Queueing Systems Vol. 100
Special Issue: 100 views on queues

In search of the least age of information

George Kesidis · **Takis Konstantopoulos** ·
Michael A. Zazanis

February 19, 2022

1 Introduction

Consider a service system (e.g., a transmission or access queue to a network) admitting *information-carrying messages*. The need to have the most recent information has been addressed in the last few years by a quantity known as *Age of Information* (AoI). If information-carrying messages require a positive amount of time to be processed then the problem of finding a policy that keeps information as fresh as possible naturally presents itself.

To define AoI in generality, consider the times at which information-carrying messages arrive into the system. Such a message may be immediately rejected upon arrival or may be accepted. If accepted, it will stay in the system until it either gets rejected for some reason or depart having being processed in its entirety. If the latter event happens, we say that the information-carrying message is *successful*. Define, for all times t, $A^*(t) :=$ the last arrival time before t of a successful information-carrying message that departed prior to time t. The AoI at time t is defined simply as $\alpha(t) := t - A^*(t)$. If the ingredients of the system (topology, storage space, arrival times, processing times, etc.) are random, then α is a stochastic process that often admits a stationary version. One is typically interested in the random variable α that is distributed like $\alpha(t)$ for all t when $\alpha(\cdot)$ is stationary,

The problem then is to achieve a least $\mathbb{E}f(\alpha)$, where f is a certain quality measure (examples: $f(\alpha) = \alpha$; $f(\alpha) = \mathbf{1}_{\alpha > t_0}$), taking into account the freedom one has in designing the system as well as deciding when to start processing an information-

G. Kesidis
School of EECS, Pennsylvania State University, USA. E-mail: gik2@psu.edu

T. Konstantopoulos
Dept. of Mathematical Sciences, University of Liverpool, UK. E-mail: takiskonst@gmail.com

M.A. Zazanis
Dept. of Statistics, Athens University of Economics & Business, Greece. E-mail: zazanis@aueb.gr

carrying message. We refer to this problem as "minimizing AoI". See Section 3 for a simple concrete version of this.

Here is the applied spiel one usually alludes to when justifying the use of such a measure. In applications such as virtual reality, online gaming, autonomous vehicles, stock market trading, "cyber physical" systems, etc., it is the freshness of information that is important rather than the correct transmission of all messages. E.g., when trading at the speed of light, one (not including the authors) is lured by the possibility of amassing wealth in microseconds if the most recent information is available. Be it as it may, minimizing AoI is an interesting mathematical problem. Adopting AoI as a performance criterion immediately poses some simplifications over traditional queueing theory performance criteria but also presents significant challenges. We refer, e.g., to [6,7] and references therein, for the advertising of AoI and some background work.

Say that we have a single queue with one server, undetermined buffer size, where all arriving messages are information-carrying. (Buffer of size n cells means that the server processes the message sitting in cell 1, if any.) Since we are not interested in obsolete information, it is reasonable to conjecture that every time a message arrives, the system should start processing it as soon as possible by even, perhaps, interrupting the currently processed message. It also seems reasonable to serve stored messages in reverse order of arrival: the most recent message must be served first (LIFO), e.g., [1]. Maintaining an unlimited buffer, and serving messages in FIFO order, particularly with infinite storage, seems to be, insofar as AoI is concerned, the worst possible thing to do.

2 Background work

Given this intuition, a number of what appear-to-be very good systems have been analyzed. Let us say that the buffer is of size 1, just enough to store the message being processed. On one extreme, every arriving message interrupts the current one and pushes it out. We call this $\mathscr{P}_1$. On the other extreme, every arriving message is immediately discarded if one is in the system. We call this $\mathscr{B}_1$ (or, in queueing theory parlance, $G/G/1/1$). In [3], we used Palm probabilities to write down fixed point equations for the distribution of α. Specializing to the GI/GI cases, we arrived at formulas for these distributions: see [2,3] for $\mathscr{P}_1$ and $\mathscr{B}_1$.

What if we have buffer of size 2? Policy-wise, we have at least two choices: $\mathscr{B}_2$ is simply a standard blocking system; $\mathscr{P}_2$ means that arriving messages keep pushing out any message stored in cell 2, while leaving the message in service undisturbed. The distribution of AoI is computed under M/GI assumptions in [2,4]; the GI/M cases are similarly derived by Markov-renewal methods. (In the same vein, we can consider $\mathscr{P}_n$ and $\mathscr{B}_n$ systems and carry out the analysis similarly, but it soon transpires that analytical expressions become formidable.)

But there are other things one can do. For example, the $\mathscr{P}_{2,\theta}$ system, analyzed in [4], has a buffer of size 2 and, if a message arrives and finds only one message in the system that has received at most θ units of service, then the new message pushes the current one out; if θ has been exceeded then the arriving message sits in cell 2. (Note that $\mathscr{P}_{2,0} = \mathscr{P}_2$, $\mathscr{P}_{2,\infty} = \mathscr{P}_1$.)

3 A concrete problem

Consider a buffer of size n and a single server. Assume all messages are information-carrying. Given the joint distribution between the message arrival and service processes and the information (if any) kept in the system at each point of time, determine the size n and the servicing policy that will minimize the expectation of the stationary AoI. The problem makes sense even when no information is kept.

What we know and a future research direction: The heuristics above are almost correct but not exactly. If the service times are i.i.d. exponential then $\mathscr{P}_1$ is best. But, perhaps surprisingly, $\mathscr{P}_1$ is not always best. Depending on the stochastic assumptions, $\mathscr{B}_1$ may be better, both in terms of smaller $\mathbb{E}\alpha$ and smaller $\mathbb{P}(\alpha > t)\ \forall t$. Among the $\mathscr{B}_n$ systems for $n \geq 2$, $\mathscr{B}_2$ is best. (Under the obvious coupling, $\alpha_{\mathscr{B}_n}(t) \geq \alpha_{\mathscr{B}_2}(t)\ \forall t$.)

One is tempted at this point to conclude that to minimize some stationary AoI metric, generally one of the three systems $\mathscr{B}_1$, $\mathscr{P}_1$ or $\mathscr{P}_2$ should be used. But this depends on how much information is kept by the system. There are cases where $\mathscr{P}_{2,\theta}$ achieves strictly smaller mean AoI for $0 < \theta < \infty$ [4]. One may similarly blend $\mathscr{B}_1$, $\mathscr{P}_1$ and $\mathscr{P}_2$ and possibly use perturbation analysis (e.g., [5]) to dynamically tune the associated parameters toward minimizing some AoI-based performance metric.

N.B. There may be a problem with the very definition of $\alpha(t) = t - A^*(t)$. As pointed out in [3], it is often better to define $\beta(t) := A(t) - A^*(t)$, where $A(t)$ is the last arrival time before t of an information-carrying message, as a new measure (naturally called *New Age of Information* (NAoI)). Indeed, NAoI is defined with respect to the system. Typically, the system has no control of the arrivals and if, say, arrivals occur with high fluctuations (say very large variance) then $\mathbb{E}\alpha$ will be very large (potentially infinite), but $\mathbb{E}\beta$ may remain low. This is why the problems mentioned above have been addressed with respect to the NAoI also; see [3].

References

1. A.M. Bedewy, Y. Sun and N.B. Shroff. Minimizing the Age of Information Through Queues. *IEEE Trans. Info. Th.* **65**(8):5125-5232, Aug. 2019.
2. Y. Inoue, H. Masuyama, T. Takine and T. Tanaka. A general formula for the stationary distribution of the age of information and its application to single-server queues. *IEEE Trans. Inf. Th.* **65**(12):8305-8324, Dec. 2019.
3. G. Kesidis, T. Konstantopoulos, and M. Zazanis. The distribution of age-of-information performance measures for message processing systems. *Queueing Systems* **95**(3):203–250, Aug. 2020.
4. G. Kesidis, T. Konstantopoulos and M. Zazanis. Age of information distribution under dynamic service preemption. http://arxiv.org/abs/2104.11393, April 2021; see also http://arxiv.org/abs/2104.08050 and http://arxiv.org/abs/2106.08473
5. P. Konstantopoulos and M. Zazanis. Sensitivity analysis for stationary and ergodic queues. *Adv Appl Prob*, 24(3):738-750, Sept. 1992.
6. Y. Sun, I. Kadota, R. Talak and E. Modiano. *Age of Information: A New Metric for Information Freshness*. Morgan & Claypool Publishers, 2020.
7. R. Yates, Y. Sun, D. Brown, S. Kaul, E. Modiano and S. Ulukus. Age of information: an introduction and survey. *IEEE J. Sel. Areas Comm.* **39**(5):1183-1210, May 2021.