Towards System Level Simulations of Public Safety Applications over 5G NR Sidelink

Samantha Gamboa[†‡], Thomas R. Henderson[§], Wesley Garey[*], Chunmei Liu[*], Richard Rouil[*]
[*]*National Institute of Standards and Technology*, Gaithersburg, MD, USA
[†]*Associate, National Institute of Standards and Technology*, Gaithersburg, MD, USA
[‡]*Prometheus Computing LLC*, Bethesda, MD, USA
[§]*Department of Electrical & Computer Engineering, University of Washington*, Seattle, WA, USA
{samantha.gamboa, wesley.garey, chunmei.liu, richard.rouil}@nist.gov, tomhend@uw.edu

***Abstract*—**

The New Radio (NR) sidelink (SL) interface was initially specified to support NR Vehicle-to-everything (V2X) applications and lately to support NR Proximity Services (ProSe) public safety applications. Network support requirements differ between the operational demands of these two types of applications. Building upon an existing ns-3 NR V2X simulator, this paper details the extensions made to the simulator to cater to NR ProSe public safety applications. We use as case study the Mission Critical Push-to-Talk (MCPTT) application, a keystone of group communication in 5G NR public safety networks. We demonstrate through simulation results that our enhancements are essential for a more accurate representation of public safety applications.

***Index Terms*—NR, SL, MCPTT, Public Safety.**

I. INTRODUCTION

The New Radio (NR) sidelink (SL) is the interface, coupled with its inherent protocols, allowing User Equipments (UEs) to engage in direct communication without relying on the network infrastructure of the Third Generation Partnership Project (3GPP) 5G NR systems. Initially, in Release 16, the NR SL was specified to support NR Vehicle-to-everything (V2X) services, enabling communication between vehicle-mounted UEs and other in-range, authorized UEs. Afterwards, Release 17 introduced the specifications for NR Proximity Services (ProSe), targeting both commercial and public safety applications that might employ the NR SL [1].

In [2], the authors developed an NR V2X simulator based on the Network Simulator 3 (ns-3). This simulator is based on 3GPP Release 16 NR V2X standards, which define the NR SL. The authors developed an NR SL model and focused on how to simulate scenarios with out-of-coverage V2X communication. Leveraging the same simulator, the authors in [3] provided a sensitivity analysis of the role of numerology and sensing-based resource selection in an NR V2X highway scenario.

In this paper, we build upon this previous ns-3 system-level simulator and extend it with NR SL capabilities that enable the performance evaluation of public safety applications. Such capabilities include: multi-subchannel operation, multiple logical channel support per UE, dynamic scheduling, and hybrid automatic repeat request (HARQ) feedback. We also incorporated the NR ProSe layer in the model, enabling the simulation of 5G NR ProSe direct discovery, unicast communication, and L3 UE-to-Network (U2N) relay. The L3 U2N relay is particularly crucial in public safety contexts; it extends network coverage to devices that would typically be off-network, reducing service interruptions for applications requiring network access.

Our case study focuses on the Mission Critical Push-to-Talk (MCPTT) public safety application, whose ns-3 model is publicly available and can be easily integrated in 5G NR simulation scenarios [4]. MCPTT is a relevant application to study because it is the baseline group communication application for First Responders in 5G NR Public Safety networks. Moreover, MCPTT has two modes of operation: off-network, designed to operate in a distributed manner over NR SL; and on-network, which is network and server focused but which can be extended to remote units with limited network coverage by using the U2N relay technology over the NR SL. Furthermore, the MCPTT application uses communication protocols that yield mixed traffic patterns with the exchange of sporadic signaling messages and periodic media traffic, support of which goes beyond that of the current ns-3 NR V2X SL model support.

In Section II we elaborate on the limitations of the current ns-3 NR V2X SL model for public safety applications and describe the extensions we developed to solve them. In Section III we describe the relevant aspects of the MCPTT application, and in Section IV we provide simulation results that demonstrate the necessity of our developed extensions to better suit the simulation of the MCPTT application. Finally, we provide some final remarks and discuss future work in Section V.

II. NS-3 NR SL EVOLUTION

In this section, we describe our extensions to the initial ns-3 NR SL simulator (introduced in [2]) and discuss how these extensions benefit the study of the performance of public safety networks. The main use

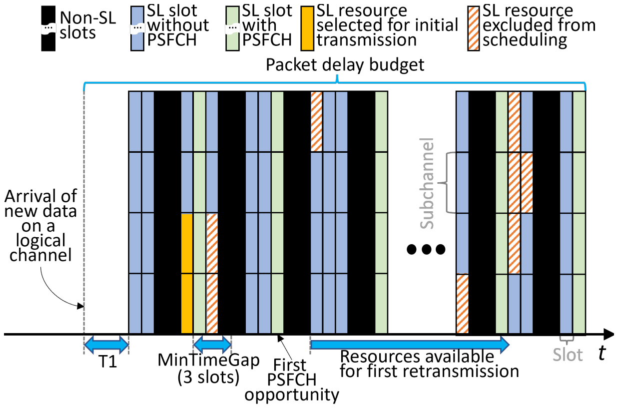


Fig. 1: Example NR SL selection window for scheduling.

case for ns-3 is to study performance aspects at the MAC layer and above, as well as aspects involving cross-layer interaction between the MAC and PHY layers. Detailed simulations of the PHY layer are more appropriate for a link level simulator such as the one described in [5].

The initial ns-3 NR SL simulator was limited at the MAC and PHY layer in the following ways:

1) **single subchannel operation:** In NR SL, the minimum frequency granularity for scheduling is the subchannel. Although resource pools with multiple subchannels can be configured in [2], several simplifications precluded the accurate use of such a configuration. Sensing of occupied subchannels that did not span all of the subchannels of a slot nevertheless led to the exclusion of the entire slot in the scheduler. In addition, limitations in the PHY model precluded the scheduling, in the same slot, of transport blocks to different destinations.

2) **single logical channel support:** The previous scheduler only supported a single logical channel within a resource pool. Multiple logical channel support is necessary to allow for different types of traffic to be supported in different ways (e.g., with different HARQ configuration) and prioritized by a scheduler.

3) **no capability for dynamic grants:** The previous simulator only supported semi-persistent scheduling (SPS)– an important scheduling mode for V2X, but aimed specifically at applications generating small amounts of periodic traffic, such as basic safety messages. Many types of traffic are not periodic but instead are intermittent and should be scheduled with one-time single-PDU dynamic grants instead of SPS grants.

4) **lack of HARQ feedback:** The previous simulator supports only blind retransmissions– retransmissions that are scheduled and sent without any feedback from the receiver. However, an important feature of NR SL is the ability to configure and send explicit feedback to suppress unnecessary retransmissions.

The above limitations have been addressed as follows. Consider the resource selection window depicted in Fig. 1. The horizontal axis displays time advancing from left to right in units of slots. The vertical axis depicts frequency resources, where the smallest granularity available for scheduling is the subchannel. Consider the left-most time point depicting the arrival of user data for which a transmission and possible retransmissions must be scheduled. The scheduler operates under the constraint of satisfying the packet delay budget: transmissions scheduled for later than this deadline should be avoided. The parameter *T1*, configured by default in ns-3 to two slots, accounts for the processing time before a newly arrived data packet can be scheduled.

In Fig. 1, the resource pool allocates three out of every five slots to SL data, and the other two slots to some other use (such as uplink or downlink data). The black slots (two out of every five) are unavailable for scheduling, leaving three out of five (primarily blue and green in the figure) available for SL scheduling. In addition, among the SL resources, some are excluded from scheduling due to anticipated future transmissions corresponding to data that was sensed in the sensing window preceding the current time. Notice that these resource exclusions do not, in general, span all of the subchannels of the slot. Our first enhancement was to make the ns-3 sensing algorithm conformant with the 3GPP standard by not excluding entire slots when only certain subchannels are sensed to be occupied.

The next enhancement was the introduction of a physical SL feedback channel (PSFCH) model to support HARQ feedback operation. HARQ feedback must be returned on a PSFCH channel, which occupies one symbol in selected SL slots. The standard specifies that PSFCH may be configured to repeat every one, two, or four SL slots. In Fig. 1, a configuration with a PSFCH period of four SL slots is shown.

PSFCH data encoding is different from that of physical SL shared channel (PSSCH) and physical SL control channel (PSCCH), and we are not aware of any link simulator that provides error models for PSFCH channels. Therefore, in our current extension, we model the PSFCH as a perfect feedback channel, and any contention in the use of it is resolved by always preferring to send available feedback over waiting to listen for feedback. In practice, slots including the PSFCH channel have three fewer symbols available for PSSCH data than those without PSFCH. As a result, the data transfer capacity of these two slot types differs, which may make it more difficult to schedule transport blocks of certain sizes. In our current scheduler, we make a conservative assumption that selected resources must fit within both slot sizes.

Despite these simplifications, our initial PSFCH model is important for the following reason: it constrains the scheduling of retransmission slots when HARQ feedback is enabled, leading to a better latency model. This constraint is according to the 3GPP standard and

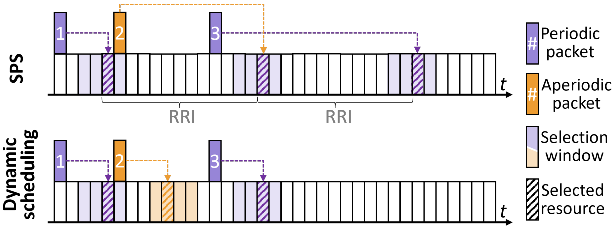


Fig. 2: SPS vs. dynamic scheduling.

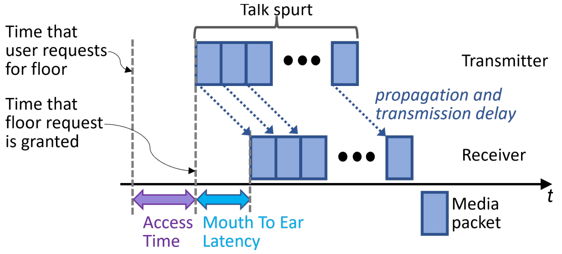


Fig. 3: MCPTT key performance indicators (KPIs).

is depicted in Fig. 1. Assume that the scheduler has to schedule a transmission spanning two subchannels, as depicted in yellow in Fig. 1. Assume also that feedback-based HARQ retransmissions are enabled. In this case, the standard specifies that the scheduling of retransmission-based resources must allow for a time gap that would permit HARQ feedback (delivered on the PSFCH) to be received by the transmitter. If positive (ACK-based) feedback is received, the retransmission can be suppressed, lowering the overall interference in the network. The time gap provides the receiver with some time to decode and process the previous transmission. This parameter is called *MinTimeGapPsfch* in the standard and may be configured to either two or three slots [6]. The transmitter assumes that feedback will not be attempted by the receiver until at least *MinTimeGapPsfch* slots have elapsed. In Fig. 1, this precludes the use of the immediately adjacent slot, but the next PSFCH slot indicated as the "first PSFCH opportunity" may be used. The standard also specifies that a period of two slots should be available to receive and process the PSFCH. Therefore, the next retransmission for this initial yellow transmission must only be scheduled after this delay has elapsed, as depicted in Fig. 1. Similarly, subsequent retransmissions (if more are desired) must respect these time gaps to allow for feedback to be delivered and processed. This is in contrast to the previous simulator's support of blind retransmissions only, which have no such scheduling constraints, but which have no mechanism to suppress unnecessary retransmissions.

Based on the inclusion of the PSFCH channel and the enforcement of minimum time gaps on the receiver and transmitter logic, HARQ operation was added as a configurable option to the logic at the MAC layer. This includes the use of a timer to protect against the case in which feedback is never received (such as if all transmissions of a transport block were lost). The number of simultaneous SL HARQ processes allowed by the 3GPP standard is sixteen from which only a maximum of four can be used for SPS grants when SL resource allocation mode 2 is used [6]. Since each HARQ process protects an individual transport block, this can limit the number of outstanding transport blocks to four and can lead to additional latency if data must wait for a HARQ process to become available. HARQ can be configured for unicast and groupcast modes of communication. In the case of groupcast, two options are specified in the standard. Option 1 defines a radius around the transmitter, and all nodes within that radius failing to decode a transport block that was announced in a successfully received control message will report a NACK to the transmitter, while nodes outside of the radius will suppress such feedback. In option 2, all receivers (regardless of range) will send an ACK or a NACK corresponding to the reception. Our simulation extensions support option 2 only, and additionally we made the policy decision to consider that the HARQ process can be freed if at least one receiver in the groupcast returns an ACK. Other policies such as waiting for K out of N possible ACKs to be received could also be easily supported.

The other important extensions are the replacement of the single logical channel, single scheduling mode (SPS) scheduler with a scheduler that supports multiple logical channels as well as dynamic single-PDU grants. Fig. 2 shows a simplified example of the limitation of SPS grants with fixed resource reservation interval (RRI) when handling aperiodic traffic. Both, the aperiodic packet and the second periodic packet experience large transmission delay as the reserved resources of the SPS grant are not enough to transmit both in a timely manner. Dynamic scheduling provides the flexibility needed to handle aperiodic traffic efficiently.

When multiple logical channels are scheduled, if resources are constrained, the order of scheduling may affect the availability of resources for each logical channel. This is configurable by a priority setting, but in the case of equal priorities, other heuristics must be introduced. Our current scheduler will randomize the order of scheduling destination layer-2 IDs in such case. Scheduling can be performed for multiple subchannels in a single slot, and the PHY layer models have been extended accordingly.

III. MCPTT

MCPTT is an application layer protocol designed by 3GPP to provide First Responders with an arbitrated method of group communication [7]. In essence, this protocol allows UEs to realize a walkie-talkie-like behavior where a user pushes a Push-to-Talk (PTT) button on their device to send voice to one or more other devices. This functionality is achieved via the Call Control [7] and Floor Control [8] services.

Call Control is responsible for group, session, and resource management. This includes realizing different types of calls and modes of operation. The type of call dictates what features are available and who the initiating user can communicate with. For example, a private call is directed at a specific user, but a group call is directed toward a group of one or more users. Each type of call can exist in two different modes of operation, either on-network or off-network. In on-network mode, all signaling is transmitted over the network to an MCPTT server, while in off-network mode, packets are exchanged directly between UEs. Thus, on-network mode is server-centric and relies on the network to manage call sessions and resources, while off-network mode is peer-to-peer.

Floor Control is responsible for arbitrating who can talk at any given time during an ongoing call. During a group call, floor participants are capable of sending requests, which if granted, would allow the user to communicate with the other users in the group. The floor arbitrator has the authority to grant, deny, or queue a user's request to talk. During an on-network call, the MCPTT server assumes the role of the floor arbitrator and the User Equipments (UEs) are the Floor Participants. However, during an off-network call, UEs must transition between the roles of floor arbitrator and floor participant. Which role an off-network UE assumes is determined by whether or not they currently have the floor. More detailed information about call and floor control protocols can be found in [4].

To measure the performance of an MCPTT application, 3GPP has defined several Key Performance Indicators (KPIs), two of which are the access time and Mouth-to-Ear (M2E) latency [9] (see Fig. 3). Access time is the time from when a user requests to speak, (e.g., a PTT button push), until they receive an indication to talk. M2E latency is the time from when a user begins speaking into their device until the time it is heard at a receiving user's device. Thus, these two KPIs are heavily influenced by network performance as they both rely on the exchange of either application signaling or application data. The access time depends on signaling since the exchange of several call control and/or floor control messages is required to access the floor. M2E latency depends on the exchange of data since it encompasses how long it takes the media/voice packets to travel from the UE that is occupying the floor to the other UEs in the call.

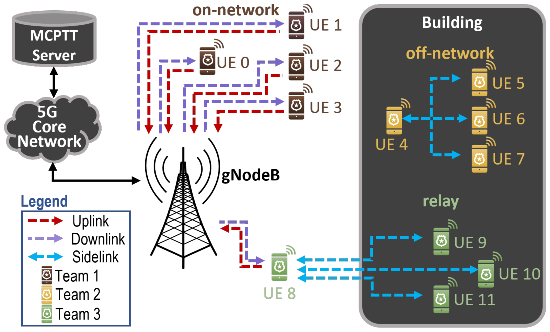


Fig. 4: MCPTT Scenario.

It is important to note what the typical traffic pattern looks like for the application signaling and data. Call control and floor control messages will typically be generated sporadically and will vary in size depending on what MCPTT actions are carried out. Media messages will typically be generated at a constant rate and size depending on the selected audio encoder and its configuration. However, the total amount of signaling or data that is generated by the MCPTT application will vary in time depending on user activity [4].

IV. SIMULATIONS

We demonstrate the utility of our developed extensions with a small simulation campaign based on public safety concepts and on the MCPTT application introduced above. All simulation programs are archived online and publicly available[1] to support reproducibility.

A. Scenario

We consider the scenario depicted in Fig. 4, where three teams of first responders are working at an incident and are using MCPTT applications to communicate with each other. Each team is part of a different MCPTT group call. The first team's members are all within gNodeB coverage and directly connected to the network. Thus, they use the on-network mode of operation for their call, and we refer to them as the *on-network* group. The members of the second team are inside a building. Due to the building's walls, they cannot connect to the gNodeB and therefore to the network. However, they communicate among themselves using the NR SL. This group uses the off-network mode of operation for their MCPTT call, and we referred to them as the *off-network* group. For the third team, one member is positioned at the entrance of the building and the other members are inside. The member at the door can connect both with the team inside using the NR SL and with the gNodeB using the traditional links, thus providing UE-to-Network Relay services to bridge the connection between the inside team members and the network. This group uses the on-network mode of operation for their call, and we refer to them as the *relay* group.

The parameters that control the pusher behavior and the media traffic flow once the floor request is granted are listed in Table I. For each set of parameters, we run the simulation until 1000 successful floor accesses (granted floor requests) are observed for each team.

---

[1] https://github.com/usnistgov/psc-ns3/tree/wfpst-2024-nr-sl-mcptt

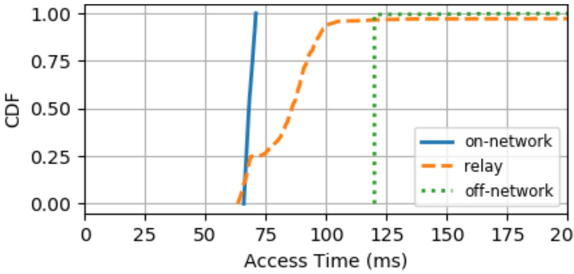
TABLE I: Simulation Parameters.

System Parameters	
Channel model	3GPP Spatial, Rural Macro
Central frequency	5.89 GHz (band n47)
Bandwidth	20 MHz UL/DL + 20 MHz SL
UE Transmit power	23 dBm[3]
Numerology (μ)	0, 2
Subchannel size	10 RBs
Number of subchannels	10 (μ=0), 2 (μ=2)
TDD pattern	DL|DL|DL|F|UL|UL|UL|UL|UL|UL
SL parameters	
Sidelink bitmap	$\{1, 1, 1, 1, 1, 1, 0, 0, 0, 1, 1, 1\}$
PSCCH MCS index	0 (fixed)
PSSCH MCS index	14 (fixed)
HARQ retransmission scheme	Blind-based, feedback-based
Number of transmissions (nTx)	1-4
Scheduling algorithm	SPS with RRI = 20 ms, Dynamic (Dyn)
Selection window size (time)	16 ms
Selection window size (slots) ($T_2 - T_1$)	16 (μ=0), 64 (μ=2)
MCPTT parameters	
Pusher PTT Inter-arrival Time	$X \sim \mathcal{N}(\mu = 10\ s, \sigma^2 = 5\ s)$
Pusher PTT Duration	$X \sim \mathcal{N}(\mu = 1\ s, \sigma^2 = 1\ s)$
Media Traffic pattern	Constant Bit Rate (CBR)
Media Packet Size	60 Bytes
Media Packet interval	20 ms
Media Data rate	24.0 kb/s

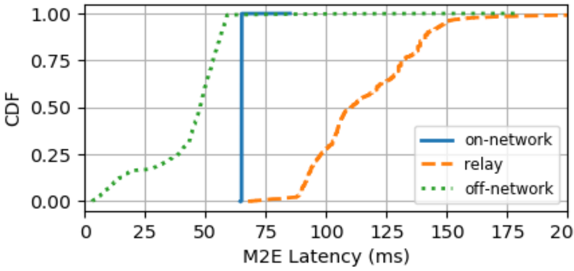
[3]Power level expressed in decibels (dB) with reference to a milliwatt (mW).

B. Results

We first focus on a configuration comparable to the MCPTT over LTE results in [4]; this helps to draw out the initial distinctions that arise between an LTE-based simulation and an NR SL-based simulation. From Fig. 5, a clear performance trend based on the group type can be observed. The on-network and relay groups use on-network MCPTT where traffic is exchanged between the MCPTT server and the UEs. Consequently, both metrics are dominated by the backhaul round-trip delay between the server and the access network, which is set to 60 ms in the scenario to be consistent with [4]. For the relay group, the relay node experiences a similar access delay to the on-network nodes, but the other three nodes in the group experience that delay plus an off-network component, which explains the discontinuity in the relay curve. The off-network group access time is the largest as the distributed protocol takes several signaling message exchanges to grant a floor request. On the other hand, the M2E latency for the off-network group is the shortest, as media traffic is broadcast on the NR SL and reaches the team members in a single hop. The relay group displays a more considerable M2E latency as the relay manages a traffic volume considerably higher than of other UEs. For instance, every media packet generated by a team member UE is directed to the MCPTT server who generates the corresponding three downlink packets directed to the other UEs. Subsequently, the relay UE transmits all the packets directed towards the remote UEs in the SL, each via the respective U2N relay unicast link. As a result, the relay M2E latency is the sum of initial off-network delay, the backhaul delay, and again an off-network component, without the benefit of broadcast. These trends are consistent with the MCPTT over LTE results in [4], which in contrast show larger off-network components due to LTE SL periodic scheduling.



(a) Access Time.



(b) M2E latency.

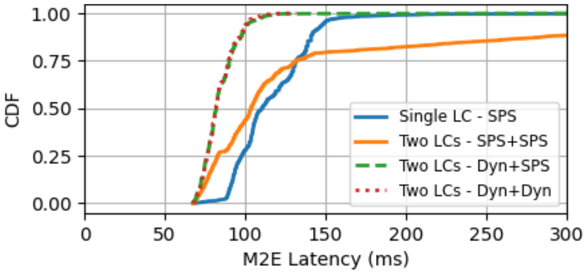
Fig. 5: MCPTT metrics for the different groups when using SPS scheduling with single LC. $\mu = 0$, nTx = 1.

Fig. 6 illustrates the M2E latency outcomes for both relay and off-network groups under various NR SL scheduling strategies. In the first strategy (Single LC - SPS), the UEs use only one LC, configured for SPS scheduling, for both signaling and media traffic. This is the legacy simulator behavior, and these curves can be found also in Fig. 5. With the second strategy (Two LCs - SPS+SPS), each UE uses two LCs for traffic, one for signaling and the other for media. Both LCs employ SPS scheduling. With the last two strategies, the UEs again use two LCs for their traffic: the LC for the signaling traffic uses the dynamic scheduler while the LC for the media traffic uses SPS (Two LCs - Dyn+SPS) or dynamic (Two LCs - Dyn+Dyn) scheduling.

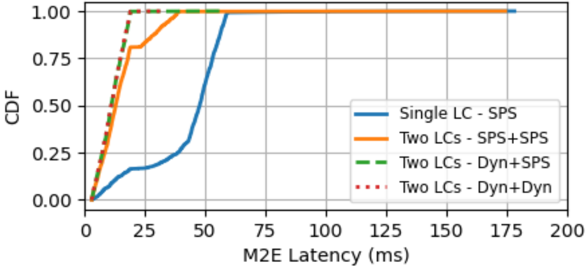
Fig. 6 suggests that the legacy scheduler strategy is sub-optimal for the MCPTT application, for both types of groups, because the single SPS grant pattern is not aligned well with the traffic arrival. A clear improvement in M2E latency is evident when distinct LCs are designated for signaling and media traffic. However, using SPS for signaling traffic is still inefficient, because the traffic is not periodic, and the repeated grants may be wasted. Another performance trend visible for the relay group is that the use of two SPS-based LCs leads to the worst latency; the demand for SL HARQ processes to serve the team's traffic on the relay node exceeded the availability (limited to four for SPS grants [6]), and

SPS grants occupy them for many cycles causing the traffic to queue while waiting for availability. Using dynamic scheduling for the signaling traffic yields the best performance for the M2E latency of both groups, and a small difference is observed based on the media traffic's scheduling method. This underscores that the scheduling of signaling traffic using dynamic scheduling minimizes overhead and enables efficient media delivery.

The above scenario configuration provides a high wireless link quality with error-free reception between team members at the physical layer. In such conditions, packet losses still occur due to scheduling collisions and half-duplex operation (UEs cannot receive and transmit simultaneously in the same slot). The tradeoffs between different SL retransmission strategies depend on higher layer protocols and KPIs and on the resource constraints in the resource pool. Performance of feedback-based and blind retransmissions due to propagation effects can be studied in a link simulator [5], while a system simulator is better suited to study MAC and higher layer effects. In the scenario presented above, the most challenging group to support from a resource contention perspective is the relay group, because the relay is responsible for using the channel frequently and cannot use broadcast to reach the other UEs. The relay is also susceptible to reaching the limit on the number of SL HARQ processes permitted by the standard. This study did not systematically explore retransmission tradeoffs, but we were able to confirm that feedback-based HARQ is successful in cancelling unnecessary retransmissions, and in the relay configuration depicted above, to improve the packet delivery ratio (PDR) for the media packets. For example, with $\mu = 2$, for a simulation lasting for 1000 floor accesses, the media PDR is approximately 95 % with no retransmissions, 89 % with blind retransmissions and nTx = 4, and 99 % for feedback-based HARQ retransmissions with up to four transmissions. The use of blind retransmissions in this case contributed to performance degradation as compared with the no retransmission case, due to the heavier use of the channel. We experimented with moving the relay node away from the building and found that these results extended to a distance of 500 meters.



(a) Relay group.



(b) Off-network group.

Fig. 6: M2E latency depending on the scheduling strategy. $\mu = 0$, nTx = 1.

V. CONCLUSION

This paper has described recent extensions to the ns-3 simulation models for 5G NR SL, supporting a more accurate MAC-layer model for studying aspects such as scheduling, latency, retransmission policies, and the impact of resource constraints or congestion. The main improvements provide a more flexible scheduler able to schedule on multiple logical channels, for both dynamic and SPS grants, and across multiple subchannels in the resource pool. Also, a unicast and groupcast SL HARQ feedback model has been added. The utility of these extensions for public safety network studies has been demonstrated with a small case study using the MCPTT application on different configurations. These simulation extensions are planned to be contributed upstream to the 5G NR ns-3 simulation module, once further documentation and testing have been completed. Our planned future work includes the addition of a sidelink-specific error model, and refactoring to improve the modularity of SL and ProSe extensions to the on-network 5G NR models.

REFERENCES

[1] 3GPP, "Proximity based Services (ProSe) in the 5G System (5GS) Release 17," 3rd Generation Partnership Project (3GPP), Technical Specification (TS) 23.304, 09 2023, v17.8.0.

[2] Z. Ali, S. Lagén *et al.*, "3GPP NR V2X mode 2: overview, models and system-level evaluation," *IEEE Access*, vol. 9, pp. 89 554–89 579, 2021.

[3] ——, "On the impact of numerology in NR V2X mode 2 with sensing and random resource selection," in *2021 IEEE Vehicular Networking Conference (VNC)*, 2021, pp. 151–157.

[4] W. Garey, T. R. Henderson *et al.*, "Modeling MCPTT and user behavior in ns-3," in *11th International Conference on Simulation and Modeling Methodologies, Technologies and Applications, SIMULTECH 2021, Online*, 2021, pp. 30–41.

[5] P. Liu, C. Shen *et al.*, "5G New Radio sidelink link-level simulator and performance analysis," in *25th International ACM Conference on Modeling Analysis and Simulation of Wireless and Mobile Systems*, Montreal, 2022, p. 75–84.

[6] 3GPP, "Medium Access Control (MAC) protocol specification Release 17," 3rd Generation Partnership Project (3GPP), Technical Specification (TS) 38.321, 09 2023, v17.6.0.

[7] ——, "Mission Critical Push To Talk (MCPTT) call control; protocol specification," 3rd Generation Partnership Project (3GPP), Technical Specification (TS) 24.379, 12 2017, v14.4.0.

[8] ——, "Mission Critical Push To Talk (MCPTT) media plane control; protocol specification," 3rd Generation Partnership Project (3GPP), Technical Specification (TS) 24.380, 09 2017, v14.4.0.

[9] ——, "Mission Critical Push To Talk (MCPTT); stage 1," 3rd Generation Partnership Project (3GPP), Technical Specification (TS) 22.179, 03 2019, v16.5.0.