APPENDIX FOR THE PAPER "TO TALK OR TO WORK: DYNAMIC BATCH SIZES ASSISTED TIME EFFICIENT FEDERATED LEARNING OVER FUTURE MOBILE EDGE DEVICES"

A. *Proof of Theorem 1*

We set the batch size in each communication round τ, $\tau \in \{0, 1, ..., T/H-1\}$ as $\lfloor \beta^{\tau} B_0 \rfloor$ and assume T/H is an integer. Therefore, in Alg. 1, we must have $H\sum_{\tau=0}^{T/H-1} \lfloor \beta^{\tau} B_0 \rfloor \geq E$, which future implies $H\sum_{\tau=0}^{T/H-1} \beta^{\tau} B_0 \geq E$. Through the summation formula of the geometric series, We can observe that $\beta^{T/H} \geq \frac{E(\beta-1)}{HB_0}+1$, which yields,

$$\begin{aligned}
T \geq & H log_{\beta}\left(\frac{E(\beta-1)}{HB_0}+1\right) \\
\overset{(a)}{\geq} & log_{\beta}\left(\frac{E(\beta-1)}{B_0}+1\right) \\
\geq & log_{\beta}\left(\frac{E(\beta-1)}{B_0}\right),
\end{aligned} \tag{1}$$

where (a) follows by $(1+x)^r \geq 1+rx$ for $x \geq -1, r \in \mathbb{R}\backslash(0,1)$.

Under Assumption 1, we have

$$\mathbb{E}\left[f\left(\overline{\mathbf{w}}^{t}\right)\right] \leq \mathbb{E}\left[f\left(\overline{\mathbf{w}}^{t-1}\right)\right] + \mathbb{E}\left[\left\langle \nabla f\left(\overline{\mathbf{w}}^{t-1}\right), \overline{\mathbf{w}}^{t} - \overline{\mathbf{w}}^{t-1}\right\rangle\right] + \frac{L}{2}\mathbb{E}\left[||\overline{\mathbf{w}}^{t} - \overline{\mathbf{w}}^{t-1}||^2\right]. \tag{2}$$

Through the convergence analysis provided in other works, we can bound the terms in the above inequality as

$$\begin{aligned}
& \mathbb{E}\left[||\nabla f\left(\overline{\mathbf{w}}^{t-1}\right)||^2\right] \\
\leq & \frac{2}{\gamma}\left(\mathbb{E}\left[f\left(\overline{\mathbf{w}}^{t-1}\right)\right] - \mathbb{E}\left[f\left(\overline{\mathbf{w}}^{t}\right)\right]\right) + 4\gamma^2 H^2 \delta^2 L^2 + \frac{L\gamma\sigma^2}{KB_t} \\
\leq & \frac{2}{\gamma}\left(\mathbb{E}\left[f\left(\overline{\mathbf{w}}^{t-1}\right)\right] - \mathbb{E}\left[f\left(\overline{\mathbf{w}}^{t}\right)\right]\right) + CH^2 + \frac{L\gamma\sigma^2}{K\left\lfloor B_0 \beta^{\lfloor t/H \rfloor}\right\rfloor} \\
\overset{(a)}{\leq} & \frac{2}{\gamma}\left(\mathbb{E}\left[f\left(\overline{\mathbf{w}}^{t-1}\right)\right] - \mathbb{E}\left[f\left(\overline{\mathbf{w}}^{t}\right)\right]\right) + CH^2 + \frac{2L\gamma\sigma^2}{KB_0 \beta^{\lfloor t/H \rfloor}},
\end{aligned} \tag{3}$$

where we set $C = 4\gamma^2\delta^2L^2$, and (a) follows by $\lfloor x \rfloor > \frac{1}{2}x$ if $x > 2$. After summing the above inequality over $t \in \{1, 2, ..., T\}$ iterations and dividing it by T, we have

$$
\begin{aligned}
&\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[||\nabla f\left(\overline{\mathbf{w}}^{t-1}\right)||^2\right] \\
\leq &\frac{2}{\gamma T}\left(f\left(\overline{\mathbf{w}}^{0}\right)-f^*\right)+CH^2+\frac{2L\gamma\sigma^2}{TKB_0}\left(H\sum_{\tau=1}^{T/H}\frac{1}{\beta^{\tau-1}}-1+\frac{1}{\beta^{T/H}}\right) \\
\leq &\frac{2}{\gamma T}\left(f\left(\overline{\mathbf{w}}^{0}\right)-f^*\right)+CH^2+\frac{2L\gamma\sigma^2 H}{TKB_0}\sum_{\tau=1}^{T/H}\frac{1}{\beta^{\tau-1}} \\
\overset{(a)}{\leq} &\frac{2}{\gamma T}\left(f\left(\overline{\mathbf{w}}^{0}\right)-f^*\right)+CH^2+\frac{2L\gamma\sigma^2 H}{TKB_0}\frac{\beta}{\beta-1},
\end{aligned}
\tag{4}
$$

where (a) follows by simplifying the sum of the proportional sequence and $\frac{1}{\beta}<1$. Next, Substituting (1) into (4) yields

$$
\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[||\nabla f\left(\overline{\mathbf{w}}^{t-1}\right)||^2\right]\leq\frac{2\left(f\left(\overline{\mathbf{w}}^{0}\right)-f^*\right)}{\gamma\log_\beta\frac{E(\beta-1)}{B_0}}+4\gamma^2H^2\delta^2L^2+\frac{2L\gamma\sigma^2H}{\log_\beta\frac{E(\beta-1)}{B_0}KB_0}\cdot\frac{\beta}{\beta-1}.
\tag{5}
$$

B. *Proof of Corollary 1*

Similar to the settings in a fixed batch size scenario, if we choose $\gamma=\frac{\sqrt{K}}{L\sqrt{\log_\beta\frac{E(\beta-1)}{B_0}}}$, $H\leq\left(\log_\beta\frac{E(\beta-1)}{B_0}\right)^{\frac{1}{4}}K^{-\frac{3}{4}}$, $K<(\log E)^{\frac{5}{3}}$, and substitute them into the (5), we have

$$
\begin{aligned}
&\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[||\nabla f\left(\overline{\mathbf{w}}^{t-1}\right)||^2\right] \\
\leq &\frac{2L}{\sqrt{K\log_\beta\frac{E(\beta-1)}{B_0}}}\left(f\left(\overline{\mathbf{w}}^{0}\right)-f^*\right)+\frac{4\delta^2}{\sqrt{K\log_\beta\frac{E(\beta-1)}{B_0}}}+\frac{\sigma^2}{K^{\frac{5}{4}}(\log_\beta\frac{E(\beta-1)}{B_0})^{\frac{5}{4}}B_0}\cdot\frac{\beta}{\beta-1} \\
= &O\left(\frac{1}{\sqrt{K\log E}}\right)+O\left(\frac{1}{\sqrt{K\log E}}\right)+O\left(\frac{1}{(K\log E)^{\frac{5}{4}}}\right) \\
\overset{(a)}{=} &O\left(\frac{1}{\sqrt{K\log E}}\right).
\end{aligned}
\tag{6}
$$

Since the term $\frac{1}{(K\log E)^{\frac{5}{4}}}$ is dominated by the term $\frac{1}{\sqrt{K\log E}}$, $O\left(\frac{1}{(K\log E)^{\frac{5}{4}}}\right)$ decays faster than $O\left(\frac{1}{\sqrt{K\log E}}\right)$. In this way, (a) will hold and Alg. 1 has the convergence rate of order $O\left(\frac{1}{\sqrt{K\log E}}\right)$.

C. *Proof of Corollary 2*

If we achieve the ϵ global convergence accuracy, i.e.,

$\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[||\nabla f\left(\overline{\mathbf{w}}^{t-1}\right)||^2\right]\leq\epsilon$, we will have

$$\epsilon = \frac{2}{\gamma T}\left(f\left(\overline{\mathbf{w}}^{0}\right)-f^*\right)+4\gamma^2H^2\delta^2L^2+\frac{2L\gamma\sigma^2H}{TKB_0}\frac{\beta}{\beta-1}.$$

Then, we have

$$\begin{aligned}
M_\epsilon =& \frac{T}{H}=\frac{2bB_0K(\beta-1)+2\gamma^2\sigma^2\beta HL}{B_0H\gamma K(\beta-1)\left(\epsilon-4\delta^2\gamma^2H^2L^2\right)} \\
=& \frac{2b}{H\gamma\left(\epsilon-CH^2\right)}+\frac{\varrho\beta}{B_0K\left(\epsilon-CH^2\right)(\beta-1)} \qquad (7)\\
=& O(\frac{1}{H^3})+O(\frac{\beta}{H^2K(\beta-1)}), \qquad (8)
\end{aligned}$$

where we define $\varrho=2L\gamma\sigma^2$, and $b=f\left(\overline{\mathbf{w}}^{0}\right)-f^*$. We have $C=4\gamma^2\delta^2L^2$, and we further choose $\epsilon>CH^2$.

Therefore, the communication round M can be represented as $\frac{A\beta}{\beta-1}+\chi$, where both A and χ are positive constants. Moreover, A and χ can be determined with $\frac{1}{H^2K}$ and $\frac{1}{H^3}$, respectively.