Statistical Learning-Based Dynamic Retransmission Mechanism for Mission Critical Communication: An Edge-Computing Approach

1st Muhammad Ahmad Raza
School of Electrical and Data Eng.
University of Technology Sydney
Sydney, Australia
MuhammadAhmad.Raza@student.uts.edu.au

2nd Mehran Abolhasan
School of Electrical and Data Eng.
University of Technology Sydney
Sydney, Australia
Mehran.Abolhasan@uts.edu.au

3rd Justin Lipman
School of Electrical and Data Eng.
University of Technology Sydney
Sydney, Australia
Justin.Lipman@uts.edu.au

4th Negin Shariati
School of Electrical and Data Eng.
University of Technology Sydney
Sydney, Australia
Negin.Shariati@uts.edu.au

5th Wei Ni
Data 61
Commonwealth Scientific and Industrial Research Org.
Sydney, Australia
wei.ni@data61.csiro.au

***Abstract*—Mission-critical machine type communication (MC-MTC) systems in which machines communicate to perform various tasks such as coordination, sensing, and actuation, require stringent requirements of ultra-reliable and low latency communications (URLLC). Edge computing being an integral part of future wireless networks, provides services that support URLLC applications. In this paper, we use the edge computing approach and present a statistical learning-based dynamic retransmission mechanism. The proposed approach meets the desired latency-reliability criterion in MC-MTC networks employing framed ALOHA. The maximum number of retransmissions Nr under a given latency-reliability constraint is learned statistically by the devices from the history of their previous transmissions and shared with the base station. Simulations are performed in MATLAB to evaluate a framed-ALOHA system's performance in which an active device can have only one successful transmission in one round composed of (Nr + 1) frames, and the performance is compared with the diversity transmission-based framed-ALOHA.**

***Index Terms*—Mission-critical machine type communication; retransmissions; edge computing; framed-ALOHA**

I. INTRODUCTION

Mission-critical machine type communication (MC-MTC) systems require stringent requirements of ultra-reliable and low-latency communications (URLLC). Remote sensing, autonomous transport, Industory 4.0, robot control, and telesurgery are among the emerging applications of MC-MTC networks. In such systems, messages from MTC devices need to be delivered successfully at the base station (BS) within a prescribed end-to-end latency of L (ms). From the physical (PHY) layer perspective, the concept of reliability is related to the packet-error rate (PER). However, reliability can also be defined as the probability of satisfying a latency bound L (ms) [1], and this notion of reliability is more useful while addressing URLLC requirements at the medium access control (MAC) layer and the other higher layers. If L_D is the latency experienced by a packet from an MTC device, and ϵ_r is the reliability constraint, the MC-MTC system is required to exhibit $\Pr\left(L_D \leq L\right) \geq 1-\epsilon_r$, where $\Pr\left(.\right)$ denotes the probability measure. Future MC-MTC networks aim to achieve $\epsilon_r \leq 10^{-5}$ and $L \leq 1$ (ms).

Several PHY and MAC layer techniques have been proposed to design URLLC based systems [2]–[4]. Short packet transmission and grant-free non-orthogonal multiple access methods can reduce latency considerably [5], while the diversity-based and retransmission schemes enhance the reliability [6]–[10]. It is identified that present approaches are primarily BS centered and employ centralized decision-making schemes where a central controller or a BS performs all the network-level decisions, which causes additional latency. It becomes very challenging to meet the required latency-reliability criterion when network parameters change dynamically. Hence, the network edge-nodes and edge-devices must have the capability of learning and adapting to the network dynamics. Moreover, current literature focuses on the schemes which involve either a single transmission or a fixed number of retransmissions and replications. However, in MC-MTC networks, the retransmissions limit needs to be adapted dynamically according to the network conditions, which is the main focus of this paper.

Edge computing is an integral part of future wireless networks that enable distributed computing, storage, and control services at the network edge-nodes. These features of edge computing can lead to provisioning a platform that is suitable for mission-critical applications. In this regard, potential enablers for edge computing-based mission-critical applications are discussed in [1]. This paper considers MC-MTC networks employing framed-ALOHA and uses an edge computing approach to demonstrate how edge-devices can

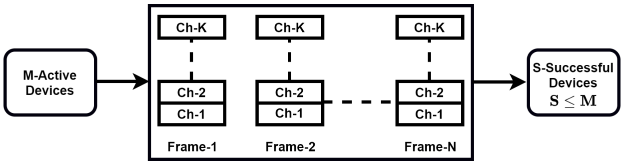

Fig. 1. Framed-ALOHA based transmission over one MCR composed of N frames.

help the edge-node/BS determine the retransmissions limit. The following are the key contributions of this paper:

- For MC-MTC networks employing framed-ALOHA, we present an edge computing-based statistical learning mechanism to predict the retransmissions limit N_r, which can meet the desired latency-reliability criterion. A sequence of $(N_r + 1)$ frames is termed as a mission-critical round (MCR), and each device uses its history of previous J MCRs to estimate the collision probability in one frame. This estimate is used to formulate a value at risk (VaR) problem to predict the retransmissions limit under a given latency-reliability constraint. Finally, each device shares the prediction of N_r with the BS.
- Through simulations, we present the performance analysis of a restricted MCR-based framed-ALOHA system in which after the first successful transmission, the device stops transmitting in the current MCR and attempts in the next MCR if it has another packet to transmit. In this regard, we compare the performance of a restricted MCR based framed-ALOHA system with the diversity transmission-based framed-ALOHA (DTFA).

II. RETRANSMISSIONS LIMIT PREDICTION

We consider a homogeneous MC-MTC network composed of W MTC-devices in which, at a given instant, $M \leq W$ active devices attempt to communicate with one BS. The uplink transmission between the MTC-devices and the BS is modeled as multi-channel slotted ALOHA or framed-ALOHA, which is also used by Long-Term Evaluation (LTE) during the contention phase [11]. As shown in Fig. 1, each frame is composed of K channels or resource blocks, and a sequence of $N = N_r + 1$ frames is called a mission-critical round (MCR). All active devices begin to transmit at the start of an MCR. It is assumed that an active device will always have a packet to transmit throughout the MCR. In each frame, every active device selects one of the K available channels randomly following a uniform distribution independently from other devices. We perform the MAC layer analysis only while considering PHY layer abstraction in which transmission fails if two or more devices select the same channel, and the failed devices attempt again in the next frame. Upon successful transmission, the device receives an acknowledgment from the BS and continues to transmit in subsequent frames. The parameter N_r is the maximum affordable retransmissions, and the value of N_r is determined at the edge dynamically. The number of active devices M at the start of each MCR is not known by the devices and the BS a priori. If processing and propagation delays are assumed constant, latency (L_D) is primarily a function of N_r. We aim to estimate the collision probability in one frame and then determine the optimal number of retransmissions at the edge, such that the desired latency-reliability criterion is met.

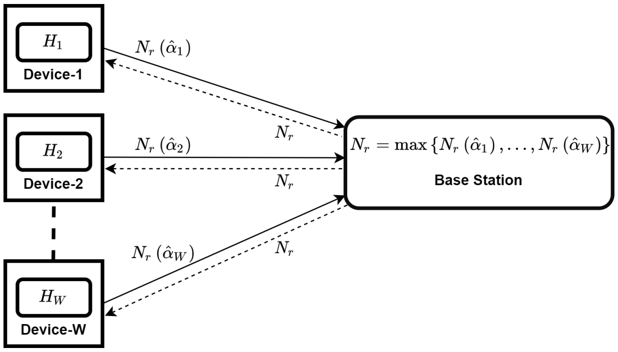

Fig. 2. MTC devices sharing locally predicted retransmissions limit with the BS.

The number of transmission attempts an MTC device performs for a successful transmission depends upon the collision probability in one frame. The probability that an MTC device of interest will collide with at least one of the $(M-1)$ devices in one frame is given as [12]:

$$\alpha := 1 - \left(1 - \frac{1}{K}\right)^{M-1} \tag{1}$$

Since, exact value of M is not known by the devices, we cannot use (1) to determine the exact value of α at the device level. However, the devices can estimate the value of α by using the history of their previous transmissions. For that purpose, each device keeps the record of its last $N_h = JN$ transmissions attempts spanned over J MCRs in a vector $H_m = \left[A_m^{(1)}, A_m^{(2)}, ..., A_m^{(N_h)}\right]$, where each element of H_m is an independent Bernoulli random variable defined as:

$$\left\{A_m^{(n)}\right\}_{\substack{m=1,2,..,W \\ n=1,2,..,N_h}} = \begin{cases} 1 & collision\ with\ other\ device/s \\ 0 & successful\ transmission \end{cases} \tag{2}$$

The estimate of collision probability at the m^{th} device denoted by $\hat{\alpha}_m$ is computed as:

$$\hat{\alpha}_m = \frac{1}{N_h} \sum_{n=1}^{N_h} A_m^{(n)} \tag{3}$$

Risk sensitive learning and control is a promising tool to address URLLC related problems [2]. We use the collision probability estimate $\hat{\alpha}_m$ to formulate a value at risk (VaR) problem to predict the number of retransmissions the MTC device is allowed under a given latency-reliability constraint. Let the random variable X_m show the number of collisions faced by m^{th} device before a successful transmission. The random variable X_m follows the geometric distribution, and the probability that a device undergoes up to N_c collisions

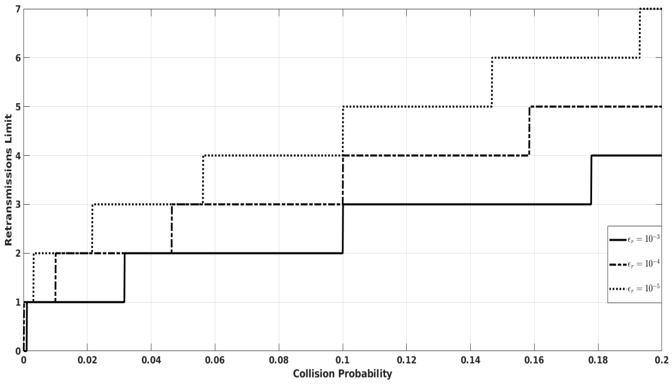

Fig. 3. Retransmissions limit against the collision probability for different values of reliability constraint ϵ_r.

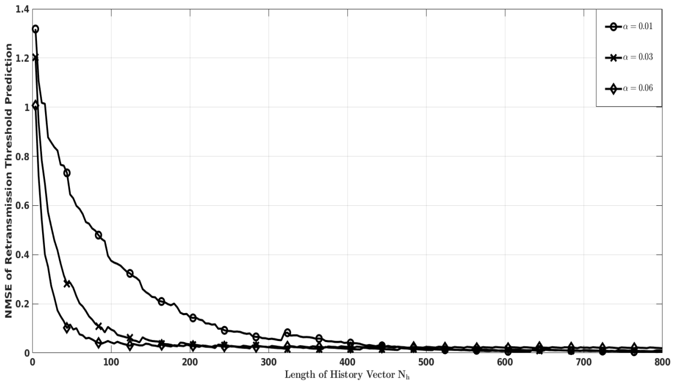

Fig. 4. NMSE of locally predicted retransmission limit $N_r\left(\hat{\alpha}_m\right)$ against length of history vector for different values of collision probability.

before having a successful transmission is given as:

$$\Pr\left(X_m \leq N_c \mid \hat{\alpha}_m\right) = 1 - \left(\hat{\alpha}_m\right)^{N_c+1} \tag{4}$$

Each retransmission adds to the latency experienced by a data packet in a given MCR, and there exists a maximum value of retransmissions under the given latency-reliability constraint, after which it will be too late to receive that data packet at the BS. Thus, the retransmissions limit N_r can be found by computing the VaR of X_m as follows:

$$N_r\left(\hat{\alpha}_m\right) = \inf_{N_c}\left\{N_c \geq 0 : \Pr\left(X_m \leq N_c \mid \hat{\alpha}_m\right) \geq 1 - \epsilon_r\right\} \tag{5}$$

An optimal value of N_r that can meet the stringent requirements of the URLLC will require that the collision probability α is kept very small. Each device predicts the retransmissions limit $N_r\left(\hat{\alpha}_m\right)$, and shares with the BS as a part of its data packet. As shown in Fig. 2, the BS keeps a record of the last update sent by each device in a vector $R = \left\{N_r\left(\hat{\alpha}_1\right), N_r\left(\hat{\alpha}_2\right), ..., N_r\left(\hat{\alpha}_L\right)\right\}$. After every F number of MCRs, the BS broadcasts its updated retransmissions limit $N_r = \max\left\{N_r\left(\hat{\alpha}_1\right), N_r\left(\hat{\alpha}_2\right), ..., N_r\left(\hat{\alpha}_L\right)\right\}$ to be used by all the devices for determining the size of the MCR, and the outage event, which is part of our future research work. A device is said to be successful in one MCR if it has at least one successful transmission in that MCR, and the probability of this event is given as:

$$p_{suc}^{(N)} = 1 - \left(\alpha\right)^N \tag{6}$$

Fig. 3 provides an insight into the number of retransmissions a system should allow for different values of reliability constraint ϵ_r against the collision probability α. It is interesting to note that a specific value of N_r can be valid for a range of α. Moreover, due to the discrete nature of N_c, Equation (5) can yield same value of N_r for two different values of $(1-\epsilon_r)$ valid over a range of collision probability. In such cases N_r corresponds to retransmissions limit for higher value of $(1-\epsilon_r)$. Since each active device updates its history vector after each attempt, the BS captures the increasing traffic load and updates value of N_r after every F number of MCRs. Thus, the network adapts to the dynamic changes by learning the network statistics, i.e., the collision probability.

III. SIMULATION RESULTS AND DISCUSSION

Extensive simulations are performed in MATLAB to evaluate performance of the proposed system. In Fig. 4, for $\epsilon_r = 10^{-5}$ and $N_r = 3$, the normalized mean squared error (NMSE) of $N_r\left(\hat{\alpha}_m\right)$ is plotted against length of the history vector H_m used to predict the retransmission threshold. The NMSE is defined as: $\text{NMSE} = \frac{\text{MSE}}{\left[N_r(\alpha)\right]^2}$, and the mean squared error (MSE) is computed as:

$$\text{MSE} = \frac{1}{N_s}\sum_{n=1}^{N_s}\left[N_r\left(\alpha\right) - N_r^{(n)}\left(\hat{\alpha}_m\right)\right]^2, \tag{7}$$

Where $N_s = 10000$ is the number of iterations performed to compute MSE against one value of N_h, and $N_r^{(n)}\left(\hat{\alpha}_m\right)$ is the retransmission limit prediction in n^{th} iteration. As shown in Fig. 3, there exists a unique value of $N_r\left(\alpha\right)$ for a specific interval of α. We pick three different values of α such that each corresponds to a different value of $N_r\left(\alpha\right)$. For each value of α, the NMSE of retransmission threshold prediction is plotted against N_h. It is shown that the NMSE of retransmission threshold prediction decreases randomly when the value of N_h is increased and becomes stable asymptotically when N_h is large.

A. *Restricted-MCR*

In restricted-MCR, all the active devices begin to transmit at the start of an MCR. However, after the first successful transmission, the device stops transmitting in the current MCR and attempts in the next MCR if it has another packet to transmit. The restricted transmission strategy helps reduce the collision probability in successive frames of an MCR, which results in latency reduction. On the other hand, in the diversity transmission-based framed-ALOHA (DTFA) method, each active device sends one replica of its message in N frames of an MCR, such that in each frame, it selects one of the K channels randomly. Devices may collide in some frames, but they can also be successful in some other frames. This

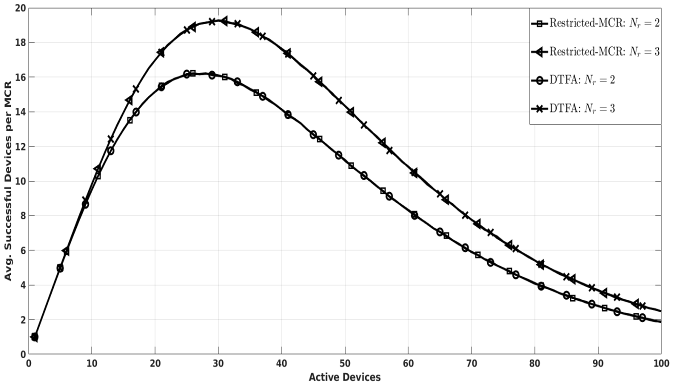

Fig. 5. Performance comparison of restricted-MCR based framed-ALOHA and DTFA in terms of average successful devices with $K = 20$.

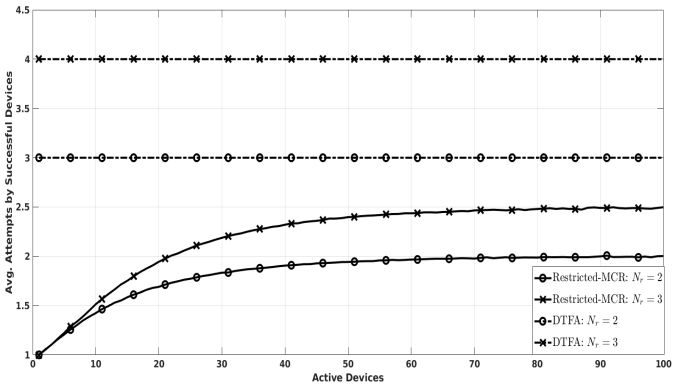

Fig. 6. Performance comparison of restricted-MCR based framed-ALOHA and DTFA in terms of average attempts by a successful device with $K = 20$.

helps increase the number of successful devices by reducing the overall collision probability. The analytical expression for the overall collision probability of the framed-ALOHA based diversity transmission scheme is provided in [13].

We compare the performance of the restricted-MCR based framed-ALOHA with the DTFA scheme. Performance is compared in terms of the average number of successful devices and average transmission attempts performed by successful devices in one restricted-MCR against a range of active devices. For the DTFA strategy, we consider the case where a device is successful if at least one transmission remains successful in one MCR. As shown in Fig. 5, for different values of N_r, both schemes depict similar behavior with respect to the average number of successful devices in one MCR. However, as illustrated in Fig. 6, the average number of transmission attempts performed by a successful device in the restricted-MCR is less than that of the DTFA method. Thus, the overall average latency can be reduced by using the restricted-MCR based retransmission scheme.

IV. CONCLUSION

By using an edge computing approach, we present a statistical learning-based dynamic retransmission mechanism for MC-MTC networks employing framed-ALOHA. Each MTC device in the network uses its history of the previous transmissions to learn statistically the number of retransmissions it can afford such that the desired latency-reliability criterion is met. In order to have a network-wide uniform value, each device shares its knowledge of retransmissions with the base station (BS). Simulations are performed in MATLAB to evaluate the performance of MC-MTC networks employing a framed-ALOHA system for the case where each active device can have only one successful transmission in one MCR called a restricted-MCR. For the same average successful devices in one MCR, the restricted-MCR based framed-ALOHA system requires a less average number of attempts as compared to the DTFA scheme. As future work, we aim to perform the analytical modeling of restricted-MCR based framed-ALOHA for MC-MTC networks.

REFERENCES

[1] M. S. Elbamby, C. Perfecto, C. Liu, J. Park, S. Samarakoon, X. Chen, and M. Bennis, "Wireless Edge Computing With Latency and Reliability Guarantees," *Proceedings of the IEEE*, vol. 107, no. 8, pp. 1717–1737, 2019.

[2] M. Bennis, M. Debbah, and H. V. Poor, "Ultrareliable and Low-Latency Wireless Communication: Tail, Risk, and Scale," *Proceedings of the IEEE*, vol. 106, no. 10, pp. 1834–1853, 2018.

[3] P. Popovski, C. Stefanovic, J. J. Nielsen, E. de Carvalho, M. Angjelichinoski, K. F. Trillingsgaard, and A. Bana, "Wireless Access in Ultra-Reliable Low-Latency Communication (URLLC)," *IEEE Transactions on Communications*, vol. 67, no. 8, pp. 5783–5801, 2019.

[4] G. J. Sutton, J. Zeng, R. P. Liu, W. Ni, D. N. Nguyen, B. A. Jayawickrama, X. Huang, M. Abolhasan, Z. Zhang, E. Dutkiewicz, and T. Lv, "Enabling Technologies for Ultra-Reliable and Low Latency Communications: From PHY and MAC Layer Perspectives," *IEEE Communications Surveys Tutorials*, vol. 21, no. 3, pp. 2488–2524, 2019.

[5] M. B. Shahab, R. Abbas, M. Shirvanimoghaddam, and S. J. Johnson, "Grant-free Non-orthogonal Multiple Access for IoT: A Survey," *IEEE Communications Surveys Tutorials*, pp. 1–1, 2020.

[6] R. Abreu, P. Mogensen, and K. I. Pedersen, "Pre-Scheduled Resources for Retransmissions in Ultra-Reliable and Low Latency Communications," in *2017 IEEE Wireless Communications and Networking Conference (WCNC)*, 2017, pp. 1–5.

[7] R. Abreu, G. Berardinelli, T. Jacobsen, K. Pedersen, and P. Mogensen, "A Blind Retransmission Scheme for Ultra-Reliable and Low Latency Communications," in *2018 IEEE 87th Vehicular Technology Conference (VTC Spring)*, 2018, pp. 1–5.

[8] T. Chiang, H. Liang, S. Wang, and S. Sheu, "On Parallel Retransmission for Uplink Ultra-Reliable and Low Latency Communications," in *2019 IEEE 90th Vehicular Technology Conference (VTC2019-Fall)*, 2019, pp. 1–5.

[9] C. Boyd, R. Kotaba, O. Tirkkonen, and P. Popovski, "Non-Orthogonal Contention-Based Access for URLLC Devices with Frequency Diversity," in *2019 IEEE 20th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, 2019, pp. 1–5.

[10] O. Galinina, A. Turlikov, S. Andreev, and Y. Koucheryavy, "Multichannel random access with replications," in *2017 IEEE International Symposium on Information Theory (ISIT)*, 2017, pp. 2538–2542.

[11] H. Thomsen, N. K. Pratas, C. Stefanovic, and P. Popovski, "Analysis of the LTE Access Reservation Protocol for Real-Time Traffic," *IEEE Communications Letters*, vol. 17, no. 8, pp. 1616–1619, 2013.

[12] M. Deghel, P. Brown, S. E. Elayoubi, and A. Galindo-Serrano, "Uplink Contention-Based Transmission Schemes for URLLC Services," in *Proceedings of the 12th EAI International Conference on Performance Evaluation Methodologies and Tools*. New York, NY, USA: ACM, 2019, p. 87–94.

[13] B. Singh, O. Tirkkonen, Z. Li, and M. A. Uusitalo, "Contention-Based Access for Ultra-Reliable Low Latency Uplink Transmissions," *IEEE Wireless Communications Letters*, vol. 7, no. 2, pp. 182–185, 2018.