2021 19th International Symposium on Modeling and Optimization in Mobile, Ad hoc, and Wireless Networks (WiOpt)

Federated Learning with Correlated Data: Taming the Tail for Age-Optimal Industrial IoT

Chen-Feng Liu and Mehdi Bennis
Centre for Wireless Communications, University of Oulu, Finland
E-mail: {chen-feng.liu, mehdi.bennis}@oulu.fi

***Abstract*—While information delivery in industrial Internet of things demands reliability and latency guarantees, the freshness of the controller's available information, measured by the age of information (AoI), is paramount for high-performing industrial automation. The problem in this work is cast as a sensor's transmit power minimization subject to the peak-AoI requirement and a probabilistic constraint on queuing latency. We further characterize the *tail behavior* of the latency by a generalized Pareto distribution (GPD) for solving the power allocation problem through Lyapunov optimization. As each sensor utilizes its own data to locally train the GPD model, we incorporate *federated learning* and propose a local-model selection approach which accounts for correlation among the sensor's training data. Numerical results show the tradeoff between the transmit power, peak AoI, and delay's tail distribution. Furthermore, we verify the superiority of the proposed correlation-aware approach for selecting the local models in federated learning over an existing baseline.**

***Index Terms*—5G and beyond, federated learning, URLLC, industrial IoT, age of information (AoI), extreme value theory.**

I. Introduction

Delivering the monitored status data with ultra-reliable low-latency communication (URLLC) and having up-to-date information at the central controller (in control systems) are pivotal in industrial Internet-of-things (IoT) networks [1]–[3]. In this regard, the age of information (AoI) [4], which is the elapsed time since the data was generated till the current time instant, has been considered as the information freshness measure for resource allocation and scheduling in industrial IoT settings [5]–[9].

A. Related Work

By assuming that the sensors update their status information over unreliable links, the work [5] focused on average AoI minimization subject to the sensors' transmit power constraints. Therein, a transmission scheduling policy was proposed. The authors in [6] studied the channel allocation problem in software-defined industrial IoT and aimed to minimize the maximal average AoI over the network. Considering that the status data is transmitted via device-to-device (D2D) communication in an industrial wireless network, Li *et al.* [7] proposed a belief-based Bayesian reinforcement learning framework in which D2D users optimize their dynamic channel and power allocation policies in a distributed manner. The objective in [7] was to maximize energy efficiency subject to AoI constraints. Moreover, a centralized [8] and a distributed [9] dynamic power allocation policy for sensors were proposed in our prior works by taking into account the statistics of the maximal AoI over time and the AoI threshold violation probability, respectively. In [8], we further investigated URLLC with respect to the information decoding error incurred by the finite blocklength transmission. Note that the end-to-end delay, including the transmission delay, queuing delay, and so forth, are incorporated in the AoI-based formulation [4]. In other words, when we allocate communication resources, the AoI performance are entangled with the delays. Furthermore, analyzing the *tail behavior* of the delay distribution is one key enabler for URLLC [10]. However, while the aforementioned works provided interesting results, little attention has been paid to the joint investigation of the AoI performance and the delay's tail distribution in state-of-the-art industrial IoT. Although AoI threshold deviation can be related to the data queue length in vehicular communication [11] in which we aimed to reduce the excess AoI/queue length, we still lacked the joint investigation of the AoI and delay.

B. Our Contribution

In this work, focusing on the uplink of an industrial IoT network with multiple sensors, we study the **power minimization problem which accounts for the peak AoI requirement and the tail distribution of the queuing delay**. Specifically, a URLLC constraint in terms of the threshold violation probability is imposed on the queuing delay whose analytic tail distribution formula is needed for allocating the sensor's transmit power via Lyapunov optimization. To address this, we invoke *extreme value theory*, by which the tail behavior can be characterized by a generalized Pareto distribution (GPD), and incorporate *federated learning* (FL) [12] in order to alleviate the sensors' overheads of finding the characteristic parameters of the GPD. The outcome of FL is affected by the correlation among the sensor's empirical data for training the GPD model. However, in most FL-aided wireless communication systems, the training data are independent [13]–[15], or the correlation among the training data is neglected [9]. Instead, we take correlation among the training data into consideration and propose a correlation-aware approach for selecting the sensors' local models in FL. We investigate the tradeoff between the average power consumption, peak AoI, and queuing delay's tail distribution by simulations. Regarding GPD-model training, the proposed model selection approach achieves a lower variance compared with the correlation-agnostic baseline.

ISBN 978-3-903176-37-9 © 2021 IFIP

II. System Model

Consider the industrial IoT network composed of a set $\mathcal{K}$ of K wireless sensors and a central controller. The sensors monitor the factory environments and send the status data to the controller. We assume that the sensors' data-sampling operations are triggered by random events. After sampling, the sensor transmits the status data immediately if the previous samples were uploaded. Otherwise, it queues in the data buffer for transmission. Let the sensor's sequentially sampled data be indexed by $n \in \mathbb{Z}^+$. Then we denote the queuing time of the nth data of sensor $k \in \mathcal{K}$ as $q_k^n \geq 0$. The total bandwidth W is orthogonally and equally allocated to all sensors. Given that the sensor k allocates transmit power P_k^n in its nth transmission, the corresponding transmission time is

$$T_k^n = \frac{KD}{W \log_2 \left(1 + \frac{K h_k^n P_k^n}{W N_0}\right)} \tag{1}$$

with data size D. Here, h_k^n is the channel gain, including path loss and channel fading, between sensor k and the controller in the nth transmission, and N_0 is the power spectral density of the additive white Gaussian noise. Fig. 1 shows the communication timeline and AoI function of sensor k. Therein, t_k^n is the time instant at which the controller receives the nth data. We denote the AoI as $a_k(t)$ which is the function of time index $t \in \mathbb{R}^+$ and measured at the controller. At time instant t_k^n, the age of the controller's newly received information, i.e., the nth data, is $q_k^n + T_k^n$. Then the information age increases linearly with time. Hence, the AoI function can be mathematically defined as

$$a_k(t) = q_k^n + T_k^n + t - t_k^n, \ \forall\, t \in [t_k^n, t_k^{n+1}), n \in \mathbb{Z}^+. \tag{2}$$

When the nth data is completely delivered to the controller, we have the peak AoI of the $(n-1)$th data (i.e., lifetime of the previous data) as

$$\begin{aligned} A_k^{n-1} &= \lim_{\tau \to 0^+} a_k(t_k^n - \tau) \\ &= a_k(t_k^{n-1}) + \max\{x_k^n - a_k(t_k^{n-1}), 0\} + T_k^n, \end{aligned} \tag{3}$$

where $x_k^n > 0$ represents the inter-arrival time between the $(n-1)$th data and nth data. Additionally, we can straightforwardly find the mathematical expression of the queuing time of sensor k's $(n+1)$th data as

$$q_k^{n+1} = \max\{q_k^n + T_k^n - x_k^{n+1}, 0\}. \tag{4}$$

Further note that x_k^{n+1} and q_k^{n+1} may be unknown when we allocate transmit power P_k^n. Finally, for each sensor $k \in \mathcal{K}$, inter-arrival time $x_k^{n+1}, \forall\, n \in \mathbb{Z}^+$, is identically distributed and can be correlated.[1] The statistics of data arrivals are identical and independent among all sensors. One applicable scenario is that various sensors separately monitor the temperatures of the identical manufacturing processes in different factories.

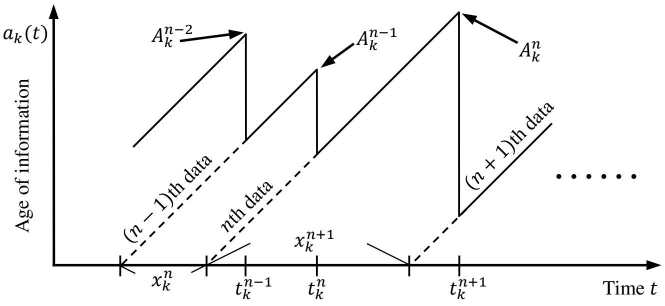

Fig. 1. Communication timeline and AoI function of sensor k.

III. Peak AoI and URLLC-Aware Power Allocation

A. Problem Formulation

Due to the continuous changes of the factory environment status, the controller's available information becomes outdated as time elapses. The aged information may further deteriorate the control system performance. In order to suppress this deficiency, we consider a cost function $f_k^n = \frac{1}{\beta}(A_k^{n-1})^\beta$ for the peak AoI and impose a long-term time-averaged constraint $\lim\limits_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} \mathbb{E}[f_k^n] \leq f_{\text{th}}, \forall\, k \in \mathcal{K}$, with a predetermined parameter $\beta \geq 1$ and the cost threshold f_{th}. Regarding the URLLC requirement, we impose a probabilistic constraint on the queuing delay in each transmission $n \in \mathbb{Z}^+$ as $\Pr\{q_k^{n+1} > q_{\text{th}} | q_k^n, h_k^n\} \leq \epsilon$, where q_{th} and ϵ are the delay threshold and tolerable threshold violation probability, respectively. Note that the concerned probability ϵ is very small. For the purpose of prolonging the battery-limited sensor's lifetime, we study a power minimization problem

$$\underset{P_k^n}{\text{minimize}} \quad \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} P_k^n \tag{5a}$$

$$\text{subject to} \quad \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \mathbb{E}[f_k^n] \leq f_{\text{th}}, \tag{5b}$$

$$\Pr\{q_k^{n+1} > q_{\text{th}} | q_k^n, h_k^n\} \leq \epsilon, \ \forall\, n \in \mathbb{Z}^+, \tag{5c}$$

$$0 \leq P_k^n \leq P_{\max}, \ \forall\, n \in \mathbb{Z}^+, \tag{5d}$$

for each sensor $k \in \mathcal{K}$, in which $P_{\max}$ is the sensor's power budget. Here, the expectation in (5b) is taken with respect to the stochastic wireless channel and inter-arrival time, whereas the conditional probability in (5c) is measured with respect to the randomness of inter-arrival time. We further note that a closed-form expression of constraint (5c) in terms of P_k^n is required for proceeding with problem (5). To address this demand, let us first rewrite (5c) as

$$\begin{aligned} \Pr\{q_k^{n+1} > q_{\text{th}} | q_k^n, h_k^n\} &= \Pr\{q_k^n + T_k^n - x_k^{n+1} > q_{\text{th}}\} \\ &= \Pr\{X > -\ln(q_k^n + T_k^n - q_{\text{th}})\} \leq \epsilon \end{aligned} \tag{6}$$

given $q_{\text{th}} > 0$, where $X = -\ln(x_k^{n+1}), \forall\, n \in \mathbb{Z}^+, k \in \mathcal{K}$. In other words, the full distribution of inter-arrival time gives the desired closed-form expression of (5c), but the distribution function of any arbitrary random variable X is not always

[1] We assume positive correlation in this work.

ISBN 978-3-903176-37-9 © 2021 IFIP

available. Since we are concerned about the tail distribution of X owing to the very small probability ϵ, we can resort to the Pickands–Balkema–de Haan theorem which asymptotically characterizes the tail behaviors of general probability distributions [16].

Theorem 1 (**Pickands–Balkema–de Haan theorem**). *Given a random variable X with the complementary cumulative distribution function (CCDF) $\bar{F}_X(x)$ and a threshold x_0, as $x_0 \to \bar{F}_X^{-1}(0)$, the conditional CCDF of the excess value $Y|_{X>x_0} = X - x_0 > 0$ can be approximated by a GPD, i.e., $\bar{F}_{Y|X>x_0}(y) = \Pr(X - x_0 > y|X > x_0) \approx (1 + \xi y/\sigma)^{-1/\xi}$, with a scale parameter $\sigma > 0$ and a shape parameter $\xi \in \mathbb{R}$.*

Thus, we consider a threshold $x_0 < -\ln\left(q_k^n + T_k^n - q_{\text{th}}\right)$ and rewrite (6) as

$$\Pr\{X > -\ln(q_k^n + T_k^n - q_{\text{th}})|X > x_0\} \leq \frac{\epsilon}{\bar{F}_X(x_0)}. \quad (7)$$

Then given $\epsilon < \bar{F}_X(x_0) \ll 1$, (7) is equivalent to the minimal transmit power requirement

$$\begin{aligned} P_k^n \geq P_{k,\min}^n = \frac{WN_0}{Kh_k^n}\Bigg[-1 + \exp\Bigg(\frac{KD\ln 2}{W} \\ \times \frac{1}{q_{\text{th}} - q_k^n + \exp\left\{\frac{\sigma}{\xi}\left[1 - \left(\frac{\epsilon}{\bar{F}_X(x_0)}\right)^{-\xi}\right] - x_0\right\}}\Bigg)\Bigg] \end{aligned} \quad (8)$$

by applying the results in Theorem 1 to (7). The characteristic parameters $\boldsymbol{\theta} \equiv (\sigma, \xi)$ of the GPD in (8) can be estimated by statistical methods while $\bar{F}_X(x_0)$ is obtained empirically. We will elaborate the approach to find $\boldsymbol{\theta}$ in Section IV. Given a specific value of $\boldsymbol{\theta}$, the power allocation problem (5) in which we replace (5c) with (8) is subsequently solved by using Lyapunov optimization [17].

B. Sensor's Transmit Power Allocation

Let us first introduce a virtual queue Z_k^n with the queue length evolution

$$Z_k^{n+1} = \max\{Z_k^n + f_k^n - f_{\text{th}}, 0\} \quad (9)$$

for the time-averaged constraint (5b). In this regard, we need to stabilize the virtual queue, i.e., $\lim\limits_{n\to\infty} \frac{\mathbb{E}[|Z_k^n|]}{n} = 0$, in order to ensure constraint (5b). Then we derive an upper bound on the conditional Lyapunov drift-plus-penalty [17] by applying $(\max\{x, 0\})^2 \leq x^2$ to (9), i.e.,

$$\begin{aligned} &\mathbb{E}\Big[\frac{1}{2}(Z_k^{n+1})^2 - \frac{1}{2}(Z_k^n)^2 + VP_k^n\Big|Z_k^n\Big] \\ &\leq \mathbb{E}\Big[\frac{1}{2}(Z_k^n + f_k^n - f_{\text{th}})^2 - \frac{1}{2}(Z_k^n)^2 + VP_k^n\Big|Z_k^n\Big] \\ &\leq \frac{1}{2}(f_{\text{th}})^2 + \mathbb{E}\Big[Z_k^n f_k^n + \frac{1}{2}(f_k^n)^2 + VP_k^n\Big|Z_k^n\Big]. \end{aligned} \quad (10)$$

To jointly stabilize the virtual queue and optimize the sensor's transmit power, we aim to minimize the upper bound (10) [17]. To this goal, the sensor k solves

$$\underset{P_{k,\min}^n \leq P_k^n \leq P_{\max}}{\text{minimize}} \quad \frac{Z_k^n\left(c_k^n + T_k^n\right)^\beta}{\beta} + \frac{\left(c_k^n + T_k^n\right)^{2\beta}}{2\beta^2} + VP_k^n \quad (11)$$

in each transmission n with the constant $c_k^n = a_k(t_k^{n-1}) + \max\{x_k^n - a_k(t_k^{n-1}), 0\}$. Here, $V \geq 0$ is a parameter trading off AoI reduction and the optimality of power consumption. Note that the convexity of problem (11) can be straightforwardly verified. Thus, via differentiation, we obtain the sensor's transmit power in the nth transmission as $P_k^{n*} = \max\{\min\{\tilde{P}_k^n, P_{\max}\}, P_{k,\min}^n\}$ in which $\tilde{P}_k^n$ satisfies

$$\begin{aligned} V = &\frac{K^2 Dh_k^n \ln 2}{W\left[\ln\left(1 + \frac{Kh_k^n\tilde{P}_k^n}{WN_0}\right)\right]^2\left(WN_0 + Kh_k^n\tilde{P}_k^n\right)} \\ &\times \Bigg[Z_k^n\Bigg(c_k^n + \frac{KD}{W\log_2\left(1 + \frac{Kh_k^n\tilde{P}_k^n}{WN_0}\right)}\Bigg)^{\beta-1} \\ &+ \frac{1}{\beta}\Bigg(c_k^n + \frac{KD}{W\log_2\left(1 + \frac{Kh_k^n\tilde{P}_k^n}{WN_0}\right)}\Bigg)^{2\beta-1}\Bigg]. \end{aligned} \quad (12)$$

After sending the status data, sensor k's updates A_k^{n-1}, Z_k^{n+1}, and q_k^{n+1} for the next transmission $n + 1$.

IV. Federated Learning with Correlated Data

A. Federated GPD-Model Learning

Assume that the sensor collects some historical data of the inter-arrival time x to estimate the GPD model $\boldsymbol{\theta}$ before proceeding with problem (5). Given the set $\mathcal{Y}_k = \{y : y|_{-\ln(x_k)>x_0} = -\ln(x_k) - x_0\}, \forall\, k \in \mathcal{K}$, of the empirical data of exceedances, each sensor k locally finds the GPD distribution which is the closest to the empirical distribution of $\mathcal{Y}_k$ in terms of the Kullback–Leibler (KL) divergence $D(\mathcal{Y}_k||\phi) = \sum_{y\in\mathcal{Y}_k} \frac{1}{|\mathcal{Y}_k|} \ln\left(\frac{1/|\mathcal{Y}_k|}{\phi(\boldsymbol{\theta}|y)\mathrm{d}y}\right)$. Here, $\phi(\boldsymbol{\theta}|y) = \frac{1}{\sigma}\left(1 + \frac{\xi y}{\sigma}\right)^{-(1+1/\xi)}$ is the likelihood function. To this goal, we minimize the KL divergence as $\max\limits_{\boldsymbol{\theta}} \frac{1}{|\mathcal{Y}_k|} \sum_{y\in\mathcal{Y}_k} \ln\phi(\boldsymbol{\theta}|y)$ which can be solved via gradient ascent. That is, each sensor k iteratively updates

$$\boldsymbol{\theta}_k^j = \boldsymbol{\theta}_k^{j-1} + \frac{\gamma}{|\mathcal{Y}_k|}\sum_{y\in\mathcal{Y}_k} \nabla_{\boldsymbol{\theta}} \ln\phi(\boldsymbol{\theta}_k^{j-1}|y) \quad (13)$$

with the learning rate γ and gradient

$$\nabla_{\boldsymbol{\theta}} \ln\phi(\boldsymbol{\theta}|y) = \left(\frac{\xi + 1}{\frac{\sigma^2}{y} + \sigma\xi} - \frac{1}{\sigma}, \frac{1}{\xi^2}\ln\left(1 + \frac{\xi y}{\sigma}\right) - \frac{1 + \frac{1}{\xi}}{\frac{\sigma}{y} + \xi}\right).$$

Additionally, we let all sensors have an identical initial value $\boldsymbol{\theta}^0$ in gradient ascent. Note that $\mathcal{Y}_k$ is composed of the exceedance data for tail distribution characterization. Hence, given a moderate[2] data-collecting time duration, the sensor may not have enough data to achieve a sufficiently accurate estimation. Although a more accurate GPD model can be obtained by aggregating all sensors' local data at the central controller, uploading the local data incurs extra transmit power which is precious for the battery-limited sensor. In order to diminish the overhead while preserving the controller's global view, we adopt the FL framework in which the sensors instead upload their locally-trained GPD models $\boldsymbol{\theta}_k^J, \forall\, k \in \mathcal{K}$, after

[2] If we consider the online GPD-model training for problem (5), the URLLC constraint (5c) cannot be addressed within this duration.

the convergence in (13) is achieved, e.g., the completion of J iterations. Then the controller finds the global GPD model $\boldsymbol{\theta}_{\rm GL}=\frac{\sum_{k\in\mathcal{K}}|\mathcal{Y}_k|\boldsymbol{\theta}_k^J}{\sum_{k\in\mathcal{K}}|\mathcal{Y}_k|}$ by weighted average [12] and feeds it back to the sensors.

B. Correlation-Aware Local-Model Selection

The global model $\boldsymbol{\theta}_{\rm GL}$ and all local models $\boldsymbol{\theta}_k^J, \forall\, k\in\mathcal{K}$, are stochastic due to the randomness of the empirical data in $\mathcal{Y}_k$. As a consequence, the variance[3] of the global model, i.e.,

$$\mathrm{Var}(\boldsymbol{\theta}_{\rm GL})=\frac{\sum_{k\in\mathcal{K}}|\mathcal{Y}_k|^2\mathrm{Var}(\boldsymbol{\theta}_k^J)}{(\sum_{k\in\mathcal{K}}|\mathcal{Y}_k|)^2}, \tag{14}$$

will affect the performance of power consumption, peak AoI, and queuing delay of the studied industrial IoT system. To deduce the details of the variance $\mathrm{Var}(\boldsymbol{\theta}_k^J)$, let us intuitively express

$$\boldsymbol{\theta}_k^J=g\bigg(\frac{\gamma}{|\mathcal{Y}_k|}\sum_{y\in\mathcal{Y}_k}\nabla_{\boldsymbol{\theta}}\ln\phi(\boldsymbol{\theta}^0|y)\bigg) \tag{15}$$

based on (13) with a function $g(\cdot)$. By further referring to [18]

$$\mathrm{Var}(g(X))\approx[g'(\mathbb{E}[X])]^2\mathrm{Var}(X), \tag{16}$$

we can derive

$$\mathrm{Var}(\boldsymbol{\theta}_k^J)\approx\frac{\kappa\gamma^2}{|\mathcal{Y}_k|^2}\mathrm{Var}\bigg(\sum_{y\in\mathcal{Y}_k}\nabla_{\boldsymbol{\theta}}\ln\phi(\boldsymbol{\theta}^0|y)\bigg) \tag{17}$$

with $\kappa=[g'(\gamma\mathbb{E}[\nabla_{\boldsymbol{\theta}}\ln\phi(\boldsymbol{\theta}^0|y)])]^2$ and, moreover,

$$\begin{aligned}\mathrm{Var}\bigg(\sum_{y\in\mathcal{Y}_k}\nabla_{\boldsymbol{\theta}}\ln\phi(\boldsymbol{\theta}^0|y)\bigg)&=|\mathcal{Y}_k|\mathrm{Var}\big(\nabla_{\boldsymbol{\theta}}\ln\phi(\boldsymbol{\theta}^0|y)\big)\\&+\sum_{y,\tilde{y}\in\mathcal{Y}_k|y\neq\tilde{y}}\mathrm{Cov}(\nabla_{\boldsymbol{\theta}}\ln\phi(\boldsymbol{\theta}^0|y),\nabla_{\boldsymbol{\theta}}\ln\phi(\boldsymbol{\theta}^0|\tilde{y})\big).\end{aligned} \tag{18}$$

Algorithm 1 Local-Model Selection for Federated Learning

```
Initialize $\boldsymbol{\eta}$ and calculate the cost $\Psi(\boldsymbol{\eta})$.
repeat
    Based on $\boldsymbol{\eta}$, find an alternative $\tilde{\boldsymbol{\eta}}$ and calculate $\Psi(\tilde{\boldsymbol{\eta}})$.
    if $\Psi(\tilde{\boldsymbol{\eta}})<\Psi(\boldsymbol{\eta})$ then
        Update $\boldsymbol{\eta}\leftarrow\tilde{\boldsymbol{\eta}}$ and $\Psi(\boldsymbol{\eta})\leftarrow\Psi(\tilde{\boldsymbol{\eta}})$.
    end if
until No alternative $\tilde{\boldsymbol{\eta}}$ with the smaller cost $\Psi(\tilde{\boldsymbol{\eta}})$ exists.
```

TABLE I
SIMULATION PARAMETERS

Para.	Value	Para.	Value	Para.	Value
K	50	W	1 MHz	$P_{\max}$	10 dBm
β	1	D	10 kbit	N_0	-174 dBm/Hz
γ	0.01	ϵ	10^{-4}	x_0	$\bar{F}_X^{-1}(0.01)$
$f_{\rm th}$	0.25	J	3000	$q_{\rm th}$	0.2 sec

If the inter-arrival time is correlated as assumed in Section II, the covariance will be larger than zero. A stronger correlation between the empirical data further increases the variance $\mathrm{Var}(\boldsymbol{\theta}_{\rm GL})$. Motivated by this, we select (a part of sensors') local models for weighted average in FL by accounting for the data correlation. To this end, let us consider a discrete-time stochastic process $\{X_t\}$. The process is long-range dependent (LRD) if the normalized auto-covariance function decays hyperbolically in the asymptotic manner, i.e., $\mathrm{K}_{XX}(m)/\mathrm{Var}(X)\sim m^{-\alpha}$ with $0<\alpha<1$. The process is short-range dependent (SRD) if the auto-covariance function decays exponentially or faster. The dependence feature is also reflected by the Hurst exponent H. For the LRD process, we have $0.5<H<1$ and $\alpha=2-2H$ [19]. The dependence is stronger as $H\to1$. Additionally, the SRD process has $H=0.5$. The Hurst exponent can be found via the rescaled range (R/S) analysis [19]. Applying $\mathrm{K}_{XX}(m)\sim\mathrm{Var}(X)m^{2H-2}$ to $\mathrm{Cov}(\cdot,\cdot)$ in (18) and incorporating (14), (17), and (18), we derive

$$\begin{aligned}&\frac{\mathrm{Var}(\boldsymbol{\theta}_{\rm GL})}{\mathrm{Var}(\nabla_{\boldsymbol{\theta}}\ln\phi(\boldsymbol{\theta}^0|y))}\\&\quad\leq\frac{\kappa\gamma^2\sum_{k\in\mathcal{K}}(|\mathcal{Y}_k|+2\sum_{i=1}^{|\mathcal{Y}_k|}\sum_{m=1}^{|\mathcal{Y}_k|-i}m^{2H_k-2})}{(\sum_{k\in\mathcal{K}}|\mathcal{Y}_k|)^2}\end{aligned} \tag{19}$$

in which the inequality is established since the exceedance data of inter-arrival time are acquired intermittently. Subsequently, referring to (19), we define a cost function

$$\Psi(\boldsymbol{\eta})=\frac{\sum_{k\in\mathcal{K}}\eta_k(|\mathcal{Y}_k|+2\sum_{i=1}^{|\mathcal{Y}_k|}\sum_{m=1}^{|\mathcal{Y}_k|-i}m^{2H_k-2})}{(\sum_{k\in\mathcal{K}}\eta_k|\mathcal{Y}_k|)^2}$$

and focus on the variance minimization problem

$$\underset{\eta_k\in\{0,1\}}{\text{minimize}}\quad\Psi(\boldsymbol{\eta}) \tag{20}$$

for selecting the local models. $\boldsymbol{\eta}=(\eta_k:k\in\mathcal{K})$ is the model selection vector. In (20), we neglect m^{2H_k-2} if sensor k has the SRD data. Note that using the time-consuming exhaustive search to solve problem (20) requires us to check all 2^K values of the objective. Alternatively, we invoke the notion of swap matching [20] in matching theory whose complexity is in the order of $\mathcal{O}(K^2)$ [21]. Let us illustrate the swap matching-based method as follows. Firstly we are given a specific vector $\boldsymbol{\eta}$. Additionally consider another vector $\tilde{\boldsymbol{\eta}}$ by either altering the value of a randomly-chosen element η_k in $\boldsymbol{\eta}$ or choosing a pair $(\eta_k,\eta_{k'})=(1,0)$ of $\boldsymbol{\eta}$ and swapping their values as $(\eta_k,\eta_{k'})=(0,1)$. If $\Psi(\tilde{\boldsymbol{\eta}})<\Psi(\boldsymbol{\eta})$, replace $\boldsymbol{\eta}$ with $\tilde{\boldsymbol{\eta}}$. We repeatedly check whether an alternative model selection vector $\tilde{\boldsymbol{\eta}}$ with the smaller cost $\Psi(\tilde{\boldsymbol{\eta}})$ exists for the current $\boldsymbol{\eta}$. The steps of the proposed correlation-aware model selection approach for FL are outlined in Algorithm 1. After finding the solution $\boldsymbol{\eta}^*$, the global model is calculated as $\boldsymbol{\theta}_{\rm GL}=\frac{\sum_{k\in\mathcal{K}}\eta_k^*|\mathcal{Y}_k|\boldsymbol{\theta}_k^J}{\sum_{k\in\mathcal{K}}\eta_k^*|\mathcal{Y}_k|}$.

V. NUMERICAL RESULTS

In simulations, we consider the path loss model $33\log x+20\log 2.625+32$ (dB) at the 2.625 GHz carrier frequency [22]

[3] For notational simplicity, $\mathrm{Var}(\boldsymbol{\theta})$ represents the variance of σ or ξ.

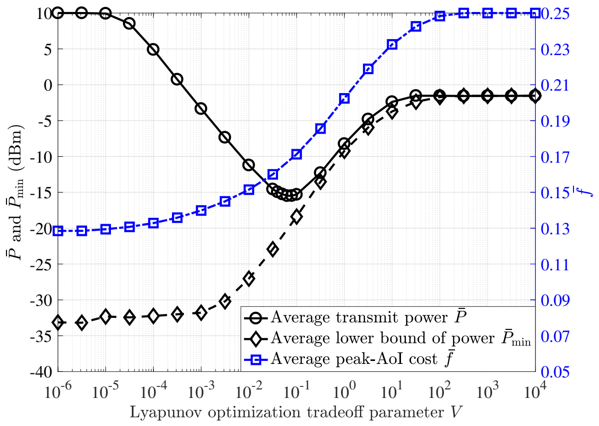

Fig. 2. Average transmit power, average lower bound of transmit power, and average peak-AoI cost versus V.

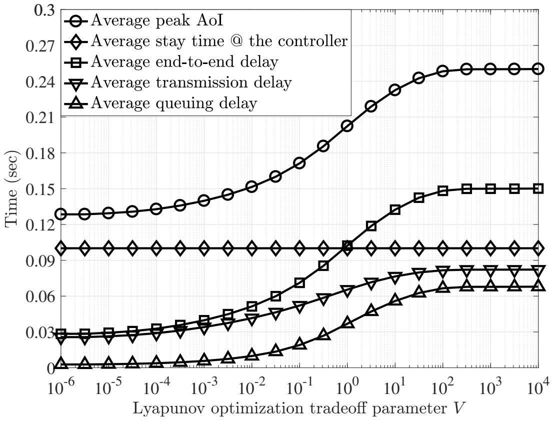

Fig. 3. Average peak AoI, average stay time of the status data at the controller, and average delays versus V.

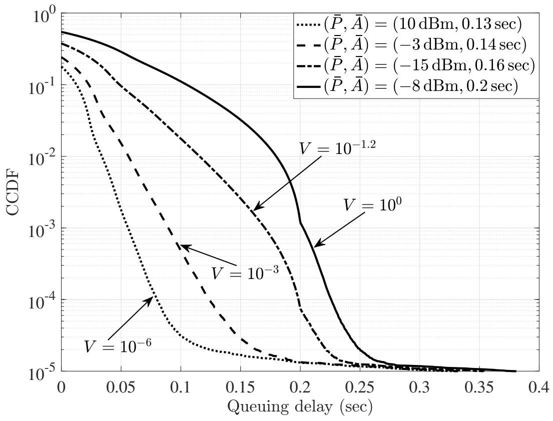

Fig. 4. CCDF of the queuing delay for various V.

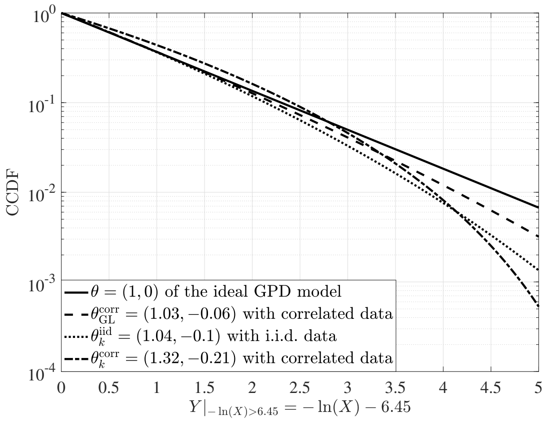

Fig. 5. CCDFs of the ideal and estimated GPD models.

in which $x=15\,\mathrm{m}$ represents the distance between the sensor and controller. The inter-arrival time follows a folded normal distribution with the mean 0.1 sec. The rest of the simulation parameters are listed in Table I.

In Figs. 2 and 3, we show the average performance of the sensor's transmit power, information age, and delays by varying the tradeoff parameter V. It can be straightforwardly understood that raising V decreases the sensor's transmit power at the expense of the higher information age as per problem (11). Note that the lower transmit power results in the higher transmission delay which consequently increases the queuing delay of the next status data. Accordingly, the average age cost $\bar{f}$, peak AoI, end-to-end delay, transmission delay, and queuing delay monotonically increase with V. Since the queuing delay is recursively related and affected by the transmission delay of the previous status data, lowering the transmit power (i.e., increasing V) has the higher impacts on the queuing delay in contrast with the transmission delay. Additionally, due to the higher power requirement (8) of the status data with a higher queuing delay, the average lower bound $\bar{P}_{\min}$ of the sensor's transmit power increases with V as shown in Fig. 2. When the tradeoff parameter V is larger than $10^{-1.2}$, constraint (8) dominates in the power minimization problem (11), making average transmit power $\bar{P}$ and $\bar{P}_{\min}$ almost coincide. Owing to this rationale, the curve of $\bar{P}$ shows cavity. Thus, unlike most Lyapunov optimization-enabled resource allocation policies in which the optimal solutions are asymptotically obtained by letting $V\to\infty$, our optimal average power consumption is achieved at a finite V, i.e., $10^{-1.2}$, in the simulated setting. Fig. 3 also shows the average stay time of the status data at the controller which is equal to the average inter-arrival time at the sensor. That is, the controller's data-updating frequency is identical to the sensor's data-sampling frequency. Let us further investigate the CCDF of the queuing delay in Fig. 4. Therein, the smaller $\bar{A}$, i.e., average peak AoI, contains not only the lower average queuing delay but also the steeper decay in the tail distribution. In contrast with the case $V=10^{-1.2}$, the probabilistic queuing delay constraint in the case $V=10^{0}$ is not satisfied even though the average power consumption is higher. This is caused by the inefficient power utilization when $V>10^{-1.2}$.

Subsequently, by considering that $K=10$ sensors have

ISBN 978-3-903176-37-9 © 2021 IFIP

TABLE II
STANDARD DEVIATION OF $\boldsymbol{\theta}_{\rm GL} = (\sigma_{\rm GL}, \xi_{\rm GL})$

$\lvert\mathcal{Y}_k\rvert$	Proposed	**FedAvg**
5	$(0.4549, 0.2951)$	$(0.4817, 0.3547)$
10	$(0.2341, 0.1428)$	$(0.2579, 0.1453)$
20	$(0.1247, 0.0832)$	$(0.1254, 0.0842)$

the status data with correlated inter-arrival time, Fig. 5 shows the CCDFs of the ideal and estimated GPD models of the exceedances $Y|_{-\ln(X)>x_0} = -\ln(X) - x_0$. Therein, the ideal GPD parameters of the simulated setting are $\boldsymbol{\theta} = (1, 0)$. All sensors have the identical correlation strength. For each sensor, the data amount of exceedances is $|\mathcal{Y}_k| = 80$. In Fig. 5, we consider the CCDF of the sensor's GPD model $\boldsymbol{\theta}_k^{\rm corr}$ which has the largest deviation in the distribution tail among all sensors. In contrast with the case $\boldsymbol{\theta}_k^{\rm iid}$ in which the inter-arrival time is independent and identically distributed (i.i.d.), the correlation has a higher impact on the estimation of the GPD parameters. As expected, the estimation accuracy with data correlation can be improved by leveraging the proposed approach for FL. In this regard, the learned global GPD model $\boldsymbol{\theta}_{\rm GL}^{\rm corr}$ is closer to the ideal GPD model. Finally, we compare our proposed model selection approach with the correlation-agnostic baseline **FedAvg** in Table II which shows the standard deviations of the learned global GPD models. Note that all sensors' correlation strengths are different, and the baseline has $\eta_k^* = 1, \forall k \in \mathcal{K}$ [12]. For each sensor, the total data number of inter-arrival time (for finding the Hurst exponent by the R/S analysis) is approximately $100 \cdot |\mathcal{Y}_k|$ since we set $x_0 = \bar{F}_X^{-1}(0.01)$. Verified by the results, our proposed local-model selection approach achieves a lower standard deviation of the global GPD model $\boldsymbol{\theta}_{\rm GL}$ when the sensor has less data samples of exceedances. In this regime, data correlation has a higher impact on the GPD-model learning.

VI. CONCLUSION

In this work, we jointly took into account the peak AoI and delay distribution tail while allocating the sensor's transmit power by Lyapunov optimization. The studied problem was formulated as a transmit power minimization in which the tail distribution is approximated as a GPD. We have further incorporated the FL framework for training the GPD model and proposed a correlation-aware local-model selection approach for FL. Finally, we have investigated the power-delay-AoI tradeoff and verified the effectiveness of our correlation-aware approach for FL.

ACKNOWLEDGMENTS

This research was supported by the Academy of Finland project MISSION, the Academy of Finland project SMARTER, the CHIST-ERA project LeadingEdge under Grant CHIST-ERA-18-SDCDN-004, the CHIST-ERA project CONNECT, the INFOTECH project NOOR, and the Nokia Bell Labs project NEGEIN.

REFERENCES

[1] 5G Alliance for Connected Industries and Automation, "White paper: 5G for connected industries and automation," 5G-ACIA, Tech. Rep., Feb. 2019, 2nd ed.

[2] G. Zhao, M. A. Imran, Z. Pang, Z. Chen, and L. Li, "Toward real-time control in future wireless networks: Communication-control co-design," *IEEE Commun. Mag.*, vol. 57, no. 2, pp. 138–144, Feb. 2019.

[3] S. Vitturi, C. Zunino, and T. Sauter, "Industrial communication systems and their future challenges: Next-generation Ethernet, IIoT, and 5G," *Proc. IEEE*, vol. 107, no. 6, pp. 944–961, Jun. 2019.

[4] S. Kaul, M. Gruteser, V. Rai, and J. Kenney, "Minimizing age of information in vehicular networks," in *Proc. 8th Annu. IEEE Commun. Soc. Conf. Sensor, Mesh Ad Hoc Commun. Netw.*, Jun. 2011, pp. 350–358.

[5] Q. Wang, H. Chen, Y. Li, Z. Pang, and B. Vucetic, "Minimizing age of information for real-time monitoring in resource-constrained industrial IoT networks," in *Proc. IEEE 17th Int. Conf. Ind. Informat.*, Jul. 2019, pp. 1766–1771.

[6] B. Liu, C. Hua, and P. Gu, "Age of information aware channel allocation for wireless industrial networks," in *Proc. 11th Int. Conf. Wireless Commun. Signal Process.*, Oct. 2019, pp. 1–6.

[7] M. Li, C. Chen, C. Hua, and X. Guan, "Learning-based autonomous scheduling for AoI-aware industrial wireless networks," *IEEE Internet Things J.*, vol. 7, no. 9, pp. 9175–9188, Sep. 2020.

[8] C.-F. Liu and M. Bennis, "Taming the tail of maximal information age in wireless industrial networks," *IEEE Commun. Lett.*, vol. 23, no. 12, pp. 2442–2446, Dec. 2019.

[9] Y.-L. Hsu, C.-F. Liu, S. Samarakoon, H.-Y. Wei, and M. Bennis, "Age-optimal power allocation in industrial IoT: A risk-sensitive federated learning approach," in *Proc. IEEE 32nd Annu. Int. Symp. Pers., Indoor, Mobile Radio Commun.*, Sep. 2021, pp. 1–6.

[10] M. Bennis, M. Debbah, and H. V. Poor, "Ultrareliable and low-latency wireless communication: Tail, risk, and scale," *Proc. IEEE*, vol. 106, no. 10, pp. 1834–1853, Oct. 2018.

[11] M. K. Abdel-Aziz, S. Samarakoon, C.-F. Liu, M. Bennis, and W. Saad, "Optimized age of information tail for ultra-reliable low-latency communications in vehicular networks," *IEEE Trans. Commun.*, vol. 68, no. 3, pp. 1911–1924, Mar. 2020.

[12] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Proc. 20th Int. Conf. Artificial Intell. Statistics*, vol. 54, Apr. 2017, pp. 1273–1282.

[13] A. M. Elbir and S. Coleri, "Federated learning for hybrid beamforming in mm-Wave massive MIMO," *IEEE Commun. Lett.*, vol. 24, no. 12, pp. 2795–2799, Dec. 2020.

[14] A. M. Elbir and S. Coleri, "Federated learning for channel estimation in conventional and IRS-assisted massive MIMO," *CoRR*, vol. abs/2008.10846, pp. 1–13, Aug. 2020.

[15] S. Wang, M. Chen, C. Yin, W. Saad, C. S. Hong, S. Cui, and H. V. Poor, "Federated learning for task and resource allocation in wireless high altitude balloon networks," *IEEE Internet Things J.*, vol. 8, 2021, to be published.

[16] S. Coles, *An Introduction to Statistical Modeling of Extreme Values*. London, U.K.: Springer, 2001.

[17] M. J. Neely, *Stochastic Network Optimization with Application to Communication and Queueing Systems*. San Rafael, CA, USA: Morgan and Claypool, Jun. 2010.

[18] H. Benaroya, S. M. Han, and M. Nagurka, *Probability models in engineering and science*. CRC Press, 2005.

[19] J. Beran, R. Sherman, M. S. Taqqu, and W. Willinger, "Long-range dependence in variable-bit-rate video traffic," *IEEE Trans. Commun.*, vol. 43, no. 2/3/4, pp. 1566–1579, Feb./Mar./Apr. 1995.

[20] E. Bodine-Baron, C. Lee, A. Chong, B. Hassibi, and A. Wierman, "Peer effects and stability in matching markets," in *Proc. 4th Int. Symp. Algorithmic Game Theory*, 2011, pp. 117–129.

[21] C.-F. Liu, M. Bennis, M. Debbah, and H. V. Poor, "Dynamic task offloading and resource allocation for ultra-reliable low-latency edge computing," *IEEE Trans. Commun.*, vol. 67, no. 6, pp. 4132–4150, Jun. 2019.

[22] Radiocommunication Sector of ITU, "P.1238-10: Propagation data and prediction methods for the planning of indoor radiocommunication systems and radio local area networks in the frequency range 300 MHz to 450 GHz," ITU-R, Tech. Rep., Aug. 2019.