Age of Information Under Random Updates

Clement Kam, Sastry Kompella
Information Technology Division
Naval Research Laboratory
Washington, DC

Anthony Ephremides
Electrical and Computer Engineering Department
University of Maryland
College Park, MD

Abstract—We consider the system where a source randomly generates status update messages and transmits them via a network cloud to the intended destination. These update messages can take different times to traverse the network, which we model as exponential service times, and may result in packets reaching the destination out of order, rendering some of the earlier transmissions obsolete. We analyze the status update age for such a system, and show that it tracks well with simulation results.

I. INTRODUCTION

The need for real-time status updates in an increasingly ubiquitous connectivity environment is recognized in [1], in which a new metric called the status update age has been defined. Status updates are periodic messages that are generated by a source, such as a sensor, in order to convey the current state of a measurement or observation. In commercial applications, these can include but are not limited to environmental sensing data [2], vehicular sensor measurements [3], etc. While real-time status updating has become increasingly popular in commercial applications fueled by the explosion of portable devices, such status update messages have been commonplace in military applications in the form of blue force tracking [4], telemetry data etc., and more recently cooperative spectrum sensing applications [5]. In such systems the status update age of a message can be loosely defined as the duration between the time of observation at the destination (of the current status) and the time at which that particular packet was originally generated at the source. These packets may contain individual measurements or observations, or sometimes a collection of measurements and/or observations can be aggregated into a single packet, especially if they are required to reach a common destination.

In this paper, we consider a system in which a source generates time-stamped status update messages that are transmitted through a cloud-based communication system to a monitor as shown in Figure 1. However, due to various reasons, the source is limited in its ability to transmit at will, and can only generate the current information in a random fashion. For example, this could be due to limited battery resources in a sensor node that relies on energy harvesting for continued operation, or because of an application-specific feature that requires a random sampling of data. Once generated, the source can immediately transmit this update message without needing to buffer it in a queue. Ideally, we would like the destination to receive the status updates in the order that they were generated. However, that may not be possible if the source and destination are separated by a network cloud, and the update packets are routed through the network via multiple paths to the destination, resulting in varying packet delivery times.

An example of the evolution of the status update age in such a system is shown in Figure 2, where the time average of the status updates is the normalized area under the sawtooth waveform. In most cases, the update packets are received by the destination in the same order as they were transmitted from the source. However, as explained earlier, because of the randomness in the packet delivery times, there is a chance that a packet, e.g., packet 3 which is generated at time \( t_3 \), may complete service after a packet generated in the future, e.g., packet 4 which reaches the destination at \( t_4 < t_3 \). In such a case, the status age is updated when packet 4 is received at the destination, but is unaffected when packet 3 is finally received at \( t_3 \). Calculating the average status age in the presence of such out-of-order packet receptions creates new challenges when compared to [1], [6], and even then provides only a piece of the puzzle.

The computation of the average status update age opens up a whole new set of interesting questions, most importantly, how does one maintain the freshness of information at the receiver, and what are the levers of control we can use to affect it? While intuition suggests that this can be achieved by making the source generate updates more rapidly, this could lead to increased congestion in the network and increase the chances of out-of-order packet reception at the destination. Note that every time a packet overtake a new set of questions, most importantly, how does one maintain the freshness of information at the receiver, and what are the levers of control we can use to affect it? While intuition suggests that this can be achieved by making the source generate updates more rapidly, this could lead to increased congestion in the network and increase the chances of out-of-order packet reception at the destination. Note that every time a packet overtake one or more previously generated packets, it results in wasted resources. However, minimizing out-of-order receptions alone could lead to outdated status information at the destination. Therefore,
this problem offers a rich set of possible options towards simultaneously reducing wastage of resources and minimizing the average age of information, which will be our focus for future work.

The remainder of the paper is organized as follows. In Section II, we formalize the definition of our system model, then in Section III we present some useful building blocks related to the system, which we use to derive the status age in Section IV and present some numerical results. Finally, we make some concluding remarks in Section V.

II. SYSTEM MODEL

We now define a system in which a source transmits packets through a network cloud to a remote destination. At transmission time, the source transmits a packet containing the current information, so there is no aging of information from waiting in a packet queue. As denoted in Figure 2, transmissions occur at times \( t_0, t_1, \ldots \), and receptions at the destination occur at times \( \tau_0, \tau_1, \ldots \).

We call the time between transmissions the interarrival time \( X_a, a = 1, 2, \ldots \), which is equal to \( t_a - t_{a-1} \). The interarrival times are modeled as random, so the source does not have control over the exact times at which it can transmit updates. In our setup, the \( X_a \)'s are i.i.d. exponential random variables with rate \( \lambda \).

We call the time spent in the network the service time \( S_a, a = 1, 2, \ldots \), which is equal to \( \tau_a - t_a \). For each interarrival time \( X_a \), the service time \( S_a \) that immediately follows is modeled as exponential with rate \( \mu \), and all the \( S_a \)'s are i.i.d. and independent of the \( X_a \)'s. We consider this to be a simplified model of the random delay due to routing through the network cloud, which is a result of various phenomena such as changing link states, competing data traffic, and other network dynamics.

We assume that packets enter the cloud instantaneously at each transmission time. Due to the randomness of the service times, packets are not necessarily received at the destination in the order in which they are transmitted, as noted in the introduction. Some packets that arrive out of order are useless to the receiver since they do not provide newer information. As a result, the receiver does not update the status for every packet received, which complicates the calculation of the status age.

Definitions. We define the status age \( \Delta(t) \) as in [1], where the age at time \( t \) is \( \Delta(t) = t - u(t) \), where \( u(t) \) is the timestamp of the most recent information at the receiver as of time \( t \). In our system, the timestamp coincides with the transmission time of the packet. Given this definition, we can see that the status age increases linearly with \( t \) but is reset to a smaller value with each packet received that contains newer information, resulting in the sawtooth pattern shown in Figure 2.

We define an informative packet as a packet received at the destination that provides newer information than what has been received up to that time. For example, in Figure 2, we say that packet 2 is an informative packet because it is received before packets 3, 4, and all future packets \( (\tau_2 < \min(\tau_3, \tau_4, \ldots)) \). However, as noted in the introduction, packet 3 is not considered informative since it is received after packet 4 \( (\tau_3 > \tau_4) \). In terms of \( X_a \)'s and \( S_a \)'s, the condition for a packet \( m \) being an informative packet is \( S_m < \min_r \left( S_r + \sum_{a=m+1}^r X_a \right) \), where \( r = m + 1, m + 2, \ldots \).

We say that a packet \( p \) is rendered obsolete if some packet \( q \) transmitted after \( p \) (i.e., \( t_q > t_p \)) is received at the destination first (i.e., \( \tau_q < \tau_p \)), e.g., packet 4 renders packet 3 obsolete. An informative packet is one that is not rendered obsolete.

III. BUILDING BLOCKS

Before we compute the status age for this system, we first derive some useful building blocks, such as the probability that a packet is an informative packet, the probability that a packet renders the previous \( n \) packets obsolete, and some statistics about the service times and interarrival times conditioned on these events.

A. Conditional Probability of a Packet Being Informative

First we compute the probability that a packet \( m \) is an informative packet, given that its service time is \( s_m \) and the interarrival times of future packets are \( x_{m+1}, x_{m+2}, \ldots \). Let \( E_1(m) \) be the event that a packet \( m \) is an informative packet, and let \( X_{\infty}^{a+b} = [X_a, X_{a+1}, \ldots, X_{a+b}] \). Then \( E_1(m) \) conditioned on its service time \( S_m \) and the future interarrival times \( X_{m+1}^\infty \) is given by

\[
\Pr\{E_1(m)|S_m = s_m, X_{m+1}^\infty = X_{m+1}^\infty\} = \Pr\left\{ \sum_{r=1}^{\infty} \left\{ s_m - \sum_{k=1}^{r} x_{m+k} \right\} \right\}
\]

\[
 = \prod_{r=1}^{\infty} \left( e^{-\mu(s_m - \sum_{k=1}^{r} x_{m+k})} \right) \left\{ s_m > \sum_{k=1}^{r} x_{m+k} \right\}
\]

\[
+ \sum_{r=1}^{\infty} \left\{ s_m < \sum_{k=1}^{r} x_{m+k} \right\}
\]
\[
\begin{align*}
\mathbb{I}\{s_m < x_{m+1}\} + \sum_{r=1}^{\infty} \left( e^{-\mu(x_{m} - \sum_{k=1}^{r+1}(r-k+1)x_{m+k}} \right) \\
\cdot \mathbb{I}\left\{ \sum_{k=1}^{r} x_{m+k} < s_m < \sum_{k=1}^{r+1} x_{m+k} \right\}
\end{align*}
\]

where \( x^{m}_{n} = [x_n, x_{n+1}, \ldots, x_{n+h}] \) and \( \mathbb{I}\{A\} = 1 \) when the Boolean expression \( A \) is true, otherwise \( \mathbb{I}\{A\} = 0 \).

**B. Conditional Probability of a Packet Rendering the Previous n Packets Obsolete**

Now assuming that there have been at least \( n \) packets transmitted in the past, we let \( E_2(n) \) be the event that the current packet renders exactly \( n \) of the previous packets obsolete (i.e., the packet transmitted before the previous \( n \) packets is an informative packet). For example, in Figure 2, packet 4 renders exactly 1 packet obsolete, meaning packet 2 is an informative packet and packet 3 is rendered obsolete. We first find the probability of this event conditioned on \( S_m \) and \( X^{m}_{m-n} \):

\[
\Pr\{E_2(n)|S_m = s_m, X^{m}_{m-n} = x^{m}_{m-n}\} = \Pr\left\{ (S_{m-n-1} < s_m + \sum_{k=0}^{n} x_{m-k}) \right\}
\]

\[\cap \left\{ \bigcap_{s=0}^{n-1} \left[ s_{m-s-1} > s_m + \sum_{k=0}^{n} x_{m-k} \right] \right\} \]

\[= (1 - e^{-\mu(s_m + \sum_{k=0}^{n} x_{m-k})}) \prod_{k=0}^{n-1} e^{-\mu(s_m + \sum_{k=0}^{n} x_{m-k})} \]

\[= e^{-\mu s_m} e^{-\mu \sum_{k=1}^{n} k x_{m+n+k}} - e^{-\mu(n+1)s_m} e^{-\mu \sum_{k=1}^{n} k x_{m+n+k-1}}.
\]

Averaging over the \( x^{m}_{m-n} \), we then get the probability conditioned on \( S_m \):

\[
\Pr\{E_2(n)|S_m = s_m\} = \frac{\lambda^n}{\prod_{k=1}^{n} (\lambda + k\mu)} \cdot \left( e^{-\mu s_m} - \frac{\lambda}{\lambda + (n+1)\mu} e^{-\mu(n+1)s_m} \right).
\]

**C. Probability of an Informative Packet Rendering Previous n Packets Obsolete**

Having computed conditional probabilities of \( E_1(m) \) and \( E_2(n) \), we will use them to compute the intersection of the two events. We note that \( \{E_1(m)|S_m = s_m, X^{\infty}_{m+1} = x^{\infty}_{m+1}\} \) is independent of the \( X^{m}_{m-n} \), so averaging the probability of \( E_2(n) \) over \( X^{m}_{m-n} \) as in (3) is valid prior to computing the probability of their intersection. We also note that \( \Pr\{E_2(n)|S_m = s_m\} \) consists of two terms with \( e^{-\mu s_m} \) and \( e^{-\mu(n+1)s_m} \). For the first term, we average \( \Pr\{E_1(m)|S_m = s_m, X^{\infty}_{m+1} = x^{\infty}_{m+1}\} \cdot e^{-\mu s_m} \) over \( s_m \):

\[
\int_0^{\infty} \Pr\{E_1(m)|S_m = s_m, X^{\infty}_{m+1} = x^{\infty}_{m+1}\} e^{-\mu s_m} f_{S_m}(s_m) \, ds_m = \frac{1}{n+1} \left( 1 - e^{-\mu(n+1)s_m} \right) + \sum_{r=1}^{\infty} \left( \frac{1}{n+r+1} \right) e^{-\mu(n+r+1)s_m} \]
E. Conditional Expectations of $\sum_{k=m-n}^{m} X_k$, $\sum_{k=m-n}^{m} X_k^2$

In this section, we solve for a variety of conditional expectations related to $X_m$. To do so, we must first derive the conditional pdf:

$$f_{\mathcal{E}(n)}(\{X_m^n | \mathcal{E}(n)\}) = \frac{Pr\{\mathcal{E}(n)|X_m^n = x_m^n\}f_{X}(x_m^n)}{Pr\{\mathcal{E}(n)\}}$$

$$= \int_{0}^{\infty} Pr\{\mathcal{E}(n)|S_m = s_m, X_m^n = x_m^n\} f_{S_m}(s_m) dx_m^n$$

where

$$\sigma(n) = \sum_{r=1}^{\infty} \left[ \frac{\lambda^r}{(n+r+1)\prod_{k=1}^{r}(\lambda + (n+k)\mu)} \right] \cdot \left(1 - \frac{\lambda}{\lambda + (n+r+1)\mu}\right)$$

To find the conditional sum of means, we compute

$$E\left[ \sum_{k=m-n}^{m} X_k | \mathcal{E}(n) \right] = \int_{m-n}^{m} X_k f_{X}(x_k | \mathcal{E}(n)) dx_k$$

$$= \sum_{k=0}^{n} \frac{1}{\lambda + k\mu} + \frac{n+1}{\lambda} \cdot \sigma(n).$$

The conditional sum of second moments can be similarly derived:

$$E\left[ \sum_{k=m-n}^{m} X_k^2 | \mathcal{E}(n) \right] = \sum_{k=0}^{n} 2\left(\frac{n}{\lambda + k\mu}\right)^2 + \frac{2(n+1)}{\lambda^2} \left(1 + \frac{\lambda}{\lambda + (n+1)\mu}\right) \cdot \sigma(n).$$

Lastly, the conditional sum of crossterms can also be derived to obtain

$$E\left[ \sum_{j=m-n}^{m} \sum_{k=j+1}^{m} 2X_j X_k | \mathcal{E}(n) \right] = \sum_{j=0}^{n} \sum_{k=j+1}^{n} \frac{2}{\lambda + j\mu + k\mu} + \frac{(n+1)\sigma(n)}{\lambda} \sum_{k=1}^{n} \frac{2}{\lambda + k\mu}.$$
Note that $\lambda$ is not equal to the arrival rate of informative packets $\lim_{T \to \infty} D(T)/T$, but when $\lambda$ is combined with $Pr\{E(n)\}$, only the informative packets are accounted for.

Finally, after substitution of the terms from the building blocks in previous section, we can express the average status age as

$$\Delta = \lambda \sum_{n=0}^{\infty} Pr\{E(n)\} \left[ \sum_{j=0}^{n} \frac{1}{\lambda + j \mu} \right. \cdot \left( \frac{(n+2)\mu}{\lambda + (n+1)\mu} \lambda + (n+2)\mu \right)
+ \frac{(n+1)\sigma(n)}{\lambda}
+ \frac{(n+2)\mu}{\lambda + (n+1)\mu} \lambda + (n+2)\mu \sum_{k=j}^{n} \frac{1}{\lambda + k \mu} \right]. \quad (6)$$

B. Upper and Lower Bounds

We will compare our status age expression with some simple upper and lower bounds. An upper bound for the status age is $\lambda$ multiplied by the average of the trapezoidal areas for each packet, $\frac{1}{2}[(X_n + S_n) - S_n^2]$, which contains the area under the curve plus some extraneous segments. This bound is given by

$$\Delta_{UB} = \lambda \left( \frac{E[X^*]}{2} + E[S]E[X] \right) = \frac{1}{\mu} + \frac{2}{\mu}. \quad (6)$$

For the lower bound, we consider altering the service time model such that the new $S_n$ can be no greater than $X_{n+1}$, the interarrival time of the next packet. This results in trapezoidal areas under the status age curve that are no greater than those for our actual system. By conditioning on the probability that the original $S_n$ is greater than or less than $X_{n+1}$, we can compute the average trapezoidal area, eventually arriving at the lower bound $\Delta_{LB} = \frac{1}{\mu} + \frac{1}{\lambda + \mu} - \frac{\lambda \mu}{\lambda + \mu} \cdot \frac{1}{\mu}$.\quad (6)

C. Numerical Results

We have numerically evaluated our expression for status age (6) for $\mu = 0.5, 1, 1.5$ and plotted the results vs. the system utilization $\rho$ in Figure 3. We have also simulated the system and computed the (simulated) age over $10^5$ time units and averaged over multiple trials, and the result is very close to the numerically-evaluated (theoretical) age. The upper and lower bounds are also included in the figure using dotted/dashed lines for each $\mu$.

We can see that the status age is minimized when the system utilization is increased, since more frequent transmissions leads to more frequent updates in this system model. However, increasing utilization comes at the cost of needless consumption of network resources. Since packets that are obsolete do not contribute to reducing the age, they end up being a waste of resources. We have also plotted in the same figure the percentage of packets that are rendered obsolete (green dotted line, right-hand y-axis), and verified the results via simulation. We see that for high utilization, over 25% of packets are wasting resources. In our future work, we will further investigate this tradeoff between minimizing the status age and needless consumption of resources.

V. Conclusion

The age of information for a status updating system through a network cloud is considered in this work. Formulating this system model, with random transmission and service processes, no waiting time, and the possibility of packets arriving out of order, opens up a new set of problems for minimizing the status age. In this work, we have computed the status age for exponential interarrival and service times with constant rates, but the computation is complicated by the out-of-order reception, with some packets having no effect on the status age. We have derived an expression for the status age and verified it by simulation, and we observe that the age decreases with increasing system utilization and service rate, as expected. However, increasing the utilization comes at the cost of increasing wasted resources spent on packets that are rendered obsolete. We plan to extend this work by considering approaches to reduce wastage of resources, as well as considering different models for the service time statistics to account for congestion.

ACKNOWLEDGMENT

This work was supported by the Office of Naval Research.

REFERENCES


