Age of Information Analysis of Multi-user Mobile Edge Computing Systems

Zhifeng Tang*, Zhuo Sun†, Nan Yang*, and Xiangyun Zhou*
*School of Engineering, Australian National University, Canberra, ACT 2600, Australia.
†School of Computer Science, Northwestern Polytechnical University, Xi'an, Shaanxi 710072, China.
Email: zhifeng.tang@anu.edu.au, zsun@nwpu.edu.cn, nan.yang@anu.edu.au, xiangyun.zhou@anu.edu.au

***Abstract*—In this paper, we analyze the age of information (AoI) performance of a multi-user mobile edge computing (MEC) system where a base station (BS) generates and transmits computation-intensive packets to user equipments (UEs). In this MEC system, we consider two computing schemes, namely, the local computing scheme and the edge computing scheme. In the local computing scheme, each packet is transmitted to the UE and then computed by the local server at the UE. In the edge computing scheme, each packet is computed by the edge server at the BS and then transmitted to the UE. Considering exponentially distributed transmission time and computation time and adopting the first come first serve queuing policy, we derive the closed-form expressions for the average AoI of these two computing schemes. Simulation results corroborate our analysis and examine the impact of system parameters on the average AoI.**

***Index Terms*—Age of information, mobile edge computing, first come first serve.**

I. Introduction

In recent years, real-time applications, such as intelligent transport systems and factory automation, have attracted a wide range of interests. In these applications, timely status updates are significantly critical for accurate monitoring and control [1, 2]. In order to fully characterize the freshness of delivered status information, the concept of age of information (AoI) was introduced as a new performance metric [3]. Specifically, AoI is defined as the elapsed time since the last successfully received status was generated by the transmitter, which is a time metric that captures both the latency and the freshness of a transmitted status. In real-time applications, data processing usually consumes a significant amount of time, due to the limited computing capacity of the local server. This seriously degrades the AoI performance. To tackle this problem, mobile edge computing (MEC) was introduced by the European Telecommunications Standard Institute (ETSI) [4]. In MEC systems, the server deployed at the network edge, called the edge server, is exploited to offload data processing from the local server [5, 6]. Owing to the powerful computing capacity of the edge server, the MEC system can significantly reduce the computation time [7–9].

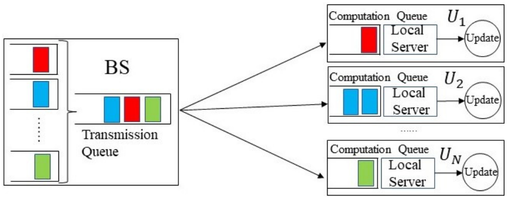
The AoI has been widely evaluated as an effective performance metric of MEC system, starting from the point-to-point system. In [10], a status sampling policy was designed to minimize the average AoI of the MEC system. By considering power allocation between transmission and computation, [11] proposed a scheduling policy to minimize the average AoI of the MEC system. Based on the AoI, [12] introduced the age of task (AoT) and designed an offloading policy to improve the AoT performance of the MEC system. In addition, [13] investigated the effect of different computing schemes on the AoI performance and found that partially offloading tasks to the edge server has the better AoI performance than other schemes in the MEC system. Building upon these efforts on the MEC system with a single user, increasing research efforts have been devoted to investigating the AoI performance of multi-user MEC systems. [14] designed the edge resource allocation to minimize the average AoI of a multi-user MEC system. [15] proposed a deep reinforcement learning based scheduling policy to minimize the average AoI of an unmanned aerial vehicles (UAV) assisted multi-user MEC system. Although the aforementioned studies have designed the scheduling policy to minimize the average AoI and analyzed the AoI performance of different MEC systems, the impact of different computing schemes on the AoI performance of a multi-user MEC system has not been investigated.

In this paper, we study the AoI performance of a multi-user MEC system. In this system, the base station (BS) transmits computation-intensive packets to multiple user equipments (UEs). The packet can be computed by either the local server at each UE, referred to as the local computing scheme, or the edge server at the BS, referred to as the edge computing scheme. We derive the closed-form expressions for the average AoI in these two computing schemes. Using simulations, we demonstrate the accuracy of our analysis results. We then characterize the impacts of various design parameters on the average AoI in two computing schemes.

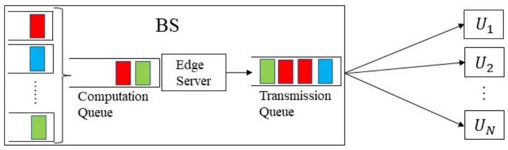
II. System Model and Average AoI

In this paper, we consider a downlink system as depicted in Fig. 1, where the BS transmits time-sensitive packets to N UEs. We denote the nth UE by U_n, where $n = 1, 2, \cdots, N$. In this system, the BS generates the packet of U_n according to a Poisson process[1] with rate λ_n. To ensure the freshness of packets, we consider an MEC system, where each packet can be computed by either the local server at UE or the edge server at the BS. Depending on which server computes the packets, we introduce two computing schemes, i.e., the

[1]We assume that the packet generation processes among UEs are independent but not identical.



(a) Local computing scheme



(b) Edge computing scheme

Fig. 1. Illustration of our considered MEC system where the BS transmits time-sensitive packets to N UEs.

local computing scheme and the edge computing scheme. In the local computing scheme, the BS directly transmits the generated packet to the UE for the local computing as shown in Fig. 1(a). In the edge computing scheme, the edge server at the BS computes the packet and then transmits the computational result of the packet to the UE as depicted in Fig. 1(b). In the MEC system, we assume that the packets in the queues are served with the first come first serve (FCFS) queuing policy. We then assume that both the computation time and the transmission time of a packet follow exponential distributions [16], where μ_n is the computation rate of the local server at U_n, μ_B is the computation rate of the edge server at the BS, and μ_D is the the transmission rate of the BS.

Our target is to analyze the average AoI of the considered MEC system. To meet this target, without loss of generality, we arbitrarily select one UE, U_n, and analyze its average AoI, Δ_n, which gives the average AoI of the system. Fig. 2 plots a sample variation of AoI for U_n, $\Delta_n(t)$, as a function of t. We assume that the observation begins at $t = 0$, where the AoI is $\Delta_n(0)$. From Fig. 2, we express the AoI of U_n at time t as

$$\Delta_n(t) = t - u_n(t), \tag{1}$$

where $u_n(t)$ is the generation time of the last received computed packet of U_n at time t. Then the time-average AoI of U_n over the observation time interval $(0, \tau)$ can be calculated as

$$\Delta_n = \frac{1}{\tau}\int_0^\tau \Delta_n(t)\mathrm{d}t. \tag{2}$$

We denote $P_{n,j}$ as the jth packet generated after time $t = 0$ of U_n, $j = 1, 2, \cdots$. We then denote Y_j as the time interval between the generation time of $P_{n,j-1}$ and the generation time of $P_{n,j}$ and denote T_j as the time interval between the generation time of $P_{n,j}$ and the time that U_n obtains the computational result of $P_{n,j}$. Therefore, we have

$$Y_j = t_j - t_{j-1} \tag{3}$$

and

$$T_j = t'_j - t_j, \tag{4}$$

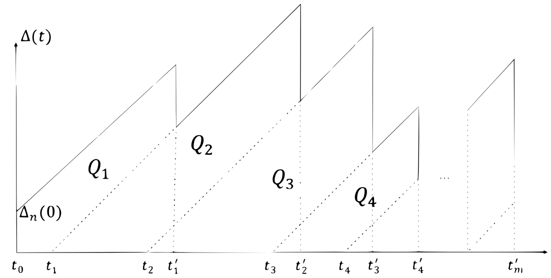


Fig. 2. The AoI variation of the selected UE, U_n.

where t_j is the generation time of $P_{n,j}$ and t'_j is the time that U_n obtains the computational result of $P_{n,j}$. We note that $Y_1 = t_1$ is obtained by setting $t_0 = 0$. Considering that U_n obtains the computational result of the mth packet at the end of this observation time interval, i.e., $\tau = t'_m$, we calculate the average AoI as

$$\Delta_n = \frac{\sum_{j=1}^{m} Q_j + \frac{T_m^2}{2}}{\tau} = \frac{2Q_1 + T_m^2}{2\tau} + \frac{m-1}{\tau} \times \left(\frac{1}{m-1}\sum_{j=2}^{m} Q_j \right), \tag{5}$$

where Q_j is the area shown in Fig. 2. From Fig. 2, we see that Q_1 is a polygon and Q_j is an isosceles trapezoid for $j \geq 2$, which can be derived from two isosceles triangles, i.e.,

$$Q_j = \frac{1}{2}(Y_j + T_j)^2 - \frac{1}{2}T_j^2 = \frac{Y_j^2}{2} + Y_j T_j. \tag{6}$$

We note from (5) that, when $\tau \to \infty$, the impact of Q_1 and T_m^2 on the average AoI is negligible, i.e., $\lim_{\tau\to\infty} \frac{2Q_1+T_m^2}{2\tau} = 0$. Moreover, due to the fact that $\tau = Y_1 + \sum_{j=2}^{m} Y_j + T_m$, we can obtain $\lim_{\tau\to\infty} \frac{\tau}{m-1} = \mathbb{E}[Y_j]$, where $\mathbb{E}[\cdot]$ is the expectation. Therefore, by integrating (6) into (5) and taking τ to infinity, we obtain the average AoI of U_n as

$$\Delta_n = \frac{\lim\limits_{m\to\infty} \frac{1}{m-1}\sum_{j=2}^{m} Q_j}{\mathbb{E}[Y_j]} = \frac{\mathbb{E}[Q_j]}{\mathbb{E}[Y_j]} = \frac{\mathbb{E}[Y_j^2] + 2\mathbb{E}[Y_j T_j]}{2\mathbb{E}[Y_j]}. \tag{7}$$

As the BS generates the packet of U_n according to a Poisson process with rate λ_n, we obtain $\mathbb{E}[Y_j^2] = 2/\lambda_n^2$ and $\mathbb{E}[Y_j] = 1/\lambda_n$ and thus

$$\Delta_n = \frac{1}{\lambda_n} + \lambda_n \mathbb{E}[Y_j T_j]. \tag{8}$$

Finally, by averaging Δ_n over all UEs, we obtain the average AoI of the MEC system as

$$\Delta = \frac{1}{N}\sum_{n=1}^{N} \Delta_n. \tag{9}$$

It is worthwhile to note that, when the arrival rate of the packet is higher than the serving rate in any queue, the average AoI of U_n goes to infinity, since T_j goes to infinity in (8). Thus, in order to ensure system stability, we assume that the arrival rate of the packet is lower than the serving rate in each queue.

III. Closed-form Expression for Average AoI

In this section, we derive the closed-form expressions for the average AoI in the local computing scheme and the edge computing scheme, respectively.

A. Local Computing Scheme

We first derive the closed-form expression for the average AoI in the local computing scheme, given in Theorem 1.

Theorem 1: In the local computing scheme, the closed-form expression for the average AoI of U_n is given by

$$\begin{aligned}\Delta_{n,l}=&\frac{1}{\lambda_n}+\frac{1}{\mu_D}+\frac{1}{\mu_n}+\frac{\lambda_{-n}}{\mu_D(\mu_D-\lambda_{-n})}+\frac{\lambda_n^2\lambda_{-n}}{\mu_D(\mu_D-\lambda_{-n})^3}\\&+\frac{\lambda_n^2}{(\mu_D-\lambda)(\mu_D-\lambda_{-n})^2}+\frac{\lambda_n^2(\mu_D+\mu_n-\lambda_n)}{\mu_D(\mu_n-\lambda_n)(\mu_D+\mu_n-\lambda)^2}\\&+\frac{\lambda_n^2(\mu_D-\lambda)(\mu_D+\mu_n-\lambda_{-n})}{\mu_n^2(\mu_D-\lambda_{-n})(\mu_n-\lambda_n)(\mu_D+\mu_n-\lambda)},\end{aligned}\quad(10)$$

where $\lambda=\sum_{n=1}^{N}\lambda_n$ and $\lambda_{-n}=\lambda-\lambda_n$.

Proof: In this scheme, we denote $X_{j,D}$ and $X_{j,U}$ as the queuing delay of $P_{n,j}$ in the transmission queue and the queuing delay of $P_{n,j}$ in the computation queue of the local server at U_n, respectively. They are given by

$$X_{j,D}=t_{j,D}-t_j \quad (11)$$

and

$$X_{j,U}=t'_j-t_{j,D}, \quad (12)$$

respectively, where $t_{j,D}$ is the time that U_n receives $P_{n,j}$. Note that the queuing delay of $P_{n,j}$ in each queue is the summation of the waiting time and the serving time of $P_{n,j}$ in the queue. Thus, we rewrite $X_{j,D}$ and $X_{j,U}$ as

$$X_{j,D}=W_{j,D}+S_{j,D} \quad (13)$$

and

$$X_{j,U}=W_{j,U}+S_{j,U}, \quad (14)$$

respectively, where $W_{j,D}$ and $S_{j,D}$ are the waiting time and the transmission time of $P_{n,j}$ in the transmission queue, respectively, and $W_{j,U}$ and $S_{j,U}$ are the waiting time and the computation time of $P_{n,j}$ in the computation queue of U_n's local server, respectively. Based on the fact that $S_{j,D}$ and $S_{j,U}$ are independent of Y_j, we calculate the average AoI of U_n in the local computing scheme, $\Delta_{n,l}$, as

$$\begin{aligned}\Delta_{n,l}&=\frac{1}{\lambda_n}+\lambda_n\mathbb{E}[Y_j(S_{j,D}+W_{j,D}+S_{j,U}+W_{j,U})]\\&=\frac{1}{\lambda_n}+\frac{1}{\mu_D}+\frac{1}{\mu_n}+\lambda_n(\mathbb{E}[Y_jW_{j,D}]+\mathbb{E}[Y_jW_{j,U}])).\end{aligned}\quad(15)$$

To obtain $\Delta_{n,l}$, we need to derive $\mathbb{E}[Y_jW_{j,D}]$ and $\mathbb{E}[Y_jW_{j,U}]$ in (15). We first derive $\mathbb{E}[Y_jW_{j,D}]$ as

$$\begin{aligned}\mathbb{E}[Y_jW_{j,D}]=&\mathbb{E}[Y_jW_{j,D}|B_{j,D}]\Pr(B_{j,D})\\&+\mathbb{E}[Y_jW_{j,D}|L_{j,D}]\Pr(L_{j,D}),\end{aligned}\quad(16)$$

where $B_{j,D}$ denotes the event that $P_{n,j}$ is generated before $P_{n,j-1}$ arrives at the transmission queue and $L_{j,D}$ denotes the event that $P_{n,j}$ is generated after $P_{n,j-1}$ arrives at the transmission queue. Here, $B_{j,D}$ and $L_{j,D}$ are two complementary events such that $\Pr(B_{j,D})+\Pr(L_{j,D})=1$. When $B_{j,D}$ happens, $W_{j,D}$ is calculated as

$$W_{j,D}=X_{j-1,D}-Y_j+\sum_{\kappa=1}^{K_{j,y}}S_{j,\kappa,D}, \quad (17)$$

where $K_{j,y}$ is the number of packets generated for other UEs during Y_j and $S_{j,\kappa,D}$ is the transmission time of the κth packet among $K_{j,y}$ packets. Based on (17), we obtain

$$\mathbb{E}[W_{j,D}|Y_j=y,B_{j,D}]=\frac{1}{\mu_D-\lambda}+\frac{\lambda_{-n}y}{\mu_D}. \quad (18)$$

According to (18), we calculate the first item in (16) as

$$\begin{aligned}&\mathbb{E}[Y_jW_{j,D}|B_{j,D}]\Pr(B_{j,D})\\&=\int_0^\infty yf_{Y_j}(y)\Pr(B_{j,D}|Y_j=y)\mathbb{E}[W_{j,D}|Y_j=y,B_{j,D}]\mathrm{d}y\\&=\frac{\lambda_n}{(\mu_D-\lambda_{-n})^2}\left(\frac{2\lambda_{-n}}{\mu_D(\mu_D-\lambda_{-n})}+\frac{1}{\mu_D-\lambda}\right).\end{aligned}\quad(19)$$

We then calculate the second item in (16). When $L_{j,D}$ happens, $W_{j,D}$ is calculated as

$$W_{j,D}=\sum_{\kappa=1}^{K_{j,e}}S_{j,\kappa,D}, \quad (20)$$

where $K_{j,e}$ is the number of packets in the transmission queue when $P_{n,j}$ is generated. Based on (20), we obtain

$$\mathbb{E}[W_{j,D}|Y_j=y,L_{j,D}]=\frac{\lambda_{-n}}{\mu_D(\mu_D-\lambda_{-n})}. \quad (21)$$

According to (21), we calculate the second item in (16) as

$$\begin{aligned}&\mathbb{E}[Y_jW_{j,D}|L_{j,D}]\Pr(L_{j,D})\\&=\int_0^\infty yf_{Y_j}(y)\Pr(L_{j,D}|Y_j=y)\mathbb{E}[W_{j,D}|Y_j=y,L_{j,D}]\mathrm{d}y\\&=\frac{\lambda_{-n}}{\mu_D(\mu_D-\lambda_{-n})}\left(\frac{1}{\lambda_n}-\frac{\lambda_n}{(\mu_D-\lambda_{-n})^2}\right).\end{aligned}\quad(22)$$

Combining (19) with (22), we obtain $\mathbb{E}[Y_jW_{j,D}]$ in (16).

Next, we derive $\mathbb{E}[Y_jW_{j,U}]$ in (15). Here, we denote $Y_{j,D}$ as the time interval when $P_{n,j-1}$ and $P_{n,j}$ arrive at the computation queue of the local server at U_n, i.e., $Y_{j,D}=t_{j,D}-t_{j-1,D}$. As $\mathbb{E}[W_{j,U}Y_j]$ is calculated as

$$\begin{aligned}\mathbb{E}[W_{j,U}Y_j]&=\int_0^\infty yf_{Y_j}(y)\mathbb{E}[W_{j,U}|Y_j=y]\mathrm{d}y\\&=\int_0^\infty\int_0^\infty yf_{Y_j}(y)f_{Y_{j,D}|Y_j}(y'|y)\\&\quad\times\mathbb{E}[W_{j,U}|Y_j=y,Y_{j,D}=y']\mathrm{d}y'\mathrm{d}y,\end{aligned}\quad(23)$$

we need to derive $f_{Y_{j,D}|Y_j}(y'|y)$ and $\mathbb{E}[W_{j,U}|Y_j=y,Y_{j,D}=y']$ to obtain $\mathbb{E}[Y_jW_{j,U}]$. Here, $W_{j,U}$ only depends on the queuing delay of $P_{n,j-1}$ at the local server, i.e.,

$$\mathbb{E}[W_{j,U}|Y_j=y,Y_{j,D}=y']=\frac{\exp(-(\mu_n-\lambda_n)y')}{\mu_n-\lambda_n}. \quad (24)$$

We then derive $f_{Y_{j,D}|Y_j}(y'|y)$ by first calculating $Y_{j,D}$ as

$$Y_{j,D} = X_{j,D} + Y_j - X_{j-1,D}. \tag{25}$$

Here, we consider two complementary events, $B_{j,D}$ and $L_{j,D}$, and derive $f_{Y_{j,D}|Y_j}(y'|y)$ as

$$f_{Y_{j,D}|Y_j}(y'|y) = p_{Y_{j,D},B_{j,D}|Y_j} + p_{Y_{j,D},L_{j,D}|Y_j}, \tag{26}$$

where $p_{Y_{j,D},B_{j,D}|Y_j}$ and $p_{Y_{j,D},L_{j,D}|Y_j}$ are given by

$$p_{Y_{j,D},B_{j,D}|Y_j} = f_{Y_{j,D}|Y_j,B_{j,D}}(y'|y, B_{j,D})\Pr(B_{j,D}) \tag{27}$$

and

$$p_{Y_{j,D},L_{j,D}|Y_j} = f_{Y_{j,D}|Y_j,L_{j,D}}(y'|y, L_{j,D})\Pr(L_{j,D}), \tag{28}$$

respectively.

When $B_{j,D}$ happens, $Y_{j,D}$ depends on the transmission time of the packets generated during Y_j at the edge server. We consider that there are $K_{j,y} = k$ packets generated during Y_j. As the transmission time of each packet follows an independent and identical exponential distribution, the total time consumed to transmit these k packets and $P_{n,j}$ follows a Gamma distribution, whose pdf is given by

$$f_{Y_{j,D}|K_{j,y}}(y'|k) = \frac{y'^k \mu_D^{k+1} \exp(-\mu_D y')}{k!}. \tag{29}$$

Thus, we calculate $p_{Y_{j,D},B_{j,D}|Y_j}$ as

$$\begin{aligned} p_{Y_{j,D},B_{j,D}|Y_j} &= f_{Y_{j,D}|Y_j,B_{j,D}}(y'|y, B_{j,D})\Pr(B_{j,D}) \\ &= \sum_{k=0}^{\infty} \Pr(K_{j,y} = k|Y_j = y) f_{Y_{j,D}|K_{j,y}}(y'|k)\Pr(B_{j,D}) \\ &= \mu_D \exp(-\mu_D(y + y') - \lambda_n y) I_0\left(2\sqrt{\lambda_{-n}\mu_D y y'}\right), \end{aligned} \tag{30}$$

where $I_0(\cdot)$ is the modified first-kind Bessel function of the zeroth order.

When $L_{j,D}$ happens, $Y_{j,D}$ depends on the time interval $Y_j - X_{j-1,D}$ and $X_{j,D}$. In particular, $X_{j,D}$ depends on the number of the packets in the transmission queue when $P_{n,j}$ is generated. We consider that when $P_{n,j}$ is generated, there are $K_{j,e} = k$ packets in the transmission queue. As the transmission time of each packet follows the independent and identical exponential distribution, the total time consumed to transmit these k packets and $P_{n,j}$ follows a Gamma distribution, whose pdf is given as

$$f_{X_{j,D}|K_{j,e}}(x_2|k) = \frac{x_2^k \mu_D^{k+1} \exp(-\mu_D x_2)}{k!}. \tag{31}$$

Thus, we calculate $p_{Y_{j,D},L_{j,D}|Y_j}$ by (28) and obtain

$$\begin{aligned} p_{Y_{j,D},L_{j,D}|Y_j} =& \frac{(\mu_D - \lambda)(\mu_D - \lambda_{-n})}{(2\mu_D - \lambda - \lambda_{-n})}\big(\exp(-(\mu_D - \lambda)(y - y')) \\ &- \exp(-(\mu_D - \lambda)y - (\mu_D - \lambda_{-n})y')\big), \end{aligned} \tag{32}$$

for $y' < y$, and

$$\begin{aligned} p_{Y_{j,D},L_{j,D}|Y_j} =& \frac{(\mu_D - \lambda)(\mu_D - \lambda_{-n})}{(2\mu_D - \lambda - \lambda_{-n})}(\exp(-(\mu_D - \lambda_{-n})(y' - y)) \\ &- \exp(-(\mu_D - \lambda)y - (\mu_D - \lambda_{-n})y')), \end{aligned} \tag{33}$$

for $y' \geq y$. By substituting (30), (32), and (33) into (26), we obtain $f_{Y_{j,D}|Y_j}(y'|y)$. Furthermore, by substituting $f_{Y_{j,D}|Y_j}(y'|y)$ and (24) into (23), we obtain

$$\begin{aligned} \mathbb{E}[W_{j,U}Y_j] =& \frac{\lambda_n(\mu_D - \lambda)(\mu_D + \mu_n - \lambda_{-n})}{\mu_n^2(\mu_D - \lambda_{-n})(\mu_n - \lambda_n)(\mu_D + \mu_n - \lambda)} \\ &+ \frac{\lambda_n(\mu_D + \mu_n - \lambda_n)}{\mu_D(\mu_n - \lambda_n)(\mu_D + \mu_n - \lambda)^2}. \end{aligned} \tag{34}$$

Finally, we obtain the final result in (10) by substituting the (16) and (34) into (15). ■

B. Edge Computing Scheme

We then derive the closed-form expression for the average AoI in the edge computing scheme, given in Theorem 2.

Theorem 2: In the edge computing scheme, the closed-form expression for the average AoI of U_n is given by

$$\begin{aligned} \Delta_{n,e} =& \frac{1}{\lambda_n} + \frac{1}{\mu_B} + \frac{1}{\mu_D} + \frac{\lambda_{-n}}{\mu_B(\mu_B - \lambda_{-n})} + \frac{\lambda_n^2 \lambda_{-n}}{\mu_B(\mu_B - \lambda_{-n})^3} \\ &+ \frac{\lambda_n^2}{(\mu_B - \lambda)(\mu_B - \lambda_{-n})^2} + \lambda_n(\Phi_{e,B_{j,E}} + \Phi_{e,L_{j,E}}), \end{aligned} \tag{35}$$

where $\Phi_{e,B_{j,E}}$ and $\Phi_{e,L_{j,E}}$ are given by (36) and (37), respectively.

Proof: In this scheme, we denote $X_{j,B}$ and $X_{j,D}$ as the queuing delay of $P_{n,j}$ in the computation queue of the edge server and the queuing delay of $P_{n,j}$ in the transmission queue, respectively. They are given by

$$X_{j,B} = t_{j,B} - t_j \tag{38}$$

and

$$X_{j,D} = t'_j - t_{j,B}, \tag{39}$$

respectively, where $t_{j,B}$ is the time that $P_{n,j}$ arrives at the transmission queue. By following the procedure in the proof of Theorem 1, we calculate the average AoI of U_n in the edge computing scheme, $\Delta_{n,e}$, as

$$\begin{aligned} \Delta_{n,e} &= \frac{1}{\lambda_n} + \lambda_n \mathbb{E}[Y_j(S_{j,B} + W_{j,B} + S_{j,D} + W_{j,D})] \\ &= \frac{1}{\lambda_n} + \frac{1}{\mu_B} + \frac{1}{\mu_D} + \lambda_n(\mathbb{E}[Y_j W_{j,B}] + \mathbb{E}[Y_j W_{j,D}])\big), \end{aligned} \tag{40}$$

where $W_{j,B}$ and $S_{j,B}$ are the waiting time and the computation time of $P_{n,j}$ in the computation queue of the edge server, respectively, and $W_{j,D}$ and $S_{j,D}$ are the waiting time and the transmission time of $P_{n,j}$ in the transmission queue, respectively. To obtain $\Delta_{n,e}$, we need to derive $\mathbb{E}[Y_j W_{j,B}]$ and $\mathbb{E}[Y_j W_{j,D}]$ in (40). Similar to the proof of Theorem 1, we obtain

$$\begin{aligned} \mathbb{E}[Y_j W_{j,B}] =& \frac{\lambda_{-n}}{\lambda_n \mu_B(\mu_B - \lambda_{-n})} + \frac{\lambda_n \lambda_{-n}}{\mu_B(\mu_B - \lambda_{-n})^3} \\ &+ \frac{\lambda_n}{(\mu_B - \lambda)(\mu_B - \lambda_{-n})^2}. \end{aligned} \tag{41}$$

We then derive $\mathbb{E}[Y_j W_{j,D}]$ in (40). We denote $Y_{j,B}$ as the time interval when $P_{n,j-1}$ and $P_{n,j}$ arrive at the transmission queue, i.e., $Y_{j,B} = t_{j,B} - t_{j-1,B}$. We then denote $B_{j,E}$ as

$$\Phi_{e,B_{j,E}} = \frac{\lambda_n(\mu_B+\mu_D-\lambda)}{\mu_B(\mu_D-\lambda)(\mu_B+\mu_D-\lambda-\lambda_{-n})^2} + \frac{\lambda_n(\mu_B-\lambda)(\mu_B-\lambda_{-n})\left(\frac{1}{(\mu_D-\lambda_{-n})^2}-\frac{1}{(\mu_B-\lambda_{-n})^2}\right)}{(\mu_B-\mu_D)(\mu_D-\lambda)(\mu_D+\mu_B-\lambda-\lambda_{-n})}$$
$$+\frac{\lambda_n\lambda_{-n}(\mu_B-\lambda)(\mu_B-\lambda_{-n})\left(\frac{2}{(\mu_D-\lambda_{-n})^3}-\frac{2}{(\mu_B-\lambda_{-n})^3}\right)}{\mu_D(\mu_B-\mu_D)(\mu_B+\mu_D-\lambda-\lambda_{-n})} + \frac{2\lambda_n\lambda_{-n}(\mu_D+\mu_B-\lambda)}{\mu_B\mu_D(\mu_D+\mu_B-\lambda-\lambda_{-n})^3}. \quad (36)$$

$$\Phi_{e,L_{j,E}} = \frac{\lambda_{-n}}{\mu_D(\mu_D-\lambda_{-n})}\left(\frac{1}{\lambda_n} - \frac{\lambda_n(\mu_B+\mu_D-\lambda)}{\mu_B(\mu_B+\mu_D-\lambda-\lambda_{-n})^2} - \frac{\lambda_n(\mu_B-\lambda)(\mu_B-\lambda_{-n})\left(\frac{1}{(\mu_B-\lambda_{-n})^2}-\frac{1}{(\mu_D-\lambda_{-n})^2}\right)}{(\mu_D-\mu_B)(\mu_B+\mu_D-\lambda-\lambda_{-n})}\right). \quad (37)$$

the event that $P_{n,j}$ arrives at the transmission queue before $P_{n,j-1}$ is transmitted to U_n and $L_{j,E}$ as the event that $P_{n,j}$ arrives at the transmission queue after $P_{n,j-1}$ is transmitted to U_n. Based on these two complementary events, we calculate $\mathbb{E}[Y_j W_{j,D}]$ as

$$\mathbb{E}[Y_j W_{j,D}] = \Phi_{e,B_{j,E}} + \Phi_{e,L_{j,E}}, \quad (42)$$

where $\Phi_{e,B_{j,E}} = \mathbb{E}[Y_j W_{j,D}|B_{j,E}]\Pr(B_{j,E})$ and $\Phi_{e,L_{j,E}} = \mathbb{E}[Y_j W_{j,D}|L_{j,E}]\Pr(L_{j,E})$. Since $\Phi_{e,B_{j,E}}$ is calculated as

$$\Phi_{e,B_{j,E}} = \mathbb{E}[Y_j(X_{j-1,D}-Y_{j,B})|B_{j,E}]\Pr(B_{j,E}) + \mathbb{E}\left[Y_j\sum_{\kappa=1}^{K_{j,y}} S_{j,\kappa,D}\middle|B_{j,E}\right]\Pr(B_{j,E}), \quad (43)$$

we follow the similar procedure in the Proof of Theorem 1 and obtain $\Phi_{e,B_{j,E}}$ given by (36). In addition, we calculate $\Phi_{e,L_{j,E}}$ as

$$\Phi_{e,L_{j,E}} = \int_0^\infty\int_0^\infty \mathbb{E}[W_{j,D}|Y_j=y, L_{j,E}]\, y f_{Y_j}(y) \times f_{Y_{j,B}|Y_j}(y'|y)\Pr(L_{j,E}|Y_{j,B}=y')\mathrm{d}y'\mathrm{d}y, \quad (44)$$

and obtain $\Phi_{e,L_{j,E}}$ given by (37). Combining (36) and (37), we obtain $\mathbb{E}[Y_j W_{j,D}]$ in (42). Finally, we obtain the final result in (35) by substituting the (41) and (42) into (40). ■

IV. Numerical Results

In this section, we present numerical results to validate our analysis in Section III. In particular, we first present numerical results in the homogeneous case where all the UEs share the same packet generation rate λ_h and the same local computation rate μ_h, i.e., $\lambda_n = \lambda_h$ and $\mu_n = \mu_h$ for $\forall n$. We then present numerical results in the heterogeneous case where the UEs have different packet generation rates and illustrate how this difference affects the average AoI.

Fig. 3 plots the average AoI of the MEC system versus the packet generation rate, λ_h. We first observe that the analytical average AoI of the MEC system tightly matches the simulation result, which demonstrates the correctness of our analytical result. We then observe that for both the local computing scheme and the edge computing scheme, the average AoI of the MEC system first decreases and then increases when λ_h increases. This observation is due to the fact that the increase in λ_h has a two-fold effect on the average AoI of the MEC system. When λ_h is small, this increase leads to a high updating rate of packets, which decreases the average AoI

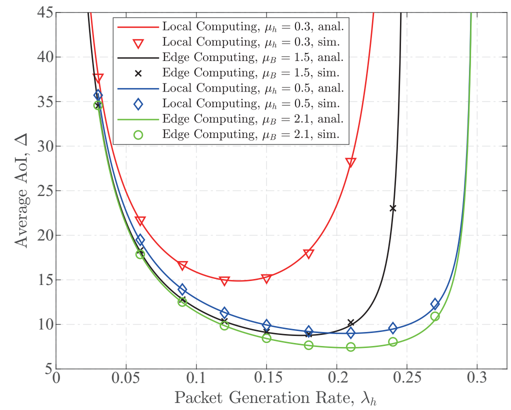


Fig. 3. The average AoI of the MEC system versus the packet generation rate, λ_h, with $N = 6$ and $\mu_D = 1.8$.

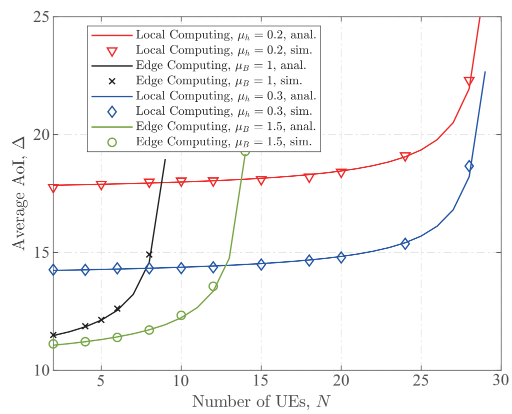


Fig. 4. The average AoI of the MEC system versus the number of UEs, N, with $\lambda_h = 0.1$ and $\mu_D = 3$.

of the MEC system. When λ_h exceeds a certain threshold, its increase leads to the significant increase in the waiting time of a packet in computation queues and the transmission queue, thereby degrading the AoI performance.

Fig. 4 plots the average AoI of the MEC system versus the number of UEs, N. We first observe that when N increases, the average AoI of the MEC system increases monotonically and this increase in the edge computing scheme is faster than in the local computing scheme. This is because that the increase in N results in the longer waiting time of a packet in both the transmission queue and the computation queue of the edge server, which increases the average AoI of the MEC system. In the edge computing scheme, the packets

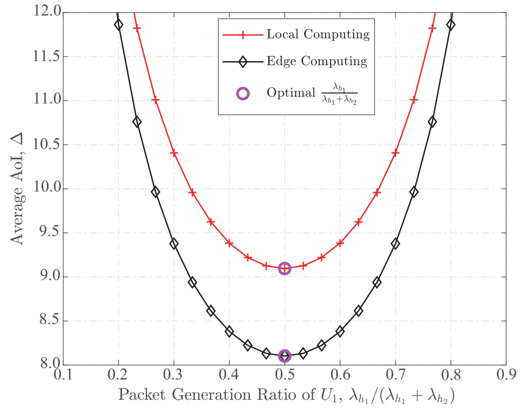


Fig. 5. The average AoI of the MEC system versus the packet generation ratio of U_1, $\lambda_{h_1}/(\lambda_{h_1}+\lambda_{h_2})$, with $N=6$, $\lambda=0.9$, $\mu_h=0.6$, $\mu_B=2$, and $\mu_D=2.4$.

computed by the edge server increases dramatically as N increases, which results in a long waiting time of a packet in the computation queue. We further observe that the edge computing scheme has a lower average AoI than the local computing scheme for small N, but a larger average AoI for large N. This is because that for a small number of UEs, compared to local computing, edge computing can provide the higher computation rate via the powerful edge server, thereby decreasing the average AoI of the MEC system. Differently, for a large number of UEs, local computing can avoid the long waiting time in the computation queue of the edge server by allocating the packets to the local server, which decreases the average AoI.

In Fig. 5, we consider the heterogeneous case where a half of UEs have the packet generation probability λ_{h_1} and the other half of UEs have the packet generation probability λ_{h_2}. We denote U_1 as a UE arbitrarily selected in the first half of UEs and U_2 as a UE arbitrarily selected in the other half of UEs. We then define the packet generation ratio of U_1 as $\frac{\lambda_{h_1}}{\lambda_{h_1}+\lambda_{h_2}}$. We assume that the BS generates two types of UEs' packets with a total rate λ, i.e., $\lambda=\frac{N}{2}(\lambda_{h_1}+\lambda_{h_2})$. From Fig. 5, we observe that for both schemes, the average AoI of the MEC system first decreases and then increases. This is because that given the total packet generation rate λ, the increase in λ_{h_1} means the increase in the packet generation rate of U_1 and the decrease in the packet generation rate of U_2. When λ_1 is small, the average AoI of U_1 dominates the average AoI of the MEC system. In this case, the increase in U_1's packet generation rate significantly reduces the average AoI of U_1, leading to a reduced average AoI. When λ_{h_1} is large, the average AoI of U_2 is large due to small λ_{h_2}, which dominates the average AoI of the MEC system. In this case, the increase in U_1's packet generation rate significantly increases the average AoI of U_2, which results in the increase in the average AoI of the MEC system. We further observe that the minimum average AoI of the MEC system is obtained when the packet generation ratio of U_1 is 0.5, i.e., $\lambda_{h_1}=\lambda_{h_2}=\frac{\lambda}{N}$. It implies that the minimum average AoI of an MEC system can be obtained when the BS uniformly generates the packet for UEs.

V. Conclusion

We analyzed the average AoI of a multi-user MEC system with two computing schemes, i.e., the local computing scheme and the edge computing scheme. We derived the closed-form expressions for the average AoI, where the packets are served in both the computation queue and the transmission queue according to the FCFS policy. From simulation results, we demonstrated the accuracy of our analysis and illustrated the impacts of different system parameters on the average AoI. Furthermore, we considered a heterogeneous case where the UEs have different packet generation rates and observed that the same packet generation rate among UEs minimized the average AoI of the MEC system.

References

[1] M. Simsek, A. Aijaz, M. Dohler, J. Sachs, and G. Fettweis, "5G-enabled tactile Internet," *IEEE J. Select. Areas Commun.*, vol. 34, no. 3, pp. 460–473, Mar. 2016.

[2] C. Li, N. Yang, and S. Yan, "Optimal transmission of short-packet communications in multiple-input single-output systems," *IEEE Trans. Veh. Technol.*, vol. 68, no. 7, pp. 7199–7203, Jul. 2019.

[3] S. Kaul, M. Gruteser, V. Rai, and J. Kenney, "Minimizing age of information in vehicular networks," in *Proc. IEEE Conf. Sensor Ad Hoc Commun. Netw.*, Salt Lake City, UT, Jun. 2011, pp. 350–358.

[4] Y. C. Hu, M. Patel, D. Sabella, N. Sprecher, and V. Young, "Mobile edge computing-A key technology towards 5G," *ETSI white paper*, vol. 11, no. 11, pp. 1–16, Sep. 2015.

[5] Y. Mao, C. You, J. Zhang, K. Huang, and K. B. Letaief, "A survey on mobile edge computing: The communication perspective," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 4, pp. 2322–2358, Aug. 2017.

[6] P. Mach and Z. Becvar, "Mobile edge computing: A survey on architecture and computation offloading," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 3, pp. 1628–1656, Mar. 2017.

[7] J. Gong, S. Zhou, and Z. Niu, "Optimal power allocation for energy harvesting and power grid coexisting wireless communication systems," *IEEE Trans. Commun.*, vol. 61, no. 7, pp. 3040–3049, Jun. 2013.

[8] S. Sardellitti, G. Scutari, and S. Barbarossa, "Joint optimization of radio and computational resources for multicell mobile-edge computing," *IEEE Trans. Signal Inf. Process. Netw.*, vol. 1, no. 2, pp. 89–103, Jun. 2015.

[9] L. Zhao, J. Wang, J. Liu, and N. Kato, "Optimal edge resource allocation in iot-based smart cities," *IEEE Netw.*, vol. 33, no. 2, pp. 30–35, Mar. 2019.

[10] B. Zhou and W. Saad, "Optimal sampling and updating for minimizing age of information in the internet of things," in *Proc. IEEE Global Commun. Conf.*, Abu Dhabi, United Arab Emirates, Dec. 2018, pp. 1–6.

[11] J. Gong, Q. Kuang, and X. Chen, "Joint transmission and computing scheduling for status update with mobile edge computing," in *Proc. IEEE Intern. Commun. Conf.*, Dublin, Ireland, Jun. 2020, pp. 1–6.

[12] X. Song, X. Qin, Y. Tao, B. Liu, and P. Zhang, "Age based task scheduling and computation offloading in mobile-edge computing systems," in *Proc. IEEE Wireless Commun. Netw. Conf.*, Marrakech, Morocco, Apr. 2019, pp. 1–6.

[13] Q. Kuang, J. Gong, X. Chen, and X. Ma, "Analysis on computation-intensive status update in mobile edge computing," *IEEE Trans. Veh. Technol.*, vol. 69, no. 4, pp. 4353–4366, Apr. 2020.

[14] H. Wu, H. Tian, S. Fan, and J. Ren, "Data age aware scheduling for wireless powered mobile-edge computing in industrial internet of things," *IEEE Trans. Industr. Inform.*, vol. 17, no. 1, pp. 398–408, Apr. 2020.

[15] X. Chen, C. Wu, T. Chen, Z. Liu, H. Zhang, M. Bennis, H. Liu, and Y. Ji, "Information freshness-aware task offloading in air-ground integrated edge computing systems," *arXiv preprint arXiv:2007.10129*, Jul. 2020.

[16] S. Sthapit, J. Thompson, N. M. Robertson, and J. R. Hopgood, "Computational load balancing on the edge in absence of cloud and fog," *IEEE Trans. Mobile Comput.*, vol. 18, no. 7, pp. 1499–1512, Jul. 2019.