How to Measure the Average and Peak Age of Information in Real Networks?

Basel Barakat
School of Engineering
University of Greenwich
Kent, United Kingdom
bb141@gre.ac.uk

Hachem Yassine
Department of Engineering Science
University of Oxford
Oxford, United Kingdom
hachem.yassine@eng.ox.ac.uk

Simeon Keates
School of Engineering and the Built Environment
Edinburgh Napier University
Edinburgh, UK
S.Keates@napier.ac.uk

Ian Wassell
Computer Laboratory
University of Cambridge
Cambridge, United Kingdom
ijw24@cam.ac.uk

Kamran Arshad
Electrical Engineering Department
Ajman University
Ajman, UAE

***Abstract*—The Age of information (AoI) was proposed in the literature to quantify the freshness of information. The majority of the work done in this area has theoretically evaluated AoI and its Peak (PAoI). In this paper, a method for obtaining the value of AoI and PAoI from experiments is proposed. We conducted an experiment emulating an M/M/1 queue and used the proposed method to evaluate AoI and PAoI. The values were compared to the expressions presented previously in the literature. Our results show that the proposed method is accurate for the M/M/1 queue. A statistical test was conducted to confirm the reliability of this conclusion.**

I. Introduction

Over the decades, continuous breakthroughs in communication technologies gave birth to a range of applications with different requirements. Many Internet of Things (IoT) applications are based on receiving updates about the status of a remote agent to help in decision making. Examples include wireless channel estimation [1], telehealth, environmental and industrial monitoring, and military battlefield coordination. For some of these applications, it is crucial that, at any point in time, the status that the decision maker has, is up to date and represents as much as possible the current status at the source. There is a subtle difference between this requirement and the traditional low latency (delay) requirement because the latter is seen purely from the perspective of network performance, while the former is seen from the destination's perspective. In other words, low-latency is not equivalent to the freshest updates at the receiver; it very much depends on how frequently the updates are being generated at the source.

To address this requirement, a new metric called Age of Information (AoI) was introduced in [7] to measure and quantify the freshness of information from the receiver's perspective. It is defined as the time since the last update received was generated. The main difference between AoI and conventional network delay metrics is that AoI is observed from the receiver's perspective, while the delay is observed from the network's perspective. The Peak Age of Information (PAoI), introduced in [3], is another metric related to AoI and represents the worst case AoI. It is defined as the maximum time elapsed since the preceding piece of information was generated. The PAoI metric has a simpler formulation and is a more mathematically tractable metric [3]. Consequently, modeling PAoI has gained attention in the literature such as in [3], [4]. Also, minimizing the PAoI by optimizing network functions was extensively investigated [5].

However, the majority of work on AoI and PAoI has been theoretical and assumed simple queuing models to derive theoretical results about these metrics. In [6], an emulation-based validation of the theoretical models was presented. More recently, [11] presented experimental results that validated the non-monotonous nature of AoI as a function of link utilization. However, in both papers, no clear explanation was provided as to how exactly the metrics were evaluated from the experiment. This paper aims to bridge this gap and provide a clear and concise tutorial for experimental researchers that wish to evaluate these metrics on real networks. The contributions of this paper can be summarized as follows:

- we provide an intuitive formulation of how the AoI and PAoI metrics can be estimated from the recorded time-stamps in an experiment,
- we validate these expressions by performing an experiment comprising an $M/M/1$ queue, and
- we present a simple methodology for conducting experiments.

The rest of the paper is organized as follows. In section II, we define the new metrics and some related quantities. In section III, we present the proposed method to estimate the delay, AoI, and PAoI. In section IV, we present a case study to validate our method and we compare the proposed estimates against their theoretical counterparts. We conclude the paper in section V.

II. Definitions and Previous Work

A. Age of Information

Consider a destination and an information source that is generating updates at discrete times (possibly by sampling a process) and then instantaneously transmitting them to the destination through a communication network. We denote by t_i the time at which the ith update was generated/transmitted at the source and by r_i the time at which it was received at the destination. We define $X_i = t_i - t_{i-1}$ as the time between the generation of updates i and $i-1$, i.e., the updates inter-arrival time. We also define the delay time (system) $T_i = r_i - t_i$ as the time it took, from the generation of ith update, until its reception at the destination. The Age of Information (AoI) at time t, denoted $\Delta(t)$, is defined as the time elapsed since the last received update at the destination was generated at the source. Mathematically, it is given by $\Delta(t) = t - u(t)$, where $u(t)$ is the generation time of the last received update at time t. Fig. 1 illustrates an example of how the information age evolves with time as a sawtooth function [7].

B. Time Average Peak AoI

For time interval $(0, \mathcal{T})$, where $\mathcal{T}$ is assumed, for simplicity, to coincide with the receipt of the nth update, i.e., $\mathcal{T} = r_n$, where r_n is the time that the last piece of information, n, was received in the time interval. The average time delay in this period can be written as

$$T_{\mathcal{T}} = \frac{1}{n}\sum_{i=1}^{n} T_i. \tag{1}$$

The peak age of an update is its age at the time of receipt of the next update [3], i.e., the peak age of the ith update is $\Delta(r_{i+1}) = X_i + T_i$. Therefore, it is possible to define the time-average peak AoI in the period $\mathcal{T}$ as follows

$$P_{\mathcal{T}} = \frac{1}{n-1}\sum_{i=1}^{n-1} \Delta(r_{i+1}). \tag{2}$$

C. Time Average AoI

The time average AoI in the interval $(0, \mathcal{T})$, denoted $\bar{\Delta}_{\mathcal{T}}$, is the area under the sawtooth function normalized by the observation period ($\mathcal{T}$), and it is given by

$$\Delta_{\mathcal{T}} = \frac{1}{\mathcal{T}}\int_0^{\mathcal{T}} \Delta(t)\, dt. \tag{3}$$

By a geometric argument, the area under the curve can be re-written as the sum of the of the areas in Fig. 1. Starting from $t = 0$, these are the areas of the polygon Q_1, trapezoids Q_i for $2 \le i \le n$, and the triangle of length T_n between t_n and r_n. the first update ($i = 1$) shown in Fig. 1. Hence,

$$\Delta_{\mathcal{T}} = \frac{Q_1 + \sum_{i=2}^{n} Q_i + T_n^2/2}{\mathcal{T}} \tag{4}$$

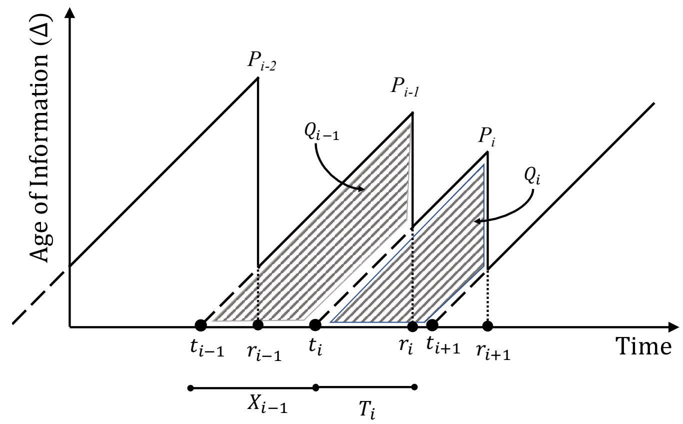

Fig. 1. Age of Information as a function of time. The updates inter-arrival time are referred to as X_i and the delay (system time) is T_i, i.e., the service time plus the queuing time. The PAoI of information i is represented by P_i. The time of generating update i is t_i and the time of receiving it is r_i.

The trapezoid area Q_i can be written as the area difference between the right isosceles triangles of legs $T_i + X_i$ and T_i, respectively. This is

$$Q_i = \frac{1}{2}(T_i + X_i)^2 - \frac{1}{2}T_i^2 \tag{5}$$

$$= T_i X_i + \frac{X_i^2}{2}. \tag{6}$$

In the literature, there is an interest in the stationary case when $\mathcal{T} \to +\infty$. It can be seen that the delay time converges to $\mathbb{E}[T]$ and the peak average AoI converges to

$$P = \lim_{\mathcal{T}\to+\infty} P_{\mathcal{T}} = \mathbb{E}[X + T] \tag{7}$$

while it can be shown [7] that the average AoI converges to

$$\Delta = \lim_{\mathcal{T}\to+\infty} \Delta_{\mathcal{T}} = \lambda\Big(\mathbb{E}[XT] + \frac{\mathbb{E}[T^2]}{2}\Big), \tag{8}$$

where X is the random variable representing the inter-arrival time of updates at the source (generation) with its rate λ, and T is the random variable representing the system time (delay) of an update. The expectations in the expressions depend on the network. To abstract the details of the underlying network, it is common to assume idealized queuing models such as the M/M/1, M/D/1, D/M/1 [3], [7]. Each model makes a different assumption about the update generation (X) and system time (T) which is composed of waiting time in the queue and service time in the network.

III. Estimation of the Metrics from Experiments

Consider a setup in which the source transmits an update at generation time (t_i) and that the receiver records its time of receipt (r_i). It is recommended that all calculations be performed after the end of the experiment as the time to calculate the metrics might affect the accuracy of the logged timings.

The delay that the ith update exhibits is calculated by,

$$T_i = r_i - t_i. \tag{9}$$

The expected delay of (9) can be estimated using the sample median of the vector that contains all the delays of all the updates, i.e.,

$$T_{(1\to n)} = \Big[T_1, T_2, ..., T_n\Big], \tag{10}$$

where n is the total number of updates communicated. The sample median is employed instead of the sample average because in some cases, the network protocol might re-transmit some packets, as in TCP/IP and HARQ protocols, if they suffered from significant errors. The re-transmission will increase the delay time hence would significantly affect the mean value.

Similarly, the average PAoI can be estimated using the sample median of the vector containing the PAoI of all the status updates communicated in the experiment, i.e.,

$$P_{(1\to n)} = \Big[P_1, P_2, ..., P_n\Big]. \tag{11}$$

To use the initial definition $P_i = X_i + T_i$, requires first finding the inter-arrival time and the delays. In the following, we provide an easier formulation using only the times of generation and receipt as shown in Fig. 2. In the figure, it is clear that the peak age of an update stretches from its time of generation until the time of receipt of the next update. Hence, P_i can be evaluated as

$$P_i = r_{i+1} - t_i. \tag{12}$$

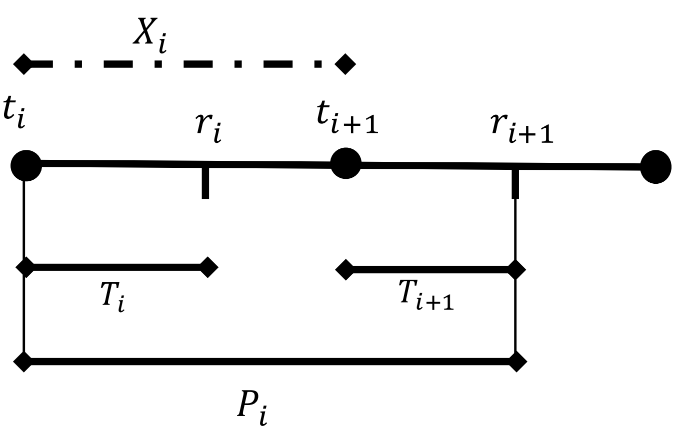

Fig. 2. Peak Age of Information measuring method illustration. Where t_i is the time that information i was generated, r_i is the time in which information i was received by the server, where X_i represents the updates inter-arrival time. The Peak Age can be considered as the difference between the time of receiving the next update and the time of generating the update.

Finally, to estimate the average AoI the areas of Q_i should be calculated. By using the area of a trapezoid it can be shown that

$$Q_i = \frac{1}{2}\bigg(\Big((t_{i+1} - t_i) + (r_i - t_i)\Big)^2 - \Big(r_i - t_i\Big)^2\bigg). \tag{13}$$

The average AoI can then be estimated as follows

$$\Delta = \frac{\sum_{i=2}^{n} Q_i}{(r_n - t_2)}. \tag{14}$$

Therefore, in this section, we can express the metrics of interest in terms of the observable time-stamps coming from an experiment.

IV. Tested Case Study

In this section, we validate our estimators by emulating an $M/M/1$ network and comparing the measured to the theoretical results. We start by reviewing the theoretical results for $M/M/1$; then we give our validation setup before presenting and discussing the results.

A. M/M/1 Queue

The method of measurement was tested on an M/M/1 queue, where the inter-arrival time and the service time follow an exponential distribution with rates λ and μ, respectively. The expected delay of such model, denoted $\mathbb{E}[T]$, is given by [10],

$$\begin{aligned}\mathbb{E}[T] &= \mathbb{E}[W + S] \\ &= \frac{\lambda}{\mu(\mu - \lambda)} + \frac{1}{\mu},\end{aligned} \tag{15}$$

where $\mathbb{E}[W] = \lambda/\mu(\mu - \lambda)$ is the expected waiting time in the queue and $\mathbb{E}[S] = 1/\mu$ is the expected service time. From (15) and the fact that $\mathbb{E}[X] = 1/\lambda$, the theoretical $M/M/1$ average peak AoI is [3],

$$\begin{aligned}P^{M/M/1} &= \mathbb{E}[X + T] \\ &= \frac{1}{\mu}\left(1 + \frac{1}{\rho} + \frac{\rho}{1-\rho}\right),\end{aligned} \tag{16}$$

where $\rho = \lambda/\mu$ is the link utilization. Finally, the average AoI of an M/M/1 is given by [10],

$$\Delta^{M/M/1} = \frac{1}{\mu}\left(1 + \frac{1}{\rho} + \frac{\rho^2}{1-\rho}\right). \tag{17}$$

B. Experimental Setup

To validate the proposed method a simple network employing an $M/M/1$ queue was emulated locally. A Client-Server model was used as shown in Fig. 3. The updates were sent using TCP/IP from the client to the server. The client transmitted the instantaneous time-stamps of when the updates were generated, i.e., t_i. Upon sending an update, the client sleeps for a random duration that is exponentially distributed with inter-arrival rate (λ). The rate λ was varied between 1 and 8 updates per second. Fig. 4 presents the flow-chart of the client's behaviour.

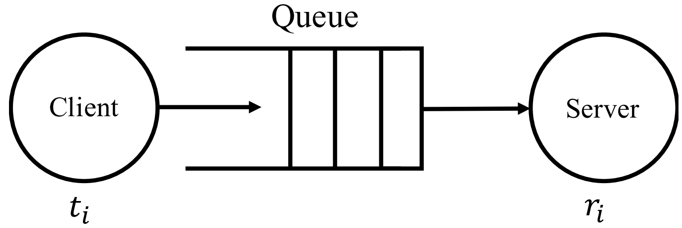

Fig. 3. Network System model showing the Client, where the time-stamp of generating updates i, i.e., (t_i). The Server saves the time of receiving the update (r_i).

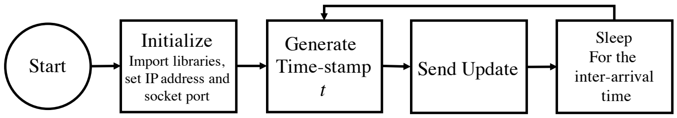

Fig. 4. Client's flow chart.

At the server side, the time of receipt r_i is recorded for each packet i along with the transmit time embedded in the update, t_i. To emulate the service time, which is exponentially distributed with rate $\mu = 10$, the server sleeping time was used. In particular, upon receipt of the update, the server will sleep for a random period (exponentially distributed), and then wake up and terminate the session with the client and record the time as the time of receipt r_i. After receiving a predefined number of updates, (we used one thousand updates per client in the experiments), the experiment terminated and the estimations presented in the previous section were performed.

The experiments were done using Apple MacOS with a 2.2 GHz Intel Core i7 processor and a 16 GB 1600 MHz DDR3 Memory. To make sure the clients and server remained tightly synchronized during the experiment, a single clock for all measurements was used. The updates were communicated using a *Python 3.6 socket* module [8]. To obtain the time-stamps, a *Python time* module [9] was used and the readings were obtained by using the object *time.time()*.

C. Results and Discussions

Fig. 5 presents the results for the delay time and compares it with the theoretical, given by (5). It can be seen that the experimental and theoretical results are in good agreement. In particular, the percentage error does not exceed 5%. The first row of Table I shows the results of a statistical test that validates the assumption that the delay (i.e., waiting time plus service time) can be accurately estimated using our method.

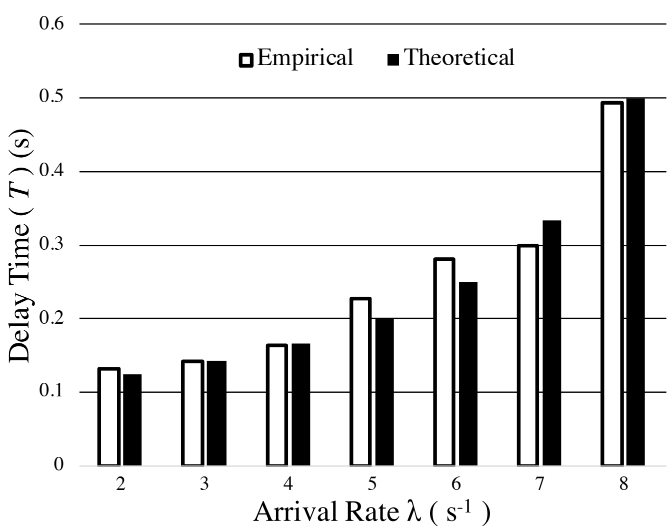

Fig. 5. Delay versus Arrival rate for M/M/1 queue calculated theoretically from (15) and measured in the experiment using the median of (10).

Next, we move to consider the PAoI estimator which was compared with its theoretical counterpart given by (16). The results in Fig. 6 show that the estimated PAoI is in very good agreement with the theoretical values. Thus, we can argue that the experimental model proposed can obtain the PAoI and delay time for the proposed system model accurately.

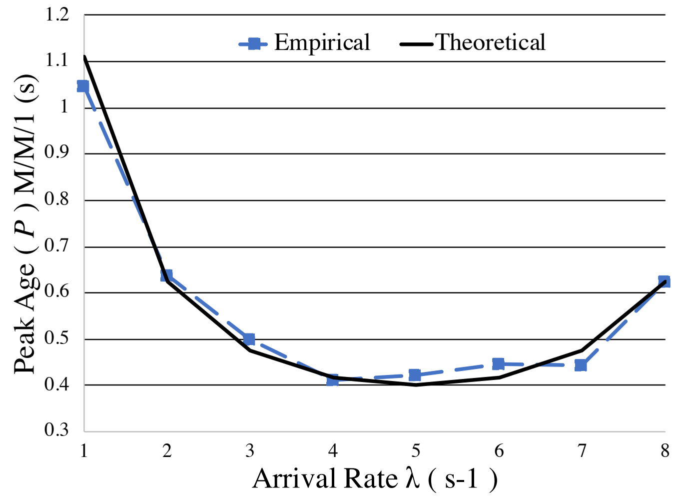

Fig. 6. Peak Age versus Arrival rate for M/M/1 queue calculated theoretically from 16 [3] and obtained experimentally using (11).

Finally, the estimated and theoretical average AoI are shown in Fig. 7 as a function of the arrival rate. The results show that the estimated and theoretical results are in agreement.

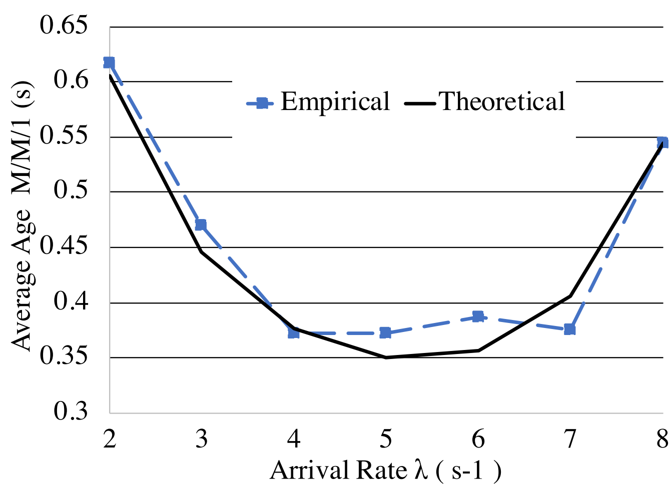

Fig. 7. Average Age versus Arrival rate for M/M/1 queue calculated theoretically from (7) [7] and obtained experimentally using (14).

To assess the significance of the results, a student *t-test* was used to calculate the p-values and the percentage errors for each of the estimators. Table I presents the *t-test* p-values and the mean percent error values for the delay time, PAoI, and AoI measurements. As shown in Table I, the difference in the means can be described as not significant. Thus, it can be concluded that the proposed method is accurate. Further, it can be concluded that the number of readings is sufficient to precisely calculate the delay time, PAoI and AoI. Hence, this method was used in our paper [12] to evaluate the PAoI of Zero-wait policy.

TABLE I
P VALUES AND PERCENT ERRORS FOR M/M/1, VALIDATING THE PROPOSED METHOD FOR EVALUATING THE PEAK AND AVERAGE AGE

Parameter	*p-Value*	Percentage error
Delay (T)	0.73	6.46
Peak (PAoI)	0.39	4.16
Age (AoI)	0.39	4.46

V. CONCLUSIONS AND FUTURE WORK

The AoI is a novel metric suggested to measure the freshness of information. A considerable amount of work has been done to evaluate and optimize AoI and PAoI theoretically. In this paper, we tried to motivate more experimental work on the AoI by making it straightforward to estimate the metrics from experiments. The proposed method was validated on an emulated M/M/1 queue. We showed that the proposed method could achieve estimates that are very close to the theoretical counterparts. The obvious next step is to test the accuracy of estimators in experiments involving different queuing models. Furthermore, it is worth investigating the implications of considering the AoI and PAoI on real-life applications such as remotely controlled robots.

REFERENCES

[1] B. Barakat and K. Arshad, “An adaptive hybrid scheduling algorithm for LTE-Advanced,” in 2015 22nd International Conference on Telecommunications (ICT), 2015, pp. 91-95.

[2] M. Costa, M. Codreanu, and A. Ephremides, “On the Age of Information in Status Update Systems With Packet Management,” IEEE Trans. Inf. Theory, vol. 62, no. 4, pp. 1897-1910, Apr. 2016.

[3] L. Huang and E. Modiano, “Optimizing age-of-information in a multi-class queueing system,” in 2015 IEEE International Symposium on Information Theory (ISIT), 2015, pp. 1681-1685.

[4] E. Najm and R. Nasser, ”Age of information: The gamma awakening,” 2016 IEEE International Symposium on Information Theory (ISIT), Barcelona, 2016, pp. 2574-2578.

[5] A. Kosta, N. Pappas, and V. Angelakis, “Age of Information: A New Concept, Metric, and Tool,” Found. Trends Netw., vol. 12, no. 3, pp. 162-259, 2017.

[6] C. Kam, S. Kompella, and A. Ephremides, “Experimental evaluation of the age of information via emulation,” in Proceedings - IEEE Military Communications Conference MILCOM, 2015, vol. 2015-Decem, pp. 1070-1075.

[7] S. Kaul, R. Yates, and M. Gruteser, “Real-time status: How often should one update?,” in Proceedings - IEEE INFOCOM, 2012, pp. 2731-2735.

[8] “18.1. socket - Low-level networking interface Python 3.3.7 documentation.” [Online]. Available: https://docs.python.org/3.3/library/socket.html. [Accessed: 06-Mar-2018].

[9] “16.3. time - Time access and conversions - Python 3.6.4 documentation.” [Online]. Available: https://docs.python.org/3/library/time.html. [Accessed: 07-Mar-2018].

[10] D. Bertsekas and R. Gallager, Data Networks. Prentice Hall, 1992.

[11] C. Snmez, S. Baghaee, A. Ergii and E. Uysal-Biyikoglu, “Age-of-Information in Practice: Status Age Measured Over TCP/IP Connections Through WiFi, Ethernet and LTE,” 2018 IEEE International Black Sea Conference on Communications and Networking (BlackSeaCom), Batumi, 2018, pp. 1-5.

[12] B. Barakat, S. Keates, I. J. Wassell, and K. Arshad, “Is the Zero-Wait Policy Always Optimum for Information Freshness (Peak Age) or Throughput?,” in IEEE Communications Letters, doi:10.1109/LCOMM.2019.2907935, (in press).