significances for these results are with 2%, 1%, 1% and 0.5% for Manual K-S, Matlab K-S, A-D and χ^2 tests with respective order.

Goodness of Fit Test Results			
Distrib.	Manual K-S	Matlab K-S	MatlabA-D	χ2
Expon.	10% Sign.	10% Sign.	15% Sign.	20% Sign.
Gauss	20% Sign.	30% Sign.	35% Sign.	25% Sign.
Stable	Null H. Rej	35% Sign.	40% Sign.	40% Sign.

A

Goodness of Fit Test Results			
Distrib.	Manual K-S	Matlab K-S	MatlabA-D	χ2
Expon.	2% Sign.	1% Sign.	1% Sign.	0.5% Sign.
Gauss	Null Rej	Null Rej	Null Rej	Null Rej
Stable	Null Rej	Null Rej	Null Rej	Null Rej

B

Table 1 *The results of the Goodness of Fit Tests with respect to Exponential, Normal, Stable distributions tested with Kolmogorov-Smirnov Manually and on MATLAB, Anderson Darling on MATLAB, and χ2 on MATLAB*

III. DISCUSSIONS

The consistent MTBF series, since its samples were collected in disciplined manner involved a threshold. This series did not have any immature test starts. As additional data were collected in the later stages of the research effort, some sample values unexpectedly arrived alike 1005. This reaffirms that as number of observations increases, the confidence to the results increases. Furthermore, that with series distributed according to an exponential random variable, any value for the outcome is possible. By results of the goodness of fit tests, QQ-plots, scatter and autocorrelation graphs, the IID property and exponential underlying distribution was found to be undeniable to certain significances. The generality of the samples is accumulated in the smaller values, emphasizing also that the exponential has its most power in the smaller sample values. Therefore, character of the existing samples is determined to a degree. The Stable Distribution sequence rather than capturing possible subsequent outcomes, it offers predictions for the data present at hand. Early quantiles of this distribution have a better fit while the tail has evident deviations.

The goodness of fit tests indicates Gaussian can be of a choice. It can be observed in the Figure 7 as well, that for the middle ranges, Gaussian and Exponential variables approaches each other. Therefore, even though goodness of fit tests would not give a direct result, for middle ranged conditions like MTBF, the Gaussian random variable can bring ease to reliability predictions.

Since the resets are almost perfect, if not leaving a heat trace from a previous run, the IID property and theoretical matches of Exponential distribution are reasonable for an Embedded computer condition. However, heat such as room temperature, heat of the CPU, how much aimed to be alleviated with uses of the fans, may be a reason of imperfections on IID results. Since it's granted that a stationary series is to have autocorrelation with only central lag is significant, it is not necessarily true that the auto correlation following this rule grants stationary property. In our observations, in some series slightly Near-Term Dependence, which is the case for Wide Sense Stationary signals are the case. We accept this situation within scope of reasonable error.

Through the study, the concentration has been on the exponential inter-arrival times and their IID character. It may be a good future work to investigate independent identical or homogenous character of the Poisson Arrival Process or an Erlangian Arrival Process based on accumulated result of times. Our study presented early results of reliability in this field, which could be useful in almost any Industrial Workshop communication where clients are most of the time lacking means of many resources. These communication mechanisms are, as well, used for some aeronautical vehicles and space baring student probe rockets [12] requiring some rigorous availabilities.

IV. CONCLUSION

The observations of 304 read counts corresponding to roughly 76 minutes of MTBF is tiny for 7/24 intense operations. Same value is 159.19 reads with all samples included. The Celeron 1.4 Ghz Embedded PC computer is for lot simpler tasks than bulky continuous transfers of historic data with lacking cache structure not being able to utilize CPU to its maximum extend.

Per the observations on Failure Arrivals in an M2M environment, with the data at hand, it is possible to tell classical assumption of modeling any arrival process with Exponential inter-arrival times is valid also for Failure Arrivals with OPC UA Client/Server Architecture History Reads as well. It is therefore possible to predict the outcomes of the future experiments over the given scenario.

CONFLICTS OF INTEREST

They reported that there was no conflict of interest between the authors and their respective institutions.

RESEARCH AND PUBLICATION ETHICS

In the studies carried out within the scope of this article, the rules of research and publication ethics were followed.

REFERENCES

[1] A. Burger, H. Koziolek, J. Ruckert, M. Platenius-Mohr, and G. Stomberg, "Bottleneck identification and performance modeling of opc ua communication models," in Proceedings of the 2019 ACM/SPEC International Conference on Performance Engineering, ser. ICPE '19. New York, NY, USA: Association for Computing Machinery, 2019, p. 231–242.

[2] S. Cavalieri and F. Chiacchio, "Analysis of opc ua performances," Computer Standards and Interfaces, vol. 36, no. 1, pp. 165 – 177, 2013.

[3] A. Eckhardt and S. Muller, "Analysis of the round-trip time of opc ua and tsn based peer-to-peer communication," in 2019 24th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA), Sep. 2019, pp. 161–167.

[4] A. Cenedese, M. Frodella, F. Tramarin, and S. Vitturi, "Comparative assessment of different opc ua open–source stacks for embedded systems," in 2019 24th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA), Sep. 2019, pp. 1127–1134.

[5] W. Kim and M. Sung, "Opc-ua communication framework for plc-based industrial iot applications: Poster abstract," in Proceedings of the Second International Conference on Internet-of-Things Design and Implementation, ser. IoTDI '17. New York, NY, USA: Association for Computing Machinery, 2017, p. 327–328.

[6] C. V. Neu, I. Schiering, A. Zorzo. Simulating and Detecting Attacks of Untrusted Clients in OPC UA Networks. CECC 2019: Proceedings of the Third Central European Cybersecurity Conference. 2019, p.1-6.

[7] K. S. Trivedi, Probability and Statistics with Reliability, Queuing and Computer Science Applications, 2nd ed. GBR: John Wiley and Sons Ltd., 2001.

[8] A. Morato, S. Vitturi, F. Tramarin and A. Cenedese, "Assessment of Different OPC UA Implementations for Industrial IoT-Based Measurement Applications," in IEEE Transactions on Instrumentation and Measurement, vol. 70, pp. 1-11, 2021

[9] O. Yucesan, A. Ozkil, "Time Complexity Comparison of Stopping at First Failure and Completely Running the Test." J Electron Test vol. 36, pp. 409–417, 2020.

[10] O. Yucesan, A. Özkil and E. Özbek , "A Reliability Assessment of an Industrial Communication Protocol on a Windows OS Embedded PC for an Oil Rig Control Application", Journal of Scientific, Technology and Engineering Research, vol. 2, no. 2, pp. 22-30, Dec. 2021.

[11] B. García-Mora, C. Santamaría Navarro, G. Rubio, Modeling dependence in the inter–failure times. An analysis in Reliability models by Markovian Arrival Processes. Journal of Computational and Applied Mathematics. 343. (2018).

[12] J. Matevska, E. Noack, M. Reinhold, and E. Diekmann, Decentralized Avionics and Software Architecture for Sounding Rocket Missions. In Proc. 2nd Workshop on Avionics Systems and Software Engineering Innsbruck, Austria, February 25, 2020.