Defending Byzantine Attacks in Ensemble Federated Learning: A Reputation-based Phishing Approach

Beibei Li, *Member, IEEE,* Peiran Wang, *Student Member, IEEE,* Qinglei Kong, *Member, IEEE,* Yuan Zhang, *Member, IEEE,* and Rongxing Lu, *Fellow, IEEE*

***Abstract*—Emerging as a promising distributed learning paradigm, federated learning (FL) has been widely adopted in many fields. Nonetheless, a big challenge for FL in real-world implementation is Byzantine attacks, where compromised clients can mislead or poison the training model by falsifying or manipulating the local model parameters. To solve this problem, in this paper, we present a reputation-based Byzantine robust-FL scheme (called FLPhish) for defending Byzantine attacks under the Ensemble Federated Learning architecture (called EFL). Specifically, we first develop a novel ensemble FL architecture, EFL, which allows FL compatible with different deep learning models in different clients. Second, we craft a phishing algorithm for the EFL architecture to identify possible Byzantine behaviors. Third, a Bayesian inference based reputation mechanism is devised to measure each client's level of confidence and to further identify Byzantine clients. Last, we strictly analyze how the FLPhish scheme defend against backdoor attacks. Extensive experiments under different settings demonstrate that the proposed FLPhish achieves great efficacy in defending Byzantine attacks in EFL. FLPhish is tested with different fractions of Byzantine clients and different degrees of distribution imbalance. [1]**

***Index Terms*—Federated learning, ensemble learning, Bayesian inference-based reputation, phishing.**

I. Conclusion

The conclusion goes here.

Acknowledgments

This should be a simple paragraph before the References to thank those individuals and institutions who have supported your work on this article.

This paper is an extended version of the paper titled 'FLPhish: Reputation-Based Phishing Byzantine Defense in Ensemble Federated Learning', which was published in IEEE ISCC 2021, and awarded 'Best Paper'.

B. Li and P. Wang are with the School of Cyber Science and Engineering, Sichuan Universsity, Chengdu, Sichuan, China 610065. Email: libeibei@scu.edu.cn; wangpeiran@stu.scu.edu.cn.

Q. Kong is with the Future Network of Intelligence Institute, The Chinese University of Hong Kong, Shenzhen, China 518172, and also with The University of Science and Technology of China, Hefei, China 230052. Email: kql8904@163.com.

Y. Zhang is with the School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu, China 610054. Email: zy_loye@126.com.

R. Lu is with the Faculty of Computer Science, University of New Brunswick, Fredericton, NB, Canada E3B 5A3. Email: rlu1@unb.ca.

TABLE I
Summary of Notations

Term	Description
s	central server in FL
c_i	the ith client in FL, $i = 1, 2, 3, ..., u$
d_i	the local dataset preserved by the ith client
C	the ensemble of all the clients
u	the number of clients
D_t	the unlabeled dataset chosen by s in each procedure
D	the unlabeled dataset preserved by s
n	the number of samples in D_t
B_t	the labeled dataset (*'bait'*) chosen by s in each procedure
B	the labeled dataset preserved by s
m	the number of samples in B_t
a_i^t	the accuracy of predictions of B_t made by c_i in tth procedure
q_i	the label of c_i to judge it is a malicious client or not
r_q	the threhold of malicious clients
x_l^t	the lth data point in D_t
b_i	the Byzantine attacker
σ	the 'trigger' in the backdoor attack
ι	the backdoor label in the backdoor attack
$\mathbf{M}$	global model preserved by s
$\mathbf{m_i}$	local models trained by the ith client
$\mathbf{k_i^t}$	the predictions (*'knowledge'*) made by the ith client in the tth procedure
$\mathbf{\hat{y}_l^t}$	the ensembled prediction of data point x_l^t
$\mathbf{\hat{y}_l^i}$	the prediction of lth data point made by ith client
$\mathbf{K_t}$	the aggregated labels (predictions) of the tth iteration's unlabeled dataset

Appendix
Proof of the Zonklar Equations

Use `\appendix` if you have a single appendix: Do not use `\section` anymore after `\appendix`, only `\section*`. If you have multiple appendixes use `\appendices` then use `\section` to start each appendix. You must declare a `\section` before using any `\subsection` or using `\label` (`\appendices` by itself starts a section numbered zero.)

References Section

You can use a bibliography generated by BibTeX as a .bbl file. BibTeX documentation can be easily obtained at: http://mirror.ctan.org/biblio/bibtex/contrib/doc/

0000–0000/00$00.00 © 2021 IEEE

The IEEEtran BibTeX style support page is: http://www.michaelshell.org/tex/ieeetran/bibtex/

Simple References

You can manually copy in the resultant .bbl file and set second argument of `\begin` to the number of references (used to reserve space for the reference number labels box).

References

[1] K. A. Bonawitz, H. Eichner, W. Grieskamp, D. Huba, A. Ingerman, V. Ivanov, C. M. Kiddon, J. Konečný, S. Mazzocchi, B. McMahan, T. V. Overveldt, D. Petrou, D. Ramage, and J. Roselander, "Towards federated learning at scale: System design," in *2019 the 2nd Systems and Machine Learning Conference (SysML)*, Stanford, CA, USA, Mar. 31-Apr. 2 2019.

References

[1] *Mathematics Into Type*. American Mathematical Society. [Online]. Available: https://www.ams.org/arc/styleguide/mit-2.pdf
[2] T. W. Chaundy, P. R. Barrett and C. Batey, *The Printing of Mathematics*. London, U.K., Oxford Univ. Press, 1954.
[3] F. Mittelbach and M. Goossens, *The LaTeXCompanion*, 2nd ed. Boston, MA, USA: Pearson, 2004.
[4] G. Grätzer, *More Math Into LaTeX*, New York, NY, USA: Springer, 2007.
[5] M. Letourneau and J. W. Sharp, *AMS-StyleGuide-online.pdf,* American Mathematical Society, Providence, RI, USA, [Online]. Available: http://www.ams.org/arc/styleguide/index.html
[6] H. Sira-Ramirez, "On the sliding mode control of nonlinear systems," *Syst. Control Lett.*, vol. 19, pp. 303–312, 1992.
[7] A. Levant, "Exact differentiation of signals with unbounded higher derivatives," in *Proc. 45th IEEE Conf. Decis. Control*, San Diego, CA, USA, 2006, pp. 5585–5590. DOI: 10.1109/CDC.2006.377165.
[8] M. Fliess, C. Join, and H. Sira-Ramirez, "Non-linear estimation is easy," *Int. J. Model., Ident. Control*, vol. 4, no. 1, pp. 12–27, 2008.
[9] R. Ortega, A. Astolfi, G. Bastin, and H. Rodriguez, "Stabilization of food-chain systems using a port-controlled Hamiltonian description," in *Proc. Amer. Control Conf.*, Chicago, IL, USA, 2000, pp. 2245–2249.

Biography Section

If you have an EPS/PDF photo (graphicx package needed), extra braces are needed around the contents of the optional argument to biography to prevent the LaTeX parser from getting confused when it sees the complicated `\includegraphics` command within an optional argument. (You can create your own custom macro containing the `\includegraphics` command to make things simpler here.)

If you include a photo:

Michael Shell Use `\begin{IEEEbiography}` and then for the 1st argument use `\includegraphics` to declare and link the author photo. Use the author name as the 3rd argument followed by the biography text.

If you will not include a photo:

John Doe Use `\begin{IEEEbiographynophoto}` and the author name as the argument followed by the biography text.