Efficient Distributed Learning Over Lossy Wireless Networks

Emilio Paolini[1,2,3], Andrea Pinto[4], Luca Valcarenghi[1], Nicola Andriolli[5], Luca Maggiani[3], and Flavio Esposito[4]
[1]*Tecip Institute, Scuola Superiore Sant'Anna, Pisa, Italy*
[2]*CNR-IEIIT, Pisa*, [3]*Sma-RTy, Italy*
[4]*Computer Science Department, Saint Louis University, St. Louis, USA*
[5]*Department of Information Engineering (DII), University of Pisa, Pisa, Italy*
emilio.paolini@santannapisa.it

***Abstract*—In the context of NextG Wireless Networks, addressing the challenges of wireless communication link reliability is paramount to ensure efficient Distributed Learning systems. However, many recent solutions have overlooked key challenges, such as packet-level losses and the impact of TCP retransmissions, which are crucial for the robustness of these systems. In this paper, we propose the integration of fountain codes into the distributed learning process to offer a robust mechanism to counteract packet loss. Specifically, we propose a cumulative strategy logic based on fountain codes specifically tailored for packet exchanges in Distributed Learning applications. Our evaluation shows that fountain codes significantly enhance the efficiency and reliability of distributed learning model updates under severe packet loss conditions, e.g., a packet reduction of $\approx 84\%$ ($\approx 60\%$) at the UE (gNB) side compared to traditional TCP methods when packet loss probability reaches 0.9 in Federated Learning context. However, under low packet loss scenarios, fountain codes computational overhead becomes non-negligible. These results highlight the potential of fountain codes to serve as a robust alternative to conventional communication protocols in distributed learning systems, particularly in environments characterized by unstable network conditions.**

***Index Terms*—fountain codes, federated learning, wireless networks, distributed learning**

I. Introduction

NextG wireless communications promise enhanced services, including better coverage and higher data rates [1]. These wireless networks will heavily rely on distributed Artificial Intelligence (AI) applications to continuously learn from massive datasets [2]. Distributed AI approaches provide lower privacy risks, bandwidth demands, and latency compared to systems that require transferring raw data to the cloud for centralized processing [3], [4]. Among the available Distributed Learning (DL) approaches, Federated Learning (FL) [5] allows devices to collaboratively learn a shared Machine Learning (ML) task while keeping the data local by exchanging only model updates, offering the potential to design, extensively scale, and fully automate context-aware AI solutions across a broad spectrum of 6G applications. Efforts are focused on the integration of a various range of applications and scenarios, necessitating robust and reliable data exchange channels [6].

However, wireless networks surrounding User Equipments (UEs) can face abrupt and sequential losses due to fluctuations in wireless channels. As discussed in [6], a possible solution may be the exploitation of heterogeneous wireless channels to provide a stable and reliable transmission channel. Although interesting, this approach may be unfeasible due to burst consecutive losses [7].

A commonly adopted strategy to counteract the unreliability of wireless channels involves the deployment of robust transport layer protocol, i.e., Transmission Control Protocol (TCP). However, this method of ensuring transmission reliability typically results in significant communication overhead, as it often requires messages to be sent multiple times [8]. Compared to TCP, User Datagram Protocol (UDP) does not ensure the reliable delivery of every packet. Nonetheless, in situations where packet loss significantly influences network performance, UDP can surpass TCP in efficiency, particularly for transmitting model updates in FL [8]. However, the lack of reliability of UDP can be problematic for DL systems, where model updates are critical to speed up model convergence.

Considering these limitations, a possible solution would be to exploit model compression techniques to reduce the number of packets to be sent over a TCP connection. However, such approaches can be resource-intensive and impractical for devices with limited processing capabilities, e.g., Internet of Things (IoT) devices [9].

Hence, given the previously mentioned challenges, it becomes critical to rethink the communication protocol for the efficient transmission of model updates in the context of NextG wireless environments. Originally designed for the purpose of multimedia streaming, Fountain Codes (FC) offer a robust method for ensuring the reliable transfer of information across devices in networks characterized by unstable connectivity [10].

978-3-903176-66-9

Network coding, particularly FC, has already proven effective in enhancing reliable packet transmission in dynamic 6G networks [11], [12]. Unlike traditional TCP-like solutions, which are optimal for point-to-point communications and rely on complex error correction and congestion control mechanisms, FC can reduce the computational burden in the communication process by reducing the number of packets transmitted. By leveraging the collective acknowledgment strategy inherent to FC, network coding offers a way to significantly improve the reliability and efficiency of model update transmissions in DL systems, thereby enabling this technology to reach its full potential in wireless environments. Thus, in this work, we have conducted a comprehensive evaluation of the application of FC in an FL wireless distributed AI application. A convergence analysis is also discussed, showing how FC helps FL algorithms converge under packet loss scenarios. Additionally, a cumulative ACK strategy logic based on FC specifically tailored for FL scenarios, is proposed. The performance of this strategy has been extensively evaluated in a digit classification problem considering a static channel loss scenario.

II. State Of The Art

A. Related Work

Recently, research efforts have been focused on many aspects related to the integration of FL techniques in the context of wireless network infrastructures, yet often overlooking key challenges such as packet-level losses and the impact of TCP retransmission. The work in [13] proposes an FL algorithm in the context of traffic estimation to maximize users data rates. In [14], the authors consider the challenges of straggling devices and imperfect Channel State Information (CSI) in wireless federated computation, developing a new approach to minimize the computing and transmission delay. While these works suggested the seamless integration of FL algorithms within wireless network infrastructures, practical challenges, such as the unreliability of wireless channels, including symbol errors, and resource constraints, e.g., bandwidth and power, significantly hinder their efficacy [15], degrading the quality and accuracy of FL updates.

A recent work [16] addresses the challenge of training FL algorithms over real-world wireless networks that experience packet losses. The authors introduce a novel algorithm that adapts the FedAvg method to work effectively over asymmetric and lossy communication channels, updating the global model by a pseudo-gradient step. The study focuses on scenarios where a client either successfully sends its update or fails to do so, not addressing packet-level losses or examining the implications of TCP retransmissions.

Hence, considering the limitations of previous works, we have carried out a comprehensive evaluation of packet-level losses for FL in unreliable wireless communication networks and proposed a FC-based approach that can solve existing mechanism issues.

A preliminary version of this research was presented in [17]. The focus was on demonstrating the feasibility of using FC in a simplified testbed environment. This work significantly extends the original demonstration by providing a comprehensive evaluation and scalability analysis, including extensive simulations that quantify the improvements brought by FC in the FL process and a detailed demonstration of model convergence.

B. Federated Learning

Federated Learning (FL) is an approach to training machine learning models across multiple devices, e.g., User Equipments (UEs), while keeping the data localized. This addresses privacy, security, and data ownership concerns that arise in centralized training methods. The fundamental idea behind FL is to enable devices to collaboratively learn a shared prediction model while keeping all the training data on the device, thus avoiding the need to send sensitive data to a central server.

An FL system can be described as a set of N devices, each with its own local dataset. Let $\mathcal{D}_k$ denote the local dataset of the k^{th} device, where $k = 1, 2, ..., N$. The goal is to train a global model $\mathbf{w}$ on the union of the datasets $\bigcup_{i=1}^{N} \mathcal{D}_k$ without exchanging data samples. The objective can be formulated as the following optimization problem:

$$\min_{\mathbf{w}} F(\mathbf{w}) = \sum_{k=1}^{N} \frac{n_k}{n} F_k(\mathbf{w}) \tag{1}$$

where $F(\mathbf{w})$ is the global objective function, $F_k(\mathbf{w})$ is the local objective function of the j^{th} device computed using its local dataset $\mathcal{D}_k$, n_k is the number of samples in $\mathcal{D}_k$, and $n = \sum_{i=1}^{N} n_k$ is the total number of samples across all devices.

Each local objective function $F_k(\mathbf{w})$ typically takes the form of an empirical risk minimization (ERM) problem:

$$F_k(\mathbf{w}) = \frac{1}{n_k} \sum_{j \in \mathcal{D}_k} L(y_j, f(\mathbf{x}_j; \mathbf{w})) \tag{2}$$

where L is a loss function measuring the discrepancy between the true label y_j and the prediction $f(\mathbf{x}_j; \mathbf{w})$ made by the model with parameters $\mathbf{w}$ for the j^{th} sample in $\mathcal{D}_k$.

C. Fountain Codes

FC are a class of erasure codes that have revolutionized the way data is transmitted over unreliable channels. These codes are characterized by their unique "rateless" property, allowing an infinite number of encoded symbols to be generated from a finite set of original data symbols [18]. The mathematical foundation and the practical applications of FCs have made them a subject of extensive research and implementation in various fields.

At the core of FC is the concept of generating encoded symbols that can be produced on-the-fly and in an unlimited manner. Let $\mathbf{d} = \{d_1, d_2, \ldots, d_k\}$ represent the original data symbols, where k is the number of symbols. The FC encoder transforms these symbols into encoded symbols $\mathbf{e} = \{e_1, e_2, \ldots, e_n\}$, with $n \geq k$. The key mathematical principle here is that any subset of the encoded symbols of size k or slightly more is highly likely to be sufficient for decoding back to the original k symbols.

One of the enhancements in this domain is the introduction of Raptor codes, which build on the FC0 framework by incorporating an initial pre-coding stage. This process mathematically transforms the original data symbols $\mathbf{d}$ into intermediate symbols $\mathbf{i}$, which are then encoded to produce the symbol set $\mathbf{e}$ based on a certain probability distribution. The encoding and decoding operations are governed by linear algebraic procedures, facilitating a linear time complexity in terms of the number of symbols, denoted as $O(n)$.

Fountain Codes have found applications in a wide range of domains due to their robustness and efficiency. They are particularly beneficial in scenarios where the channel conditions are unpredictable and where it is unfeasible to determine the exact number of transmission errors in advance. Some of the notable applications include reliable data storage and retrieval systems, ensuring data integrity and availability [19].

III. System Design

In this section, we describe the proposed system model, highlighting how FC can be implemented in the DL process. The proposed systems for FL is shown in Fig. 1.

Following the converged Radio Access Network (RAN) and Core Network (CN) architecture [20], we place the FL Parameter Server (PS) at the next generation eNB (gNB). Both the gNB and the UEs exploit a FC-based scheme built on top of UDP connections to transmit the parameters of the FL process. Specifically, each device involved in the FL will exploit a FC-based encoder to produce repair symbols that will be transmitted over the unreliable wireless communication channel.

A key strength of FC used in this work is their resilience to packet loss and bandwidth variability, ensuring that all devices in a wireless network receive the necessary model updates, even under fluctuating network conditions, common in NextG wireless settings. This scenario is depicted in Fig. 2, where even if some packets, i.e., source symbols or repair symbols, are lost during transmission, the receiver can still reconstruct the original model updates K as long as it receives enough encoded packets.

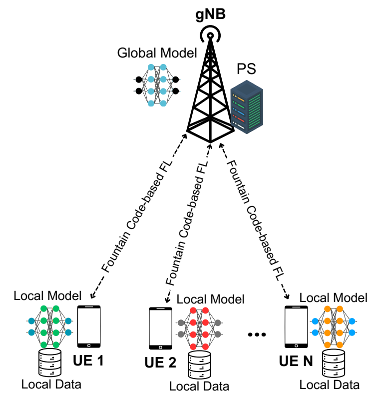


Fig. 1: The Wireless Federated Learning architecture features a Parameter Server (PS) at the gNB and clients at the User Equipment (UE), which use the proposed Fountain Codes to enhance communication with the base station.

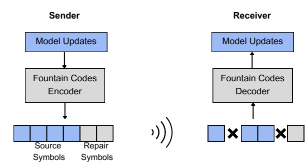


Fig. 2: Example of how the receiver can reconstruct the model updates under packet loss.

Another aspect to consider is the size of individual packets that can be sent across the network. Indeed, this is limited by the Maximum Transmission Unit (MTU) size, and this reflects on the number of packets sent in the DL process. Given the size constraints imposed by the MTU in wireless networks, it is impractical to encapsulate an entire model update within a single packet.

Furthermore, in the context of unreliable communication channels, the transmitted packets can grow exponentially due to retransmissions, resulting in fewer available resources for the transmission of new data.

Hence, to counteract the effects of unreliable packet delivery, we propose a cumulative ACK strategy based on FC. The receiver handles incoming data packets over a network by using a socket to receive and decode the data. The logic includes a mechanism for sending a ’STOP’ signal to halt further transmissions once the data is successfully decoded. It also accounts for the possibility that the ’STOP’ command may be lost through a timeout-based check for retransmissions to ensure reliable communication. The sender transmits packets sequentially through a socket to the receiver's address and pauses to wait for a 'STOP' acknowledgment (ACK) signal, indicating that the receiver has successfully decoded the data. If the 'STOP' ACK is not received within a brief timeout, the sender retransmits the remaining packets, ensuring data reliability and completeness at the receiver's end.

IV. Evaluation

A. Parameter Server

We consider a static channel loss scenario for the experiments. The experiments are based on rateless codes, i.e., RaptorQ, since they are particularly suited for packet loss recovery for the erasure channel [21]. We systematically varied the probability of packet loss across a broad spectrum, ranging from 10% to 90%, in increments of 10%. This approach allowed us to simulate a wide array of network conditions, from relatively stable to highly unreliable, thereby enabling a comprehensive evaluation of the algorithm's robustness and adaptability to adverse transmission environments.

The experiments consider a Convolutional Neural Network (CNN) with 3 convolutional layers, consisting of 10, 20, and 40 filters, respectively, each followed by a max pooling layer. Additionally, 2 fully-connected layers with 50 and 10 neurons are utilized. The objective of the ML algorithm is to perform digit classification on the benchmark MNIST dataset [22]. Results are reported in Sec. IV-A2.

1) Convergence Analysis: For the convergence analysis of the proposed algorithms, we assume the following.

Assumption 1. $\{F_k\}_{k \in N}$ are *L-smooth*:

$$\forall v, w \quad F_k(v) \leq F_k(w) + \langle \nabla F_k(w), v - w \rangle + \frac{L}{2}\|v - w\|^2$$

Assumption 2. $\{F_k\}_{k \in N}$ are *μ-strongly convex*:

$$\forall v, w \quad F_k(v) \geq F_k(w) + \langle \nabla F_k(w), v - w \rangle + \frac{\mu}{2}\|v - w\|^2$$

Assumption 3. Let ξ_k be a random batch sampled from k-th device's local data uniformly at random. The variance of the stochastic gradients in each device is bounded:

$$\mathbb{E}[\|\nabla F_k(w, \xi_k) - \nabla F_k(w)\|^2] \leq \sigma_k^2$$

Assumption 4. Each UE experiences packet loss characterized by probability p_k. Each client updates the model based on its local data:

$$w_{t+1,k} = w_(t, k) - \eta_t \nabla F_k(w_{t,l})$$

The server aggregates the received updates using:

$$w_{t+1} = w_t + \sum_{k=1}^{N} \alpha_k \Delta_k = w_t - \eta_t \sum_{k=1}^{N} \alpha_k (1 - p_k) \nabla F_k(w_t)$$

Hence, the expected update from client k is:

$$\mathbb{E}[\Delta_k] = (1 - p_k)\nabla F_k(w_t)$$

and the expected global update is:

$$\mathbb{E}[w_{t+1}] = w_t - \eta_t \sum_{k=1}^{N} \alpha_k (1 - p_k) \mathbb{E}[\nabla F_k(w_t)]$$

Since each F_k is L-smooth, also F is L-smooth and thus:

$$F(v) \leq F(w) + \langle \nabla F(w), v - w \rangle + \frac{L}{2}\|v - w\|^2$$

setting $v = w_{t+1}$ and $w = w_t$ we have:

$$F(w_{t+1}) \leq F(w_t) + \langle \nabla F(w_t), w_{t+1} - w_t \rangle + \frac{L}{2}\|w_{t+1} - w_t\|^2$$

using the update rule:

$$F(w_{t+1}) \leq F(w_t) - \eta_t \langle \nabla F(w_t), \sum_{k=1}^{N} \alpha_k (1 - p_k) \nabla F_k(w_t) \rangle + \frac{L}{2} \left\| \eta_t \sum_{k=1}^{N} \alpha_k (1 - p_k) \nabla F_k(w_t) \right\|^2 \tag{3}$$

Since gradient contributions are independent across clients, we can rewrite the last term as:

$$\eta_t^2 \sum_{k=1}^{N} \alpha_k^2 (1 - p_k)^2 \|\nabla F_k(w_t)\|^2$$

Applying the expectation to all terms we get:

$$\mathbb{E}[F(w_{t+1})] \leq F(w_t) - \eta_t (1 - \overline{p}) \|\nabla F(w_t)\|^2 + \frac{L\eta_t^2 \sigma^2}{2}$$

To ensure convergence, we select n_t so that it decreases over time and satisfies the typical conditions for stochastic gradient descent: (i) $sum_{t=1}^{\infty} \eta_t = \infty$ and (ii) $sum_{t=1}^{\infty} \eta_t^2 \leq \infty$. The application of FC reduces packet loss probability by allowing the decoding of complete updates from a subset of received packets. This can be modeled as $\overline{p}_{\mathrm{FC}} < \overline{p}$. Additionally, they add computational complexity and in some cases noise due to approximation in decoding, e.g., interpolation technique. This is modeled by σ_{FC}, whose value may be slightly higher but with less variance.

2) Static Channel Loss: In the static channel experiment, we first analyzed the number of packets sent by both sides of the FL system, namely the gNB and UE. We examined the packets sent per round, defining a round as the process where packets are sent from one side to the other, encompassing a single direction of communication. Results are depicted in Fig. 3 and Fig. 4.

In both cases, the FC approach outperforms the traditional TCP connection. At both the gNB and UE, the disparity in the number of packets sent between the two methods widens as the packet loss probability increases. The highest reduction in sent packets is achieved at the

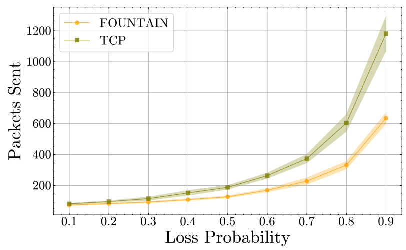


Fig. 3: Sent packets per round: gNB side.

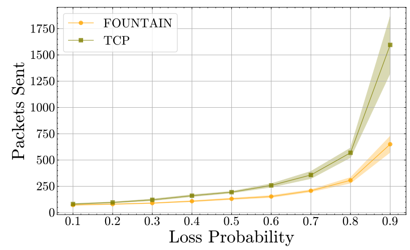


Fig. 4: Sent packets per round: UE side.

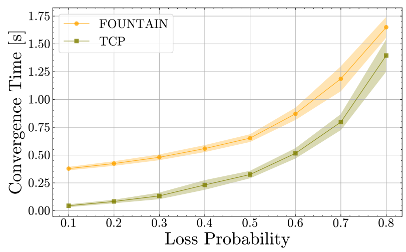


Fig. 5: Convergence time: gNB side.

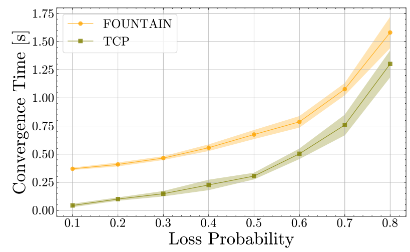


Fig. 6: Convergence time: UE side.

highest packet loss probability, $p = 0.9$, with the UE seeing a reduction of approximately 84% and the gNB seeing a reduction of approximately 60%. Conversely, the lowest reduction in sent packets occurs at $p = 0.1$ for both sides, with the UE achieving a reduction of about 12% and the gNB achieving a reduction of about 10%. These results highlight how FC can significantly reduce the number of packets exchanged in a FL process, especially under highly unreliable communication links.

The convergence time for the FL communication round results are shown in Fig. 5 and Fig. 6 for the gNB and the UE, respectively. An opposite behavior is observed compared to the previous analysis. Indeed, here the TCP-based solution outperforms the FC-based communication. This is because, even by exploiting FC with linear complexity, the overhead related to the decoding of the received packets impacts the final computation time. The highest differences in convergence time can be observed for the lowest loss probabilities. Indeed, in both UE and gNB, the increase in time due to FC in $p = 0.1$ is $\approx 157\%$. However, when the loss probability reaches very high values, i.e., in very unreliable channels, the overhead due to FC is smaller, reaching $\approx 19\%$ and $\approx 16\%$ for UE and gNB, respectively, when $p = 0.9$.

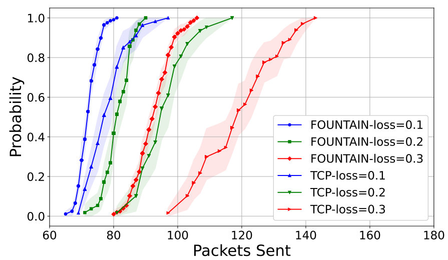
The increased computational overhead associated with decoding under FC, which significantly impacts convergence times, especially at low loss probabilities, suggests a trade-off. This overhead becomes less pronounced at very high loss probabilities, indicating that FC may be more advantageous in severely unreliable networks.

Furthermore, we have evaluated the Cumulative Distribution Function (CDF) for the considered loss probabilities. Results for the UE and gNB sides are reported in Fig. 7. On the UE side, the results again demonstrate the superior efficiency of FC in successfully transmitting data with fewer packets, especially as packet loss increases. Although both protocols require a higher number of packets as packet loss increases, FC consistently outperforms TCP-based communication, with more pronounced benefits in highly unreliable network conditions. Furthermore, the confidence intervals indicate that FC behavior is also more predictable, affirming its potential as a robust alternative to TCP, provided the computational overhead is manageable.

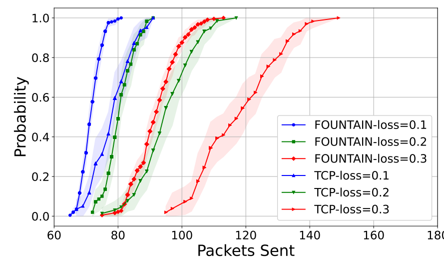
In the results depicting gNB-side performance, the FC-based protocol outperforms TCP at all observed packet loss probabilities. FC requires fewer packets to achieve a successful round of FL, even as packet loss severity increases. This performance trend suggests that FC are notably more efficient and reliable for FL systems, particularly in unstable network conditions, making it a preferred protocol in scenarios where robust communication is crucial.

Finally, Fig. 8 illustrates the scalability of the proposed approach by comparing the difference in packets sent between TCP-based and FC-based methods as the number of clients increases. The analysis considers scenarios with $N \in 1, 2, 4, 8, 16$, where N represents the number of clients participating in the FL process.

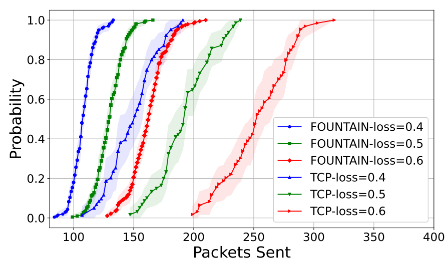
Regarding the scalability results for packets sent, reported in Fig. 8a, we observe stable behavior as the number of clients increases across all loss probability configurations. As already discussed, higher loss probabilities tend to have a higher reduction in packet sent.



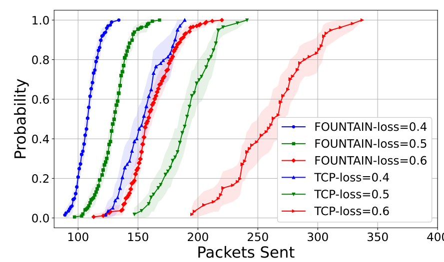
(a) UE side - CDF for $p \in \{0.1, 0.2, 0.3\}$



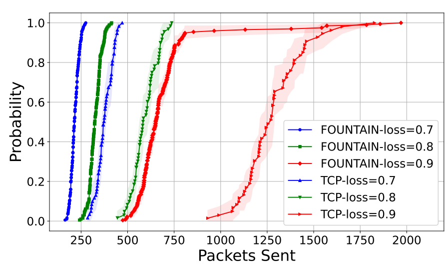
(b) gNB side - CDF for $p \in \{0.1, 0.2, 0.3\}$



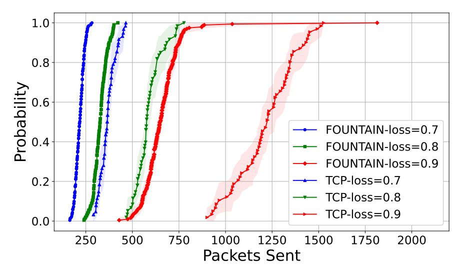
(c) UE side - CDF for $p \in \{0.4, 0.5, 0.6\}$



(d) gNB side - CDF for $p \in \{0.4, 0.5, 0.6\}$

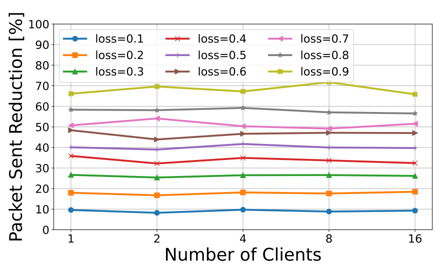


(e) UE side - CDF for $p \in \{0.7, 0.8, 0.9\}$

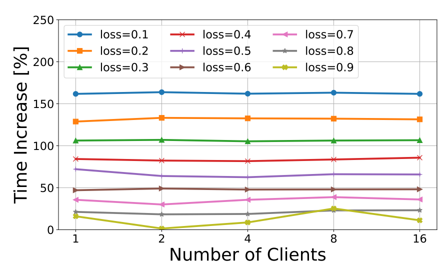


(f) gNB side - CDF for $p \in \{0.7, 0.8, 0.9\}$

Fig. 7: CDF for the considered loss probabilities.



(a) Packet Sent Reduction.



(b) Time increase.

Fig. 8: Scalability results for varying number of clients.

Indeed, the highest packet sent reduction is reached for $p = 0.9$ with 8 clients, corresponding to $\approx 72\%$.

Scalability results for the time needed for FL round, reported in Fig. 8b, confirm once again that FC can be exploited in scenarios where the packet loss is high. Hence, it is possible to observe that the time increase is relatively small for all clients when p is very high. Specifically, with 2 clients and $p = 0.9$ the lowest time increase is obtained, reaching $\approx 1\%$. Both scalability graphs confirm the previous observations made regarding the trade-off of applying FC in the context of FL. Indeed, all results point out that in high loss scenarios their exploitation can be beneficial to the FL process. However, when the loss is very low the overhead added by FC computations can become a bottleneck, evidencing that classical TCP solutions perform better.

V. Conclusion

In the context of NextG wireless networks, a critical challenge is the unreliability and variability of communication channels, which can severely hamper the efficiency and effectiveness of learning tasks. In this paper, we have focused on the problem of how packet loss affects the DL process in FL scenario, posing significant issues to achieving timely and accurate model updates across distributed networks.

To address these challenges, we propose the integration of FC within the DL systems as a novel solution, introducing a cumulative ACK logic. Results have shown scenarios in which the utility of FC over traditional TCP connections becomes evident. While FC demonstrate superior performance in scenarios characterized by high packet loss probabilities, e.g., sent packets reduction $\approx 84\%$ at the UE and $\approx 60\%$ at the gNB at a loss probability of $p = 0.9$ for FL, they may not always offer the best solution. Specifically, in environments where packet loss is minimal, i.e., $p = 0.1$, the benefits of FC diminish significantly and their overhead becomes non negligible, indicating that in such scenarios TCP is still the best option.

Therefore, while FC present a promising alternative to TCP for DL in high-loss wireless networks, careful consideration of their application is necessary, taking into account both the potential for packet transmission efficiency and the computational demands they impose, even in very optimized implementations. In conclusion, this paper paves the way for the application of FC in DL under network losses, and it sets the stage for further exploration into optimizing FC parameters through federated reinforcement learning, enabling adaptive responses to dynamic channel conditions.

VI. Acknowledgment

This work is supported by the CLEVER project (project number 101097560), funded by the Key Digital Technologies Joint Undertaking and its members, including top-up funding by the Italian MUR. It is also supported by the European Union under Italy's National Recovery and Resilience Plan (NRRP) of NextGenerationEU, ("Telecommunications of the Future" partnership PE00000001 - program "RESTART"), and by the Italian MUR under the FoReLab Project (Departments of Excellence). This work has been partially supported by NSF Awards #2133407 and #2201536.

References

[1] S. Dang, O. Amin, B. Shihada, and M.-S. Alouini, "What should 6G be?" *Nature Electronics*, vol. 3, no. 1, pp. 20–29, 2020.

[2] S. Talwar, N. Himayat, H. Nikopour, F. Xue, G. Wu, and V. Ilderem, "6G: Connectivity in the era of distributed intelligence," *IEEE Communications Magazine*, vol. 59, no. 11, pp. 45–50, 2021.

[3] E. Baccour, N. Mhaisen, A. A. Abdellatif, A. Erbad, A. Mohamed, M. Hamdi, and M. Guizani, "Pervasive AI for IoT applications: A survey on resource-efficient distributed artificial intelligence," *IEEE Communications Surveys & Tutorials*, 2022.

[4] P. S. Bouzinis, P. D. Diamantoulakis, and G. K. Karagiannidis, "Wireless Federated Learning (WFL) for 6G Networks Part I: Research Challenges and Future Trends," *IEEE Commun. Lett.*, vol. 26, no. 1, pp. 3–7, 2021.

[5] K. Bonawitz, H. Eichner, W. Grieskamp, D. Huba, A. Ingerman, V. Ivanov, C. Kiddon, J. Konečnỳ, S. Mazzocchi, B. McMahan *et al.*, "Towards federated learning at scale: System design," *Proceedings of machine learning and systems*, vol. 1, pp. 374–388, 2019.

[6] Y. Zhang, W. Zhao, P. Dong, X. Du, W. Qiao, and M. Guizani, "Improve the reliability of 6G vehicular communication through skip network coding," *Veh. Commun.*, vol. 33, p. 100400, 2022.

[7] G. Sun, Y. Zhang, D. Liao, H. Yu, X. Du, and M. Guizani, "Bus-trajectory-based street-centric routing for message delivery in urban vehicular ad hoc networks," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 8, pp. 7550–7563, 2018.

[8] H. Ye, L. Liang, and G. Y. Li, "Decentralized federated learning with unreliable communications," *IEEE J. Sel. Topics Signal Process.*, vol. 16, no. 3, pp. 487–500, 2022.

[9] H. Kim, M. U. K. Khan, and C.-M. Kyung, "Efficient neural network compression," in *Proc. IEEE/CVF CVPR*, 2019.

[10] S. Chanayai and A. Apavatjrut, "Fountain codes and their applications: Comparison and implementation for wireless applications," *Wireless Personal Communications*, vol. 121, no. 3, pp. 1979–1994, 2021.

[11] Y. Zhang, P. Dong, Y. Yu, X. Du, H. Luo, T. Zheng, and M. Guizani, "A bignum network coding scheme for multipath transmission in vehicular networks," in *Proc. IEEE GLOBECOM*, 2018, pp. 206–212.

[12] C. Yu, W. Quan, K. Liu, M. Liu, Z. Xu, and H. Zhang, "DRL-based fountain codes for concurrent multipath transfer in 6G networks," in *Proc. IEEE INFOCOM Workshops*, 2022.

[13] O. Habachi, M.-A. Adjif, and J.-P. Cances, "Fast uplink grant for NOMA: A federated learning based approach," in *Ubiquitous Networking: 5th International Symposium, UNet 2019, Limoges, France, November 20–22, 2019, Revised Selected Papers 5*. Springer, 2020, pp. 96–109.

[14] S. Ha, J. Zhang, O. Simeone, and J. Kang, "Coded federated computing in wireless networks with straggling devices and imperfect CSI," in *2019 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2019, pp. 2649–2653.

[15] J. Park, S. Samarakoon, M. Bennis, and M. Debbah, "Wireless network intelligence at the edge," *Proceedings of the IEEE*, vol. 107, no. 11, pp. 2204–2239, 2019.

[16] A. Rodio, G. Neglia, F. Busacca, S. Mangione, S. Palazzo, F. Restuccia, and I. Tinnirello, "Federated Learning with Packet Losses," in *2023 26th International Symposium on Wireless Personal Multimedia Communications (WPMC)*. IEEE, 2023, pp. 1–6.

[17] E. Paolini, L. Valcarenghi, N. Andriolli, L. Maggiani, and F. Esposito, "Enabling Lightweight Federated Learning in NextG Wireless Networks," in *2024 IEEE 10th International Conference on Network Softwarization (NetSoft)*. IEEE, 2024, pp. 304–306.

[18] D. J. MacKay, "Fountain Codes," *IEE Proceedings-Communications*, vol. 152, no. 6, pp. 1062–1068, 2005.

[19] S. S. Borkotoky and M. B. Pursley, "Analytical Techniques for Performance Evaluation of Fountain-Coded File Distribution in Packet Radio Networks," *IEEE/ACM Transactions on Networking*, 2023.

[20] E. Paolini, L. Valcarenghi, L. Maggiani, and N. Andriolli, "Real-Time Network Packet Classification Exploiting Computer Vision Architectures," *IEEE Open Journal of the Communications Society*, vol. 5, pp. 1155–1166, 2024.

[21] A. Ali, K. S. Kwak, N. H. Tran, Z. Han, D. Niyato, F. Zeshan, M. T. Gul, and D. Y. Suh, "RaptorQ-based efficient multimedia transmission over cooperative cellular cognitive radio networks," *IEEE Trans. Veh. Technol.*, vol. 67, no. 8, pp. 7275–7289, 2018.

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