On the Performance of Federated Learning Network

Godwin Idoje(✉), Tasos Dagiuklas, and Muddesar Iqbal

Computer Science Department, London South Bank University, London SE1 0AA, UK
idojeg2@lsbu.ac.uk
https://www.suitelab.org

Abstract. Federated Learning is a decentralised network platform where the edge nodes train their local models and send their updated weights to the server. The server combines all the various local weights received and sends the aggregated model back to the edge nodes for further training, and this process continues until convergence is achieved. This study models the Federated Learning (FL) network. The Traffic speed (TS), Round trip time (RTT), and Bandwidth delay-product (BDP) parameters have been considered for modelling the Federated Learning network. Through experimentation, it can be inferred that the TS has a high impact and high correlation on the BDP within the network, and the RTT has a low impact on the BDP. The decentralised and classical machine learning models' predictions have been compared. It has been observed that the decentralised machine learning model's prediction outperforms the classical machine learning model's prediction. The link experiences low latency because only the updated weights are transmitted within the link and not the raw data.

Keywords: Federated Learning · Bandwidth delay product · aggregate model · Round Trip time · Traffic speed

1 Introduction

Federated Learning is a decentralised Machine Learning framework where the edge nodes train their local models and send their updated weights to the server. The server combines all the various local updated weights received and sends the aggregated model back to the edge nodes for further training. This process continues until convergence is achieved [1]. The authors in [2] used the Network traffic Federated Learning extreme Learning machine models to analyse local network traffic data. Their model achieved a higher accuracy when compared with their benchmark models. The authors in [3] have used stochastic models, optimisation models and differential equations to model the optimisation of the federated Learning network. The stochastic models have been used to resolve uncertainty and variability data issues, while the convergence and stability challenges have been resolved using the deterministic models. This study discusses

H. Gao et al. (Eds.): CollaborateCom 2023, LNICST 563, pp. 41–56, 2024.
https://doi.org/10.1007/978-3-031-54531-3_3

the modelling of the network of queues of the federated Learning platform while considering the Bandwidth delay product (BDP) as a performance metric.

The authors in [4] discuss that Federated learning is a method in machine learning where a model is trained across numerous distributed edge devices or servers, each containing its own set of local data samples, without disclosing these data. This approach safeguards data privacy by storing it locally on a server or edge device during the model training process. Given legal constraints, this technique holds particular significance for hospitals, as it enables collaborative machine learning model development without transferring all the training data, such as patient records, to a centralised location. According to the authors in [5], Federated learning emerges as a privacy-centric approach to machine learning, which finds valuable application in intelligent healthcare. It involves orchestrating multiple hospitals to conduct deep learning training collaboratively without exchanging data. This federated learning technique standardises the individual training procedures by globally averaging feature vectors. Throughout the federated training process, the transmission of model parameters is unnecessary, and local clients merely upload the average feature vectors of each class. Clients can opt for distinct local models based on their computational capacities.

The authors in [6] discuss that Federated learning enables deep learning algorithms to gain insights from a wide range of data present in various databases. This innovative approach allows deep learning models to be trained using local patient data from different medical centres, with only model parameters shared among the facilities. The authors in [7] discuss that Federated Learning is the evaluation of models in a decentralised platform where the communication cost is reduced because only the updated weights of the local models are sent to the server. It can be inferred from their research that the Long short-term memory models considered for the federated Learning network have been five times faster in achieving prediction values than the centralised network. The authors in [8] discuss that the FL model has been used to aggregate the updated weights from health institutions (electronic records and primary & secondary health centres). The FL evaluation of the dataset has provided data privacy and security for health institutions.

It can be inferred from [1–8] that the FL technique enables data training while the dataset is domiciled at the local edge device. This method provides privacy for the data and reduces the communication cost of the network. Figure 1 and Fig. 2 depict the architecture of the decentralised and centralised network, respectively. The decentralised network only sends the updated weights to the server for aggregation, unlike the centralised network, where the raw datasets are sent to the server in the cloud for analysis. The decentralised network provides data privacy and security since the server in the cloud does not see the raw dataset for evaluation.

Our Contribution. Our paper models the FL network as a network of queues within a federated learning network, considering the network's Bandwidth delay product (BDP), Traffic speed and round trip time (RTT) as a performance metric. The convergence of decentralised and the classical centralised models has been compared. The decentralised model converged better than the classical centralised model. It has been observed that comparing the decentralised and centralised networks, the Bandwidth delay product of the decentralised network has been able to match a higher proportion of the predicted BDP with the original BDP values. Section 2 discusses the related works. Section 3 discusses the methodology adopted in conducting this research, while Sect. 4 discusses the results obtained from the experimentation. Section 5 narrates the conclusion and the Area of further work.

2 Related Work

The authors in [9] discuss that the congestion window within a network link, where the bandwidth capacity and the round-trip time (RTT) are considered, requires a maximum transmission rate and minimum delay scenario for optimal operation within the network. It can be inferred that the higher the transmission rate, the lower the delay within the network. The authors in [10] discuss the communication cost of the FL network reduction by introducing the federated sparse compression (FSC) algorithm. It can be inferred from their research that better generalisation and prediction have been achieved by using CapsNet to train data in edge devices. The authors in [11] have investigated bottleneck queue level (BQL) performance in a high bandwidth-delay product link. It has been observed that varying the RTT within the network has a very minimal effect on the link performance. They further iterate that the packet loss experienced within the network does not affect the congestion control performance.

The authors in [12] used the content popularity prediction of privacy-preserving (CPPPP) scheme based on federated learning and Wasserstein generative adversarial network (WGAN) to improve the cache hit ratio and resolve the data leakage during model training in an FL platform. It can be inferred that the transmission time using the FL scheme for caching has been reduced. [13] discuss that using Federated averaging + CNN + MobileNet models for the classification of breast cancer images has improved the classification accuracy of breast cancer detection using the federated averaging +CNN model. It can be inferred that their model classification results outperformed other centralised network models. According to [14], FL has been used to analyse chest X-ray images. The federated averaging models have been able to classify the infected lungs and healthy lungs from the image datasets. It can be inferred that their proposed model has reduced the bias in the prediction models because it combines all the updated weights and features of the various edge node models.

The authors in [15] discuss the slow convergence of Mobile edge nodes using Federated Learning for heterogeneous nodes. Their research results outperformed the existing centralised network in resource usage, accuracy, and convergence. [16] discuss that they analysed an Augmented Intelligence of Things (AIoT) network using BDP for a heterogeneous platform, and they affirmed that their solution improved the network's transmission rate by a factor of 3.72. The download rate has been boosted by 3.94 fold. Their solution improves the data transmission rate by reducing the round-trip time's impact and boosting the congestion window optimisation when data loss is experienced. It can be inferred that when the solution reduces the impact of the round-trip time within the network, the bandwidth-delay product is invariably affected positively. According to the authors in [17], many cache misses are experienced when the network's latency increases. Bandwidth and BDP capacities increase, which invariably results in about a 24% drop in throughput-per-core. [18] discuss that within a centralised network, when the RTT varies, as soon as the RTT drops below half of the average RTT, the bottleneck bandwidth round-trip propagation time (BBR) experiences payload collapse because of the congestion within the network.

3 Modelling of the FL Network

The BDP has been the performance metric for the accuracy of the federated Learning Network. The modelling has been achieved by making some assumptions to obtain the performance metric. The modelling Assumptions are the following;

The Arrival of the local model from the edge nodes is independent of each other and discrete events.

The Arrival rate of the local models follows Poisson distribution.

The inter-arrival rate time is independent, and we assume the service rate is said to be exponentially distributed. The edge nodes of the FL network link experience delay. This delay and bandwidth are represented as d_i and b_i, respectively, for the edge node.

Let b_i represent the effective bandwidth of the i^{th} node.

Let d_i represent the effective delay of the i^{th} node. The effective bandwidth of the server node of the Federated Learning network is represented as BW_eff and the delay at the server node is represented as DL_eff

Let BW_eff represent the effective bandwidth at the server.

Let DL_eff represent the effective delay at the server.

$$\sum_{j=1}^{k} b_i^j = b_i | min_j(d_i^j) d_i \tag{1}$$

$$(B_i, D_i) = (\sum_j b_i^j min_j(d_i^j)) \tag{2}$$

Equation 1 represents the summation of the bandwidth for the edge nodes that experience some delay. Equation 2 depicts the aggregate effective bandwidth and effective delay at the edge nodes. $BW_{eff} = min(B_s, B_i)$

$DL_{eff} = (D_s, D_i)$

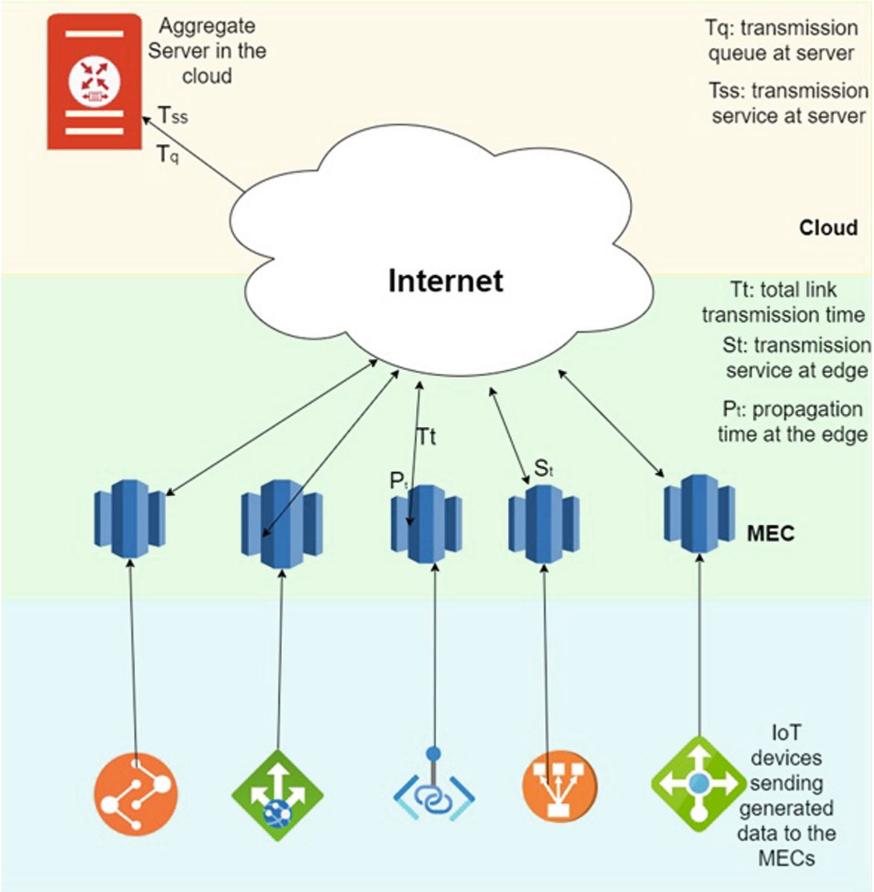

Fig. 1. Architecture of the Federated Learning Network

The Bandwidth Delay product can be calculated as

$$BW_{eff} \times DL_{eff} = \sum_{i=1}^{n} min\,(B_s\ ,\ B_i)\ \times\ D_s \sum_{i=1}^{n} D_i$$

$$= \left(\sum_{i=1}^{n} min \left(B_s\ ,\ \sum_{j=1}^{k} b_i^j \right) \times\ D_s \sum_{i=1}^{n} \min\left(d_i^j \right) \right) \tag{3}$$

Equation 3 shows the product of the effective bandwidth and delay at the server node and the aggregate bandwidth and delay at the edge nodes. It can be inferred that the product of the bandwidth, measured in Megabits per second (Mb/s) and the delay, measured in seconds (s) within the federated learning network, produces the total packets transmitted within the network.

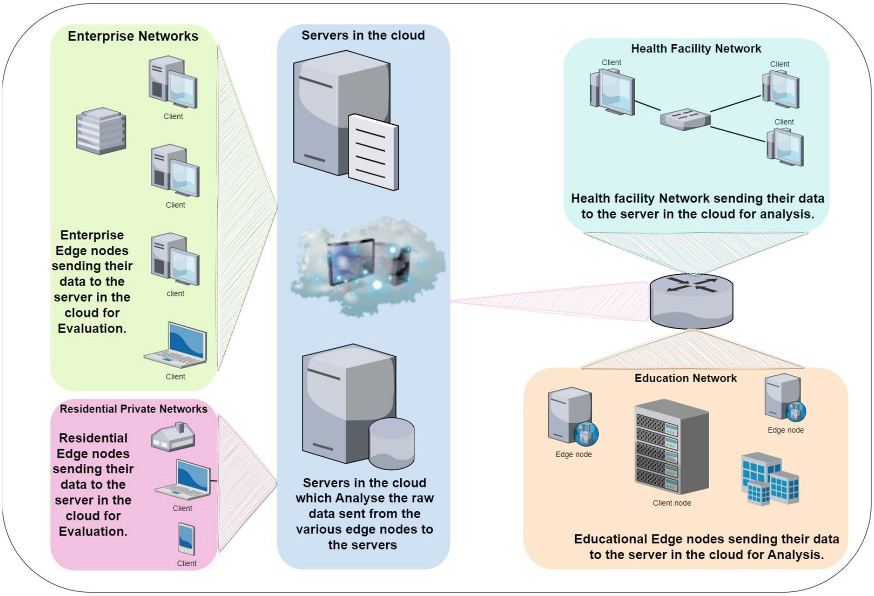

Fig. 2. Architecture of the Centralised Network

3.1 Methodology

The network was emulated using the GNS3 network tool, and the network architecture is shown in Fig. 1. The architecture has 3 tier layers. A Testbed has been set up, and the network's Traffic speed (TS) and round trip time (RTT) have been captured for analysis. The performance of the edge nodes, combined and classical machine learning models for the bandwidth-delay product has been investigated, and the predictions of the combined and classical models have been analysed. The generated data have been sent to the second tier layer containing the MEC nodes. The captured datasets are numerical. They were collected for a duration of Ninety days using the Paessler PRTG network monitor software installed on The database server within the network. The data has been pre-processed using the Python sci-kit library to remove Not A Number (NAN) values. These have been trained, and the local model updates have been sent to the server. The server aggregated all the various models from the respective edge nodes, creating a global model. The server sends the combined model to the edge nodes to train its local model with the new global model until it achieves convergence.

3.2 Time Complexity

The Big O notation is used to determine the time complexity of the Model mathematically.

The iteration

$$H = \frac{n}{2^H} \tag{4}$$

Taking the iteration to equal 100, which has been used for the training of the model during the training in the testbed,

$$n \div 2^H = 100$$

Therefore,

$$n = 100 * 2^H$$

taking the logarithm of both sides of the equation and note that 100 is a constant

$$logn = log2^H$$

$$log\ n = H \times log_2 2, \quad where \quad log_2 2 = 1$$

$$log\ n = H$$

from the model developed the Big O notation for the FedAve algorithm is $O(logn)$.

It can be inferred that the time complexity for the model is in order of log n, O(log n). The Federated averaging time complexity from Eq. 4 above evaluation is time efficient. It can be inferred that the model will be less space-efficient since time and space complexity are always inversely proportional.

4 Results and Discussion

Figure 3 shows the univariate distribution of the TS, RTT and BDP from the emulated network. It can be inferred that there is a positive correlation between TS and the BDP. The RTT shows a bi-modal distribution having two peak values, and the RTT is skewed to the right, implying the RTT mean values are more than the median values in the dataset. The bi-modal peaks shown in Fig. 3 result from the mixture of the samples from the experimentation. Another reason is the errors within the dataset obtained from the experimentation of the network. The BDP and TS both show they have a single mode, indicating that a single sample of the dataset population represents the mean, and the probability density for the TS and BDP are skewed to the left, indicating the mean for TS and BDP are less than the median of the dataset.

From Fig. 3, it can be deduced that the correlation between the BDP and RTT, BDP and TS experience a lot of noise and has a positive correlation. The chart shows many outliers in the BDP and RTT, unlike the BDP and TS, which have fewer outliers. This indicates that round-trip time has a low influence on the BDP. Taking the TS and RTT from Fig. 3 indicates a positive correlation but a low influence on the BDP due to the high outlier, as shown in Fig. 2. The

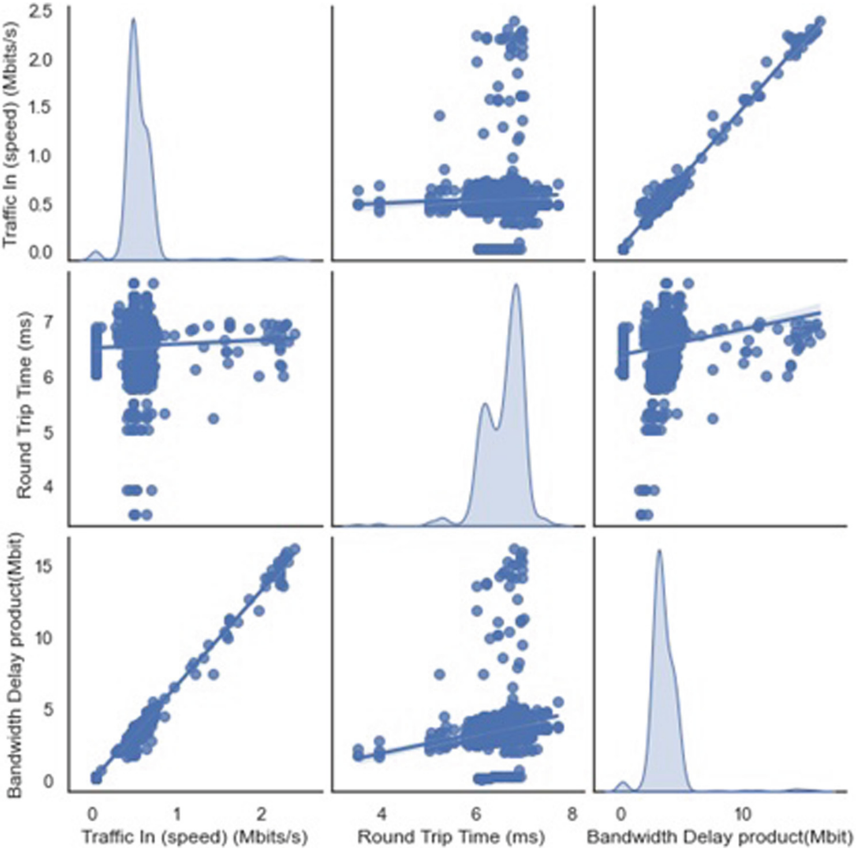

Fig. 3. Pairwise Plot of the Bandwidth Delay Product.

kernel density estimate indicates that the round-trip time is bi-modal while the traffic speed and BDP are unimodal. Figure 3 shows the pairwise plot to express the correlation between the variables, and it cannot be conclusive evidence of the relationship between the TS, RTT and BDP within the dataset. Further analysis is carried out on the BDP dataset using the federated and classical machine learning models. The positive correlation between the BDP and TS indicates a direct relationship. An increase in the TS will invariably increase the BDP. The outliers observed from the dataset captured from the experimentation are caused by the emulation tool used for the experiment. The tool inaccuracies and challenges caused the outliers.

To further substantiate our analysis of the BDP dataset, a heatmap has been developed, as shown in Fig. 4. The heatmap gives numerical values to the correlation between the variables in the BDP dataset. It can be inferred from Fig. 4 that the correlation value between TS and the RTT is 0.042, substantiating the findings in Fig. 3 that the correlation between the two variables is very weak. The correlation value between RTT and the BDP is 0.18, higher than that of

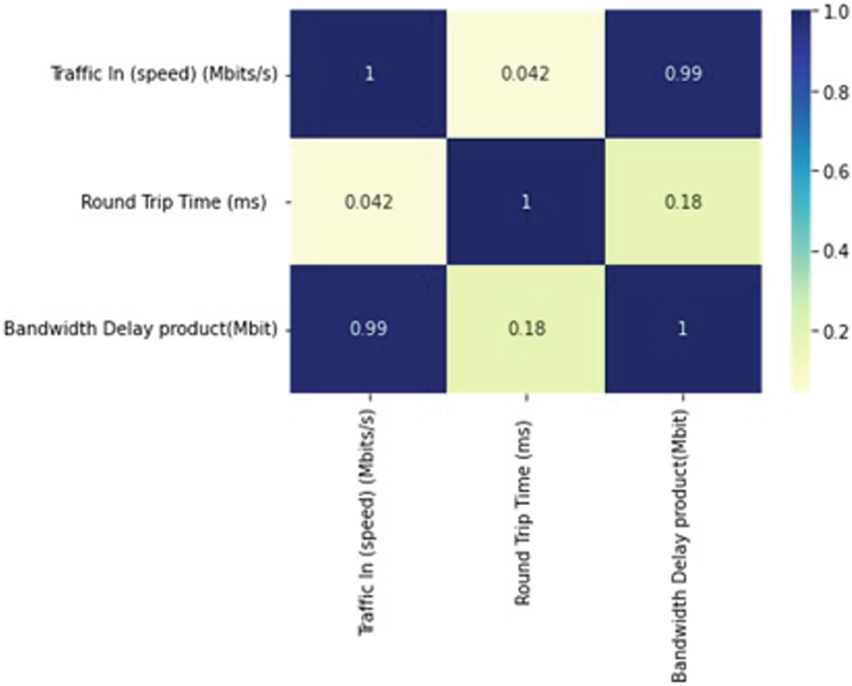

Fig. 4. Heat map correlation of the Bandwidth delay Product.

RTT and TS, but it also shows a weak correlation between the two variables. The correlation values between the BDP and TS have a correlation value of 0.99, which is extremely high, indicating the traffic speed within a federated learning network can influence the BDP values within the network. The diagonal values show the correlation of each to itself.

It can inferred from Figs. 5 to 28 that the convergence of each edge node varies for each node indicating the diversity of the traffic within the edge nodes. Figures 5 to 28 show the different epochs at which each edge node converges indicating the different patterns of the traffic within each edge node. The aggregate model predictions and the Original BDP values are shown in Fig. 29. It can be observed that a high portion of the predictions is very accurate with the original BDP value. This indicates that the aggregate model can combine the individual node models and produce an acceptable prediction of the BDP. This is shown in Fig. 29, where a plot of the predictions against the original BDP values is shown. It can be inferred that the Federated Learning model can predict a high proportion of the BDP original values. The research investigates the predictions of the classical machine learning model, as shown in Fig. 30, where a low proportion of the predicted BDP values accurately match the original BDP values. It can be observed from Figs. 29 and 30 that the aggregate model of the Federated Learning platform has been able to predict a high proportion of the original BDP values.

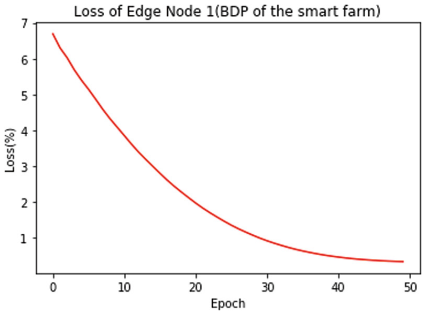

Fig. 5. Loss of edge Node 1

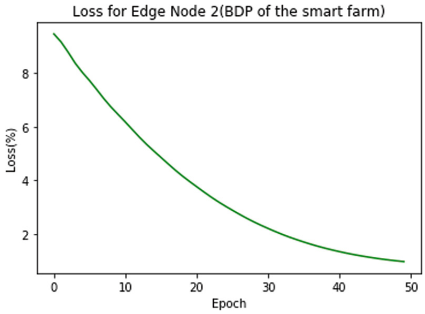

Fig. 6. Loss of edge node 2

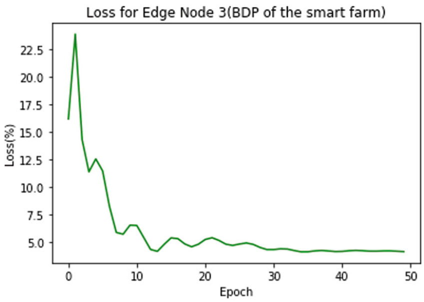

Fig. 7. Loss of edge Node 3

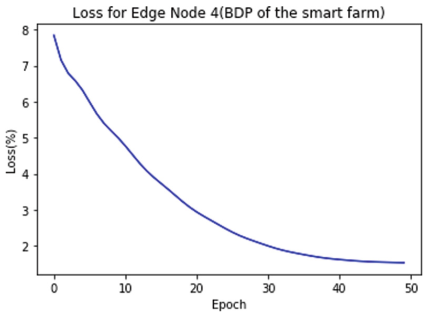

Fig. 8. Loss of edge node 4

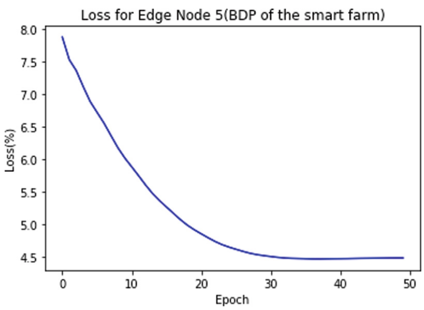

Fig. 9. Loss of edge Node 5

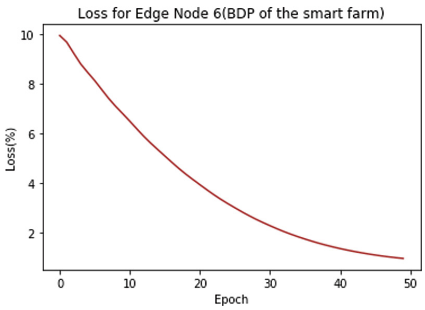

Fig. 10. Loss of edge node 6

Fig. 11. Loss of edge Node 7

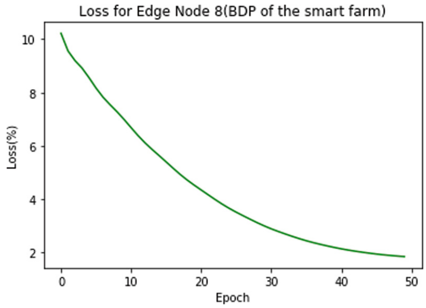

Fig. 12. Loss of edge node 8

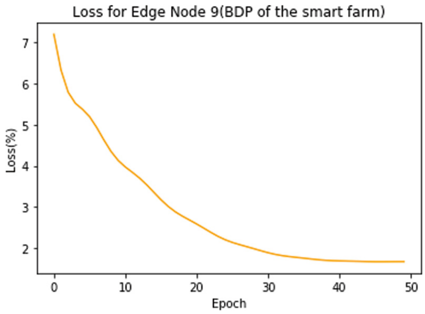

Fig. 13. Loss of edge Node 9

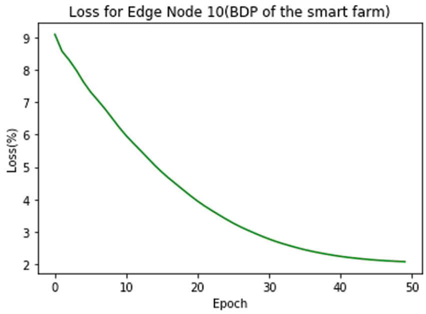

Fig. 14. Loss of edge node 10

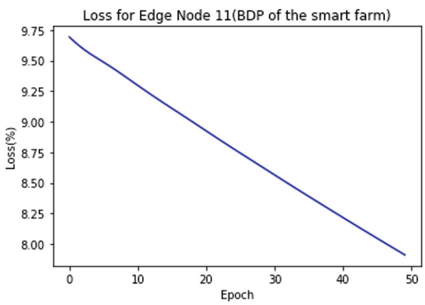

Fig. 15. Loss of edge Node 11

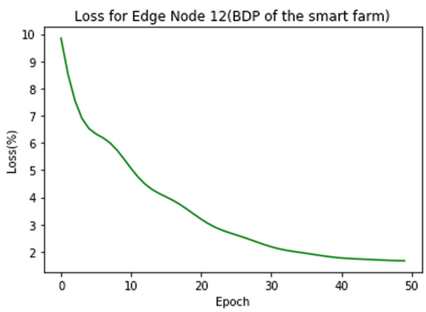

Fig. 16. Loss of edge node 12

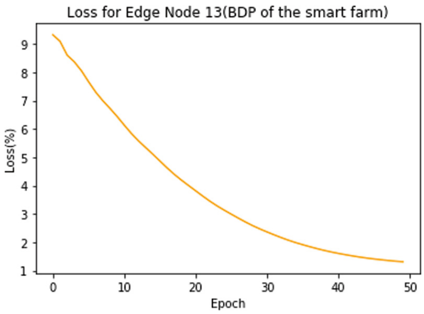

Fig. 17. Loss of edge Node 13

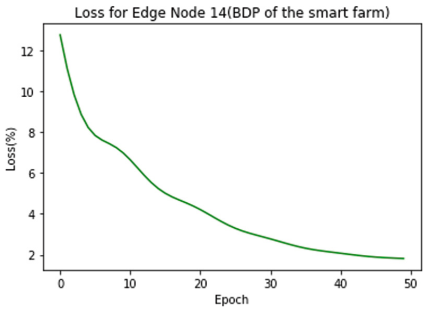

Fig. 18. Loss of edge node 14

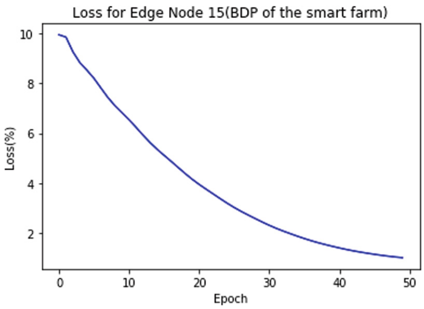

Fig. 19. Loss of edge Node 15

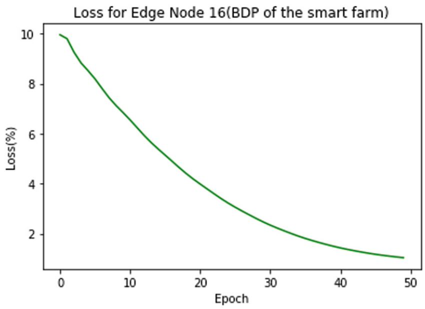

Fig. 20. Loss of edge node 16

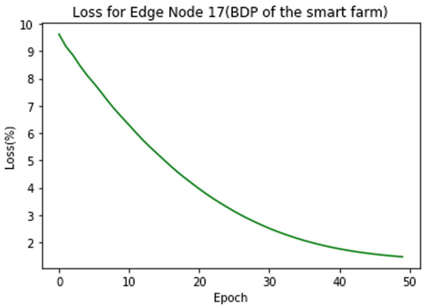

Fig. 21. Loss of edge Node 17

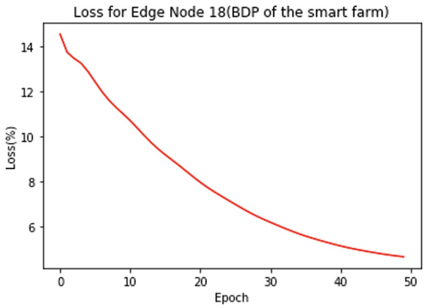

Fig. 22. Loss of edge node 18

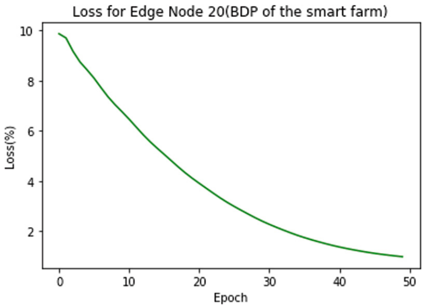

Fig. 23. Loss of edge Node 20

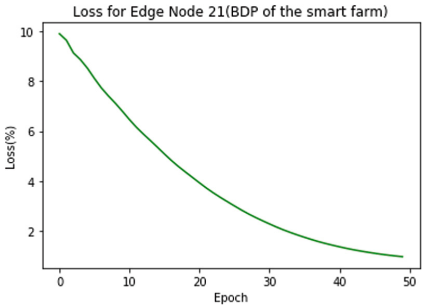

Fig. 24. Loss of edge node 21

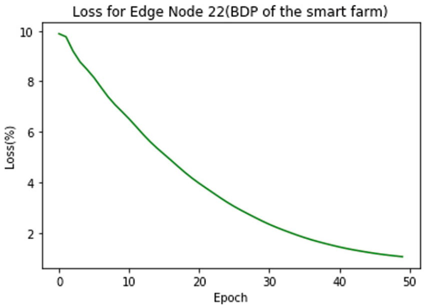

Fig. 25. Loss of edge Node 22

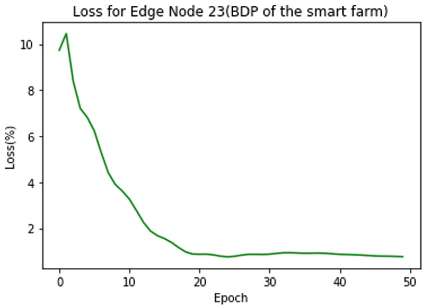

Fig. 26. Loss of edge node 23

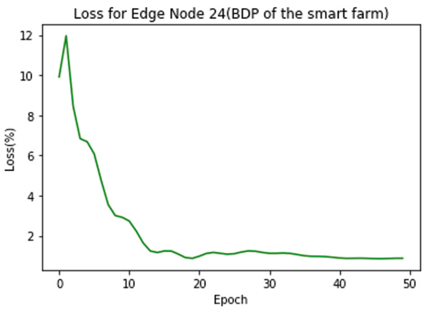

Fig. 27. Loss of edge Node 24

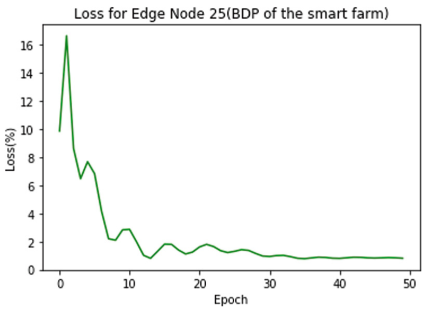

Fig. 28. Loss of edge node 25

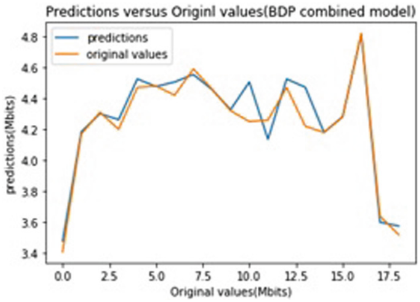

Fig. 29. Prediction versus Original Aggregate model of BDP network

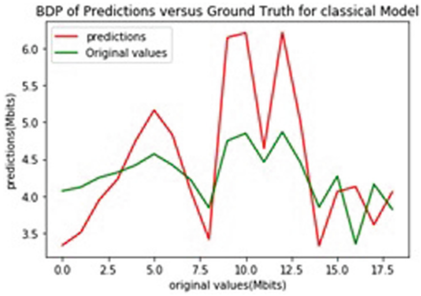

Fig. 30. Prediction versus Original Classical model of BDP network

The Aggregate model of the decentralised network outperforms the centralised machine learning model in the accuracy of the predictions produced, which can be seen in Figs. 29 and 30. The decentralised model outperformed the centralised model by producing a higher accuracy of the predicted BDP than the centralised model accuracy of the predicted BDP as shown in Figs. 29 and 30.

5 Conclusion

This research has investigated the modelling of the network of queues of a federated learning network. Mathematical modelling of the performance metrics, such as the bandwidth-delay product, has been used to depict the performance of the FL network within a federated learning network using GNS3 simulation for experimentation. It has been observed that the bandwidth-delay product is an important parameter that affects the convergence of the federated Learning network. The FL network produced a higher accuracy of the predictions when compared with the centralised machine learning model using the BDP as a performance metric. The predictions of the decentralised machine learning model outperformed the classical machine learning model. It can be inferred that the BDP has a high correlation with the TS, while the correlation of the BDP with the RTT is rather low. Further analysis of the Bandwidth Delay product can be investigated using parallel learning models and a comparison with the Federated split learning model and classical machine learning model.

References

1. Kim, S.W.: Covert communication over federated learning channel. In: 2023 17th International Conference on Ubiquitous Information Management and Communication (IMCOM). IEEE (2023). https://doi.org/10.1109/IMCOM56909.2023
2. Qiu, C.: A network traffic classification method based on federated learning and extreme learning machine. In: 2023 IEEE International Conference on Control, Electronics and Computer Technology (ICCECT). IEEE (2023). https://doi.org/10.1109/ICCECT57938.2023.10140851

3. Kumar, B., Singh, S., Grover, R., Isabels, K.R., Garg, A., Dattatraya, B.C.: Analysis of mathematical modelling deterministic and stochastic problems in federated learning. In: 2023 3rd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), Greater Noida, India, 2023, pp. 1700–1704 (2023). https://doi.org/10.1109/ICACITE57410.2023.10183114
4. Korkmaz, A., Alhonainy, A., Rao, P.: An evaluation of federated learning techniques for secure and privacy-preserving machine learning on medical datasets. In: 2022 IEEE Applied Imagery Pattern Recognition Workshop (AIPR) (2022). https://doi.org/10.1109/AIPR57179.2022.10092212
5. Lai, W., Yan, Q.: Federated learning for detecting COVID-19 in chest CT images: a lightweight federated learning approach. In: 2022, 4th International Conference on Frontiers Technology of Information and Computer (ICFTIC), Qingdao, China, pp. 146–149 (2022). https://doi.org/10.1109/ICFTIC57696.2022.10075165
6. Abidin, N.Z., Ismail, A.R.: Federated deep learning for automated detection of diabetic retinopathy. In: 2022 IEEE 8th International Conference on Computing, Engineering and Design (ICCED), Sukabumi, Indonesia, pp. 1–5 (2022). https://doi.org/10.1109/ICCED56140.2022.10010636
7. Da Costa, L.F., Furtado, L.S., Rocha, P.H., Rego, P.A., Trinta, F.A.: Time series prediction in IoT: a comparative study of federated versus centralized learning. In: 2023 IEEE 20th Consumer Communications & Networking Conference (CCNC). IEEE (2023). https://doi.org/10.1109/CCNC51644.2023.10060467
8. Gupta, A., Maurya, C., Dhere, K., Chaurasiya, V.K.: Wellness detection using clustered federated learning. In: 2022, IEEE 6th Conference on Information and Communication Technology (CICT), Gwalior, India, pp. 1–5 (2022). https://doi.org/10.1109/CICT56698.2022.9997827
9. Bhanumathi, V., Dhanasekaran, R.: TCP variants - a comparative analysis for a high bandwidth-delay product in a mobile ad-hoc network. In: 2010, the 2nd International Conference on Computer and Automation Engineering (ICCAE) (2010). https://doi.org/10.1109/ICCAE.2010.5451683
10. Wang, A., Zhao, Y., Yang, L., Wu, H., Iwahori, Y.: Heterogeneous defect prediction algorithm combined with federated sparse compression. IEEE Access **11**, 23739–23753 (2023). https://doi.org/10.1109/ACCESS.2023.3253765
11. Gupta, S., Singh, Y.: Comparative analysis of newer congestion control algorithms in high BDP networks (2022). https://doi.org/10.56726/IRJMETS30391
12. Wang, K., Deng, N., Li, X.: An efficient content popularity prediction of privacy preserving based on federated learning and Wasserstein GAN. IEEE Internet Things J. **10**(5), 3786–3798 (2022). https://doi.org/10.1109/JIOT.2022.3176360
13. Nguyen Tan, Y., Tinh, Y.P., Lam, P.D., Nam, N.H., Khoa, T.A.: A transfer learning approach to breast cancer classification in a federated learning framework. IEEE Access 27462–27476 (2023). https://doi.org/10.1109/ACCESS.2023.3257562
14. Pereira, K., Parikh, A., Kumar, P., Devadkar, K.: Healthcare diagnostics service using federated learning. In: 2023 International Conference for Advancement in Technology (ICONAT). IEEE (2023). https://doi.org/10.1109/ICONAT57137.2023.10080053
15. Sun, W., Zhao, Y., Ma, W., Guo, B., Xu, L., Duong, T.Q.: Accelerating convergence of federated learning in MEC with dynamic community. IEEE Trans. Mob. Comput. (2023). https://doi.org/10.1109/TMC.2023.3241770
16. Zong, L., Qiao, D., Wang, H., Bai, Y.: Sustainable cross-regional transmission control for the industrial augmented intelligence of things. IEEE Trans. Industr. Inf. (2022). https://doi.org/10.1109/TII.2022.3230674

17. Cai, Q., Chaudhary, S., Vuppalapati, M., Hwang, J., Agarwal, R.: Understanding Host Network Stack Overheads. ACM (2021). ISBN 978-1-4503-8383-7/21/08. https://doi.org/10.1145/3452296.3472888
18. Kumar, R., et al.: TCP BBR for ultra-low latency networking: challenges, analysis, and solutions. In: 2019 IFIP Networking Conference (IFIP Networking), pp. 1–9. IEEE (2019). ISBN 978-3-903176-16-4