Supplement to Time-Varying Noise Perturbation and Power Control for Differential-Privacy-Preserving Wireless Federated Learning

APPENDIX A

PROOF OF LEMMA 1

The privacy loss of the Algorithm 1 is defined by $\exp(\psi(\gamma))$ [16], where $\psi(\gamma)$ is the γth moment, which is the logarithm of the moment generating function. Applying the decomposability theorem [16, Theorem 2], the γth moment of Algorithm 1 is upper bounded by

$$\psi(\gamma) \leq \sum_{t=1}^{T} \psi_t(\gamma) \leq \sum_{t=1}^{T} \frac{K}{N} \frac{\gamma(\gamma+1)\Delta_f^2}{2(\sigma_t^2+\xi_t^2)}, \tag{15}$$

where $\psi_t(\gamma)$ denotes the γth moment at the tth FL iteration of Algorithm 1 and Δ_f is defined in (4). For any $\epsilon > 0$, applying the tail bound of the moment [16, Theorem 2] gives

$$\widetilde{\delta} = \min_{\lambda} \exp\Big(\psi(\gamma) - \gamma\epsilon\Big) \tag{16a}$$

$$= \min_{\lambda} \exp\Big(\sum_{t=1}^{T} \frac{K}{N} \frac{\gamma(\gamma+1)\Delta_f^2}{2(\sigma_t^2+\xi_t^2)} - \gamma\epsilon\Big). \tag{16b}$$

Defining $p(\gamma) = \sum_{t=1}^{T} \frac{K\Delta_f^2}{2N(\sigma_t^2+\xi^2)}\gamma(\gamma+1) - \gamma\epsilon$, the problem (16b) is equivalent to $\gamma^\star = \operatorname{argmin}_\gamma q(\gamma)$. It is not difficult to verify that the latter problem is convex. Therefore, the optimal solution $\gamma^\star$ can be obtained by equating the first-order derivative to 0, i.e., $p'(\gamma^\star) = 0$. Specifically, $\sum_{t=1}^{T} \frac{K\Delta_f^2}{2N(\sigma_t^2+\xi^2)}(2\gamma^\star+1) - \epsilon = 0$, leading to

$$\gamma^\star = \omega\epsilon - \frac{1}{2}, \tag{17}$$

where the last equality follows from (4) and the definition of ω in Lemma 1. Substituting (17) into $p(\gamma)$ gives $p(\gamma^\star) = -\frac{(2\omega\epsilon+1)^2}{8\omega}$, implying that $\tilde{\delta} = \exp\left(-\frac{(2\omega\epsilon+1)^2}{8\omega}\right)$. Therefore, the proposed Algorithm 1 ensures (ϵ, δ)-DP for all $\delta \geq \tilde{\delta}$, which completes the proof.

APPENDIX B

PROOF OF LEMMA 3

We note that the term $\mathbb{E}[\|\widehat{\mathbf{w}}_{t+1} - \mathbf{w}^\star\|^2]$ is independent of the DP noise $\mathbf{z}_t$ and wireless channel noise $\mathbf{n}_t$. Hence, applying the conventional result in FL in [18, Lemma 2] to $\mathbb{E}[\|\widehat{\mathbf{w}}_{t+1} - \mathbf{w}^\star\|^2]$ gives

$$\mathbb{E}[\|\widehat{\mathbf{w}}_{t+1} - \mathbf{w}^\star\|^2] \leq (1+N\eta_t^2)(1-\mu\eta_t)^E \mathbb{E}[\|\mathbf{w}_t - \mathbf{w}^\star\|^2] + E(E-1)^2 L^2 \frac{\sigma_f^2}{N} e\eta_t^2 + \frac{E^2\sigma^2\eta_t^2}{2} + E^2(E-1)L^2\sigma^2 e\eta_t^4. \tag{18}$$

This completes the proof.

Appendix C
Proof of Lemma 4

We have

$$\mathbb{E}[\|\widehat{\mathbf{w}}_{t+1}-\overline{\mathbf{w}}_{t+1}\|^2] = \mathbb{E}\Big[\Big\|\mathbf{w}_t+\frac{1}{N}\sum_{n=1}^{N}\mathbf{u}_{n,t}-\frac{1}{N}\sum_{n=1}^{N}\mathbf{w}_{n,t}^{(E)}\Big\|^2\Big], \tag{19a}$$

$$= \mathbb{E}\Big[\Big\|\frac{1}{N}\sum_{n=1}^{N}\mathbf{u}_{n,t}-\frac{1}{N}\sum_{n=1}^{N}(\mathbf{w}_{n,t}^{(E)}-\mathbf{w}_t)\Big\|^2\Big], \tag{19b}$$

$$= \mathbb{E}\Big[\Big\|\frac{1}{N}\sum_{n=1}^{N}\Big(c_{k,t}(\mathbf{w}_t^{(E)}-\mathbf{w}_t)-(\mathbf{w}_{n,t}^{(E)}-\mathbf{w}_t)\Big)\Big\|^2\Big], \tag{19c}$$

$$\leq \frac{1}{N}\sum_{n=1}^{N}\mathbb{E}\Big[\Big\|c_{k,t}(\mathbf{w}_t^{(E)}-\mathbf{w}_t)-(\mathbf{w}_{n,t}^{(E)}-\mathbf{w}_t)\Big\|^2\Big], \tag{19d}$$

where the last inequality is due to the Cauchy-Schwarz inequality. Applying Lemma 2 for $p=2$ gives

$$\mathbb{E}\Big[\Big\|c_{k,t}(\mathbf{w}_t^{(E)}-\mathbf{w}_t)-(\mathbf{w}_{n,t}^{(E)}-\mathbf{w}_t)\Big\|^2\Big] \leq \frac{\mathbb{E}[\|\mathbf{w}_{n,t}^{(E)}-\mathbf{w}_t\|^2]}{C}. \tag{20}$$

Plugging (20) into (19d) yields

$$\mathbb{E}[\|\widehat{\mathbf{w}}_{t+1}-\overline{\mathbf{w}}_{t+1}\|^2] \leq \frac{1}{NC}\sum_{n=1}^{N}\mathbb{E}[\|\mathbf{w}_{n,t}^{(E)}-\mathbf{w}_t\|^2]. \tag{21}$$

From the standard result of local SGD in [18, Eqn. (59)], we have

$$\mathbb{E}\Big[\Big\|\mathbf{w}_{n,t}^{(E)}-\mathbf{w}_t\Big\|^2\Big] \leq 2E^2L^2\eta_t^2\mathbb{E}[\|\mathbf{w}_t-\mathbf{w}^\star\|^2]+2E^2\sigma_f^2\eta_t^2+2(E-1)E^2L^2\sigma_f^2e\eta_t^4, \forall n. \tag{22}$$

Substituting (22) into (21), we have the following,

$$\mathbb{E}[\|\widehat{\mathbf{w}}_{t+1}-\overline{\mathbf{w}}_{t+1}\|^2] \leq \frac{1}{C}\Big(2E^2L^2\eta_t^2\mathbb{E}[\|\mathbf{w}_t-\mathbf{w}^\star\|^2]+2E^2\sigma_f^2\eta_t^2+2(E-1)E^2L^2\sigma_f^2e\eta_t^4\Big). \tag{23}$$

This completes the proof.

APPENDIX D
PROOF OF LEMMA 5

We let $\widehat{\mathbf{g}}_{t+1} = \frac{1}{N}\sum_{n=1}^{N}\mathbf{u}_{n,t}$. Then, the following holds

$$\begin{aligned}
\mathbb{E}[\|\mathbf{w}_{t+1} - \widehat{\mathbf{w}}_{t+1}\|^2] &= \mathbb{E}_{\mathcal{K}}\Big[\|\frac{1}{K}\sum_{k\in\mathcal{K}}(\sqrt{\rho}\mathbf{u}_{k,t} + \mathbf{z}_t + \mathbf{n}_t) - \widehat{\mathbf{g}}_{t+1}\|\Big], \\
&= \mathbb{E}_{\mathcal{K}}\Big[\|\frac{1}{K}\sum_{k\in\mathcal{K}}\sqrt{\rho}\mathbf{u}_{k,t} - \widehat{\mathbf{g}}_{t+1}\|^2\Big] + \frac{\sigma_t^2 + \xi_t^2}{K}, && \text{(24a)} \\
&\leq \mathbb{E}_{\mathcal{K}}\Big[\|\frac{1}{K}\sum_{k\in\mathcal{K}}(\rho+1)(\|\mathbf{u}_{k,t}\|^2 + \|\widehat{\mathbf{g}}_{t+1}\|^2)\Big] + \frac{\sigma_t^2 + \xi_t^2}{K}, && \text{(24b)} \\
&= (\rho+1)\mathbb{E}_{\mathcal{K}}\Big[\|\frac{1}{K}\sum_{k\in\mathcal{K}}\|\mathbf{u}_{k,t}\|^2\Big] + \frac{\rho+1}{K}\|\widehat{\mathbf{g}}_{t+1}\|^2 + \frac{\sigma_t^2 + \xi_t^2}{K}, \\
&= (\rho+1)\frac{1}{N}\sum_{n=1}^{N}\|\mathbf{u}_{n,t}\|^2 + \frac{\rho+1}{K}\|\widehat{\mathbf{g}}_{t+1}\|^2 + \frac{\sigma_t^2 + \xi_t^2}{K}, \\
&\leq (\rho+1)\frac{1}{N}\sum_{n=1}^{N}\|\mathbf{u}_{n,t}\|^2 + \frac{\rho+1}{KN}\sum_{n=1}^{N}\|\mathbf{u}_{n,t}\|^2 + \frac{\sigma_t^2 + \xi_t^2}{K}, && \text{(24c)} \\
&= \Big(\frac{\rho+1}{N} + \frac{\rho+1}{KN}\Big)\sum_{n=1}^{N}\|\mathbf{u}_{n,t}\|^2 + \frac{\sigma_t^2 + \xi_t^2}{K}, && \text{(24d)}
\end{aligned}$$

where (24a) is due to the fact that the DP noise $\mathbf{z}_t$ a communication noise $\mathbf{n}_t$ are independent to each other and they are independent of $\{\mathbf{u}_{k,t}\}$ and $\widehat{\mathbf{g}}_{t+1}$ and (24b) and (24c) are due to the Cauchy-Schwarz inequality.

Next, we further upper bound the r.h.s. of (24d). We have the following,

$$\begin{aligned}
\mathbb{E}[\|\mathbf{u}_{n,t}\|^2] &= \mathbb{E}[\|c_{k,t}(\mathbf{w}_{n,t}^{(E)} - \mathbf{w}_t)\|^2], && \text{(25a)} \\
&\leq 2\mathbb{E}\Big[\Big\|c_{k,t}(\mathbf{w}_{n,t}^{(E)} - \mathbf{w}_t) - (\mathbf{w}_{n,t}^{(E)} - \mathbf{w}_t)\Big\|^2\Big] + 2\mathbb{E}\Big[\Big\|\mathbf{w}_{n,t}^{(E)} - \mathbf{w}_t\Big\|^2\Big], && \text{(25b)} \\
&\leq \Big(\frac{2}{C} + 2\Big)\mathbb{E}[\|\mathbf{w}_{n,t}^{(E)} - \mathbf{w}_t\|^2], && \text{(25c)} \\
&\leq \Big(\frac{2}{C} + 2\Big)(2E^2L^2\eta_t^2\mathbb{E}[\|\mathbf{w}_t - \mathbf{w}^\star\|^2] + 2E^2\sigma_f^2\eta_t^2 + 2(E-1)E^2L^2\sigma_f^2 e\eta_t^4), && \text{(25d)}
\end{aligned}$$

where (25b) follows from Cauchy-Schwarz inequality, (25c) is due to Lemma 2 for $p = 2$, and (25d) follows from (22). From (25d) and (24d), we attain

$$\begin{aligned}
\mathbb{E}[\|\mathbf{w}_{t+1} - \widehat{\mathbf{w}}_{t+1}\|^2] &\leq \Big(\rho + 1 + \frac{\rho+1}{K}\Big)\Big(\frac{2}{C} + 2\Big)\Big(2E^2L^2\eta_t^2\mathbb{E}[\|\mathbf{w}_t - \mathbf{w}^\star\|^2] + 2E^2\sigma_f^2\eta_t^2 + 2(E-1)E^2L^2\sigma_f^2 e\eta_t^4\Big) + \frac{\sigma_t^2 + \xi_t^2}{K}, && \text{(26a)} \\
&= \alpha\Big(2E^2L^2\eta_t^2\mathbb{E}[\|\mathbf{w}_t - \mathbf{w}^\star\|^2] + 2E^2\sigma_f^2\eta_t^2 + 2(E-1)E^2L^2\sigma_f^2 e\eta_t^4\Big) + \frac{\sigma_t^2 + \xi_t^2}{K}, && \text{(26b)} \\
&\leq \alpha\Big(2E^2L^2\eta_t^2\mathbb{E}[\|\mathbf{w}_t - \mathbf{w}^\star\|^2] + 2E^2\sigma_f^2\eta_t^2 + 2(E-1)E^2L^2\sigma_f^2 e\eta_t^4\Big) + \max_t \frac{\rho}{\theta_t K}, && \text{(26c)}
\end{aligned}$$

where the last inequality follows from the constraint in (6d). This completes the proof.

Appendix E

Proof of Theorem 1

We have $\eta_t = \frac{4}{E\mu t} \le \eta_{t_0} \le \min\left\{\frac{1}{E\mu}, \frac{E\mu}{4N}, \frac{\mu}{4E((\frac{6}{C}+6\alpha)L^2+\mu^2)}\right\}$. Hence, the following holds,

$$
\begin{aligned}
m_{1,t} &= 3(1+N\eta_t^2)(1-\mu\eta_t)^E + (\frac{6}{C}+6\alpha)E^2L^2\eta_t^2, && (27a)\\
&= (1+N\eta_t^2)(1-\mu\eta_t)^2 3(1-\mu\eta_t)^{E-2} + (\frac{6}{C}+6\alpha)E^2L^2\eta_t^2, && (27b)\\
&\le (1+N\eta_t^2)(1-2E\mu\eta_t+E^2\mu^2\eta_t^2) + (\frac{6}{C}+6\alpha)E^2L^2\eta_t^2, && (27c)\\
&\le (1+N\eta_t^2)(1-E\mu\eta_t+E^2\mu^2\eta_t^2) + (\frac{6}{C}+6\alpha)E^2L^2\eta_t^2, && (27d)\\
&= 1-E\mu\eta_t+E^2\mu^2\eta_t^2 + N\eta_t^2(1-E\mu\eta_t+E^2\mu^2\eta_t^2) + (\frac{6}{C}+6\alpha)E^2L^2\eta_t^2, && (27e)\\
&\le 1-E\mu\eta_t+E^2\mu^2\eta_t^2 + N\eta_t^2 + (\frac{6}{C}+6\alpha)E^2L^2\eta_t^2, && (27f)\\
&\le 1-\frac{E\mu\eta_t}{2}, && (27g)\\
&= 1-\frac{2}{t}, && (27h)
\end{aligned}
$$

where (27c) follows from the fact that $3(1-\mu\eta_t)^{E-2} \le 1$ when E is large, (27f) is due to the fact that $(1-E\mu\eta_t+E^2\mu^2\eta_t^2) \le 1$, (27g) is due to the facts that $E^2\eta_t^2((\frac{6}{C}+6\alpha)L^2+\mu^2) \le \frac{E\mu\eta_t}{4}$ and $N\eta_t^2 \le \frac{E\mu\eta_t}{4}$, and (27h) follows from the fact that $\eta_t = \frac{4}{E\mu t}$. From (27g) and (12), we obtain

$$
\mathbb{E}[\|\mathbf{w}_{t+1}-\mathbf{w}^\star\|^2] \le \Big(1-\frac{2}{t}\Big)\mathbb{E}[\|\mathbf{w}_t-\mathbf{w}^\star\|^2] + m_{2,t}. \tag{28}
$$

Now, we show the inequality in (13) by the induction method. When $t=t_0$, it is straightforward to see that the result in (13) holds. Assume that (13) holds up to $t=s>t_0$, i.e., $\mathbb{E}[\|\mathbf{w}_s-\mathbf{w}^\star\|^2] \le \frac{t_0}{s}\mathbb{E}[\|\mathbf{w}_{t_0}-\mathbf{w}^\star\|^2] + sM_0 + \frac{M_1}{E^2\mu^2 s} + \frac{M_2}{E^4\mu^4 s^2}$. Applying (28) for $t=s+1$ leads to

$$
\begin{aligned}
\mathbb{E}[\|\mathbf{w}_{s+1}-\mathbf{w}^\star\|^2] &\le (1-\frac{2}{s})\mathbb{E}[\|\mathbf{w}_s-\mathbf{w}^\star\|^2] + m_{2,s}, && (29a)\\
&\le (1-\frac{2}{s})\Big(\frac{t_0}{s}\mathbb{E}[\|\mathbf{w}_{t_0}-\mathbf{w}^\star\|^2] + sM_0 + \frac{M_1}{E^2\mu^2 s} + \frac{M_2}{E^4\mu^4 s^2}\Big) + m_{2,s}, && (29b)\\
&\le \frac{t_0}{s+1}\mathbb{E}[\|\mathbf{w}_{t_0}-\mathbf{w}^\star\|^2] + (s+1)M_0 + \frac{M_1}{E^2\mu^2(s+1)} + \frac{M_2}{E^4\mu^4(s+1)^2}, && (29c)
\end{aligned}
$$

where (29a) is due to (28), (29b) follows from the induction hypothesis, and (29c) follows from the facts that $(1-\frac{2}{s})\frac{1}{s} = \frac{s-2}{s(s+1)} \le \frac{1}{s+1}$ and $(1-\frac{2}{s})\Big(sM_0 + \frac{M_1}{E^2\mu^2 s} + \frac{M_2}{E^4\mu^4 s^2}\Big) + m_{2,s} \le (s+1)M_0 + \frac{M_1}{E^2\mu^2(s+1)} + \frac{M_2}{E^4\mu^4(s+1)^2}$. This implies that the result in (13) also holds for $t=s+1$, which completes the proof.

References

[1] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Artificial intelligence and statistics*. PMLR, 2017, pp. 1273–1282.

[2] G. Zhu, Y. Du, D. Gündüz, and K. Huang, "One-bit over-the-air aggregation for communication-efficient federated edge learning: Design and convergence analysis," *IEEE Transactions on Wireless Communications*, vol. 20, no. 3, pp. 2120–2135, 2021.

[3] Y. Wang, Y. Xu, Q. Shi, and T.-H. Chang, "Quantized federated learning under transmission delay and outage constraints," *IEEE Journal on Selected Areas in Communications*, vol. 40, no. 1, pp. 323–341, 2022.

[4] T. Gafni, N. Shlezinger, K. Cohen, Y. C. Eldar, and H. V. Poor, "Federated learning: A signal processing perspective," *IEEE Signal Processing Magazine*, vol. 39, no. 3, pp. 14–41, 2022.

[5] L. Zhu, Z. Liu, and S. Han, "Deep leakage from gradients," in *Proceedings of the 33rd International Conference on Neural Information Processing Systems*. Red Hook, NY, USA: Curran Associates Inc., 2019.

[6] J. Geiping, H. Bauermeister, H. Dröge, and M. Moeller, "Inverting gradients - how easy is it to break privacy in federated learning?" in *Advances in Neural Information Processing Systems*, H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, and H. Lin, Eds., vol. 33, 2020, pp. 16 937–16 947.

[7] C. Dwork and A. Roth, "The algorithmic foundations of differential privacy," *Found. Trends Theor. Comput. Sci.*, vol. 9, no. 3–4, p. 211–407, aug 2014.

[8] N. Agarwal, A. T. Suresh, F. X. X. Yu, S. Kumar, and B. McMahan, "cpSGD: Communication-efficient and differentially-private distributed SGD," in *Advances in Neural Information Processing Systems*, S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, Eds., vol. 31. Curran Associates, Inc., 2018.

[9] K. Wei, J. Li, M. Ding, C. Ma, H. H. Yang, F. Farokhi, S. Jin, T. Q. S. Quek, and H. Vincent Poor, "Federated learning with differential privacy: Algorithms and performance analysis," *IEEE Transactions on Information Forensics and Security*, vol. 15, pp. 3454–3469, 2020.

[10] X. Zhang, X. Chen, M. Hong, S. Wu, and J. Yi, "Understanding clipping for federated learning: Convergence and client-level differential privacy," ser. Proceedings of Machine Learning Research, vol. 162. Proceedings of the 39th International Conference on Machine Learning, 17–23 Jul 2022, pp. 26 048–26 067.

[11] W.-N. Chen, P. Kairouz, and A. Özgür, "Breaking the communication-privacy-accuracy trilemma," *IEEE Transactions on Information Theory*, vol. 69, no. 2, pp. 1261–1281, 2023.

[12] X. Yuan, W. Ni, M. Ding, K. Wei, J. Li, and H. V. Poor, "Amplitude-varying perturbation for balancing privacy and utility in federated learning," *IEEE Transactions on Information Forensics and Security*, vol. 18, pp. 1884–1897, 2023.

[13] Y. Koda, K. Yamamoto, T. Nishio, and M. Morikura, "Differentially private aircomp federated learning with power adaptation harnessing receiver noise," in *GLOBECOM 2020 - 2020 IEEE Global Communications Conference*, 2020, pp. 1–6.

[14] D. Liu and O. Simeone, "Privacy for free: Wireless federated learning via uncoded transmission with adaptive power control," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 1, pp. 170–185, 2021.

[15] M. S. E. Mohamed, W.-T. Chang, and R. Tandon, "Privacy amplification for federated learning via user sampling and wireless aggregation," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 12, pp. 3821–3835, 2021.

[16] M. Abadi, A. Chu, I. Goodfellow, H. B. McMahan, I. Mironov, K. Talwar, and L. Zhang, "Deep learning with differential privacy," in *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, ser. CCS '16. Association for Computing Machinery, 2016, p. 308–318.

[17] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.

[18] A. Reisizadeh, A. Mokhtari, H. Hassani, A. Jadbabaie, and R. Pedarsani, "Fedpaq: A communication-efficient federated learning method with periodic averaging and quantization," in *International Conference on Artificial Intelligence and Statistics*, 2019. [Online]. Available: https://api.semanticscholar.org/CorpusID:203593931

[19] Y. Nesterov, *Introductory Lectures on Convex Optimization: A Basic Course*, 1st ed. Springer Publishing Company, Incorporated, 2014.

[20] R. Das, S. Kale, Z. Xu, T. Zhang, and S. Sanghavi, "Beyond uniform lipschitz condition in differentially private optimization," in *Proceedings of the 40th International Conference on Machine Learning*, ser. ICML'23. JMLR.org, 2023.

[21] D. Q. Nguyen and T. Kim, "Supplement to Time-Varying Noise Perturbation and Power Control for Differential-Privacy-Preserving Wireless Federated Learning," 2023. [Online]. Available: https://github.com/quandku/Supplementary-for-Asilomar-2023

[22] Y. LeCun and C. Cortes, "MNIST handwritten digit database," 2010. [Online]. Available: http://yann.lecun.com/exdb/mnist/