System-Level Evaluation of 5G NR UE-Based Relays

Samantha Gamboa[†‡], Aziza Ben Mosbah[†‡], Wesley Garey[*], Chunmei Liu[*], Richard Rouil[*]
[*]*Wireless Networks Division, National Institute of Standards and Technology*, Gaithersburg, MD, United States
[†]*Associate, Wireless Networks Division, National Institute of Standards and Technology*, Gaithersburg, MD, United States
[‡]*Prometheus Computing LLC*, Bethesda, MD, United States
{samantha.gamboa, aziza.benmosbah, wesley.garey, chunmei.liu, richard.rouil}@nist.gov

***Abstract*—The notion of User Equipment (UE)-based relays started with the introduction of Proximity Services (ProSe) and Device-to-Device (D2D) direct communication over the Sidelink (SL), and evolved with the specification of the Fifth Generation (5G) New Radio (NR) interface. While the main goal of UE-based relays is to extend coverage, many parameters play a role in determining its effective range and benefits. In this paper, we study a military use case involving UE-to-Network (U2N) relays providing connectivity to a squad of soldiers with limited coverage. We configure advanced NR SL functionalities that include diverse numerologies, sensing-based resource selection, and blind retransmissions. We evaluate the end-to-end network performance in terms of Packet Delivery Ratio (PDR) and latency. We show, through system-level simulations, that fine-tuning the different NR SL parameters is critical to achieve the expected performance in scenarios where the relay UE is serving a variable number of UEs.**

***Index Terms*—Mobile Networks, Modeling and Simulation, Fifth Generation (5G), New Radio (NR), Sidelink (SL), Proximity Services (ProSe), UE-to-Network (U2N) relay**

I. Introduction

Proximity Services (ProSe) emerged as the initial technology enabling Device-to-Device (D2D) direct communication between User Equipments (UEs) in Third Generation Partnership Project (3GPP) Long Term Evolution (LTE) mobile networks [1]. LTE ProSe was introduced in Release 12, leading to the development of the LTE Sidelink (SL) which is the interface that allows UEs to communicate directly without using the network infrastructure. In Release 16, the SL interface was adapted to New Radio (NR), the new air interface developed by 3GPP for Fifth Generation (5G) mobile networks. Then, Release 17 introduced the first NR ProSe functionalities, with direct discovery, direct communication, and UE-to-Network (U2N) relays. With U2N relays, an in-network UE acts as a relay for UEs in close proximity with limited or no coverage from a 5G Evolved Node B (gNB). This allows the relay UE to extend the network reach by relaying network data to and from the remote UEs using direct communication over SL. Currently, another UE-based technology, the UE-to-UE (U2U) relay, is in development for Release 18 among other enhancements to the U2N relay features [2]. Release 18 specifications focus on single-hop U2U relaying, providing connection between a source and a target UE with only one U2U relay UE between them. However, the 3GPP work item states that specifications should take into account forward compatibility to support multi-hop in later releases [3].

The UE-based relay technologies are key to support mission critical operations for first responders and tactical military campaigns. The ability to extend network coverage to remote areas by using U2N relays is essential for situational awareness as well as efficient and prompt command delivery for missions in remote areas where network infrastructure is lacking. Moreover, U2U relays allow intra-squad connectivity in sparse and challenging terrains. However, current SL specifications only consider one-hop U2U relays, whereas with multi-hop the opportunities to greatly extend network coverage and support more complex use cases such as subterranean communication, human-machine integration, and unmanned aerial systems, have led to work items that call for 3GPP to support it [4].

Two U2N relay architectures are supported: Layer 3 (L3) and Layer 2 (L2) relaying. In the L3 U2N relay architecture, data forwarding occurs at the network layer and remote UEs connected to the relay UE are transparent to the Next Generation Radio Access Network (NG-RAN) of the U2N relay UE. In the L2 U2N relay architecture, remote UEs must first connect to the relay UE and then establish a connection with the NG-RAN via the relay UE to be considered in-network. Both technologies are designed to extend network coverage using direct communication over SL, but provide different levels of access control and implementation complexity. However, given that remote UEs can be out-of-coverage, SL resource allocation is performed in a distributed manner for the U2N relay and remote UEs based on preconfiguration parameters (Mode 2), regardless of the architecture. This also applies to the U2U relay functionality where source, target, and U2U relay UEs will use direct communication over the shared SL for each hop. The NR SL has an extensive list of parameters that need to be preconfigured and they directly affect the end-to-end performance of the system. In this paper, we focus on a subset of these parameters, such as resource pool configuration, sensing-based resource selection, and blind Hybrid Automatic Repeat Request (HARQ) transmissions to study the expected performance for a given configuration and the corresponding trade-offs. To the best of our knowledge, this is the first end-to-end system-level study that considers 3GPP's 5G NR U2N relays.

Our main contributions are summarized as follows:

- Elaborate on the complexity of the SL preconfiguration, since some input parameters (e.g., the numerology, resource selection window, and number of transmissions) are crucial to the overall system performance
- Utilize system-level simulations to evaluate the performance of a system using the U2N relay functionality in a tactical military scenario where scout soldiers are deployed in an area with limited coverage
- Identify the performance trade-offs associated with a given SL configuration within the context of the 1 to N communication that the U2N relay architecture imposes

The paper is organized as follows: In Section II, we look at related work in the literature to examine current and emerging solutions. In Section III, we describe the NR SL features considered in this paper. Then, the impact of these features on the network performance will be analyzed in Section IV using a military use case. Finally, in Section V, we conclude our paper and discuss future work.

II. Related Work

In the literature, a vast number of papers are written on 5G NR, with many focusing on emerging Cellular Vehicle-to-Everything (V2X) (C-V2X) solutions that may use the NR SL for communication. In [5], the authors provide an overview of the 3GPP study and work items for 5G NR SL, which promise performance improvements for direct mode communication in comparison with LTE systems.

The authors from [6] explore the impact of the numerology on the end-to-end latency, proving that an increase in numerology does not necessarily mean a decrease in latency, and that the resulting performance highly depends on the traffic pattern considered. While this is an interesting conclusion, the paper focuses on the Uplink (UL) and Downlink (DL) flows, but not on the direct communication between UEs. With U2N relays, in the addition to the UL and DL , the SL is used, which adds to the system complexity.

The authors in [7] introduce a Network Simulator 3 (ns-3)-based, NR V2X, system-level simulator based on Release 16 and the NR V2X standardization efforts at that time. The implementation covers the baseline NR SL and V2X broadcast communications for out-of-coverage scenarios. The study performed in this paper is based on a model that builds on their simulator to include more advanced resource scheduling and a more standard-compliant sensing algorithm. Other additions include an NR ProSe layer to support direct discovery, direct unicast communication, and the L3 U2N relay used for the evaluation in this paper.

The authors in [8] explore the impact of the numerology and sensing in the packet inter-reception delay in an NR V2X highway scenario. We perform a similar analysis in this paper but considering the ProSe U2N relay architecture, whose 1 to N unicast connection characteristics result in different and much higher utilization of the NR SL than the NR V2X scenarios.

The authors in [9] propose and evaluate a blind retransmission-based scheme. For smaller inter-vehicle distances and/or lower speeds, retransmissions burden an already fully loaded system. The scenario only focuses on LTE C-V2X Mode 3, which is only available for vehicles under cellular coverage and where the resource allocation is performed by the gNB, while in this work we consider NR SL Mode 2 where resource allocation is UE-selected and channel sensing is an option.

In [10], the authors present a high-level, standardization-based overview of the architectures and protocol stacks for both 3GPP Release 17 L2 relay and L3 relay. This also includes details regarding their related discovery and (re)selection features. Yet, no performance study is conducted.

Since the NR SL (Releases 16 and 17) does not support multi-hop at the lower layers in the LTE stack, Chukhno et al. consider a static scenario with pre-defined sequential hops between UEs, and conduct a preliminary simulation study for public safety and factory automation applications [11]. They conclude that, compared to its LTE Advanced predecessor, the NR ProSe technology allows for higher reliability and throughput, along with lower power consumption and latency thanks to SL relay capabilities. In [12], Narayanan et al. study multiple enhancements to Release 16's L3 relay-related procedures to effectively support massive Internet of Things (IoT) applications. They are also able to maintain roughly the same success probability while reducing the energy consumption and signaling overhead for an IoT traffic scenario. Finally, in [13], Kawamatsu et al. suggest a new single-hop routing method. This method is based on the use of gNBs and Road Side Units (RSUs) as relays for vehicular communications. The idea is that these two entities can be used to maintain a balance between traveled distance and link quality.

Despite the benefits stated above, none of these research efforts evaluate how coverage can be extended using a standards-defined relay service, even though this is what network operators and device vendors are more likely to implement and use. In this paper, we evaluate the standardized 5G NR single-hop L3 U2N relays, with a focus on how various 3GPP-compliant, NR SL parameters and features, such as the numerology, sensing-based resource selection, and blind retransmissions affect network performance.

III. 5G NR SL

In this section, we provide a general description for each of the 5G NR and SL features used in this evaluation study. This includes an overview of L3 U2N relays, SL parameters and features, as well as device limitations and operations.

A. Layer 3 (L3) UE-to-Network (U2N) Relay

As illustrated in Fig. 1, a ProSe L3 U2N relay UE provides indirect connectivity to the 5G network for remote UEs that may be outside of NG-RAN coverage. When the relay UE receives traffic in the DL from the 5G core that is designated for one of the remote UEs it is serving, it forwards that unicast traffic over the NR SL to the designated UE. When the relay UE receives traffic from a remote UE that should be relayed to the 5G core, it does so by forwarding that traffic in the UL [14] [15].

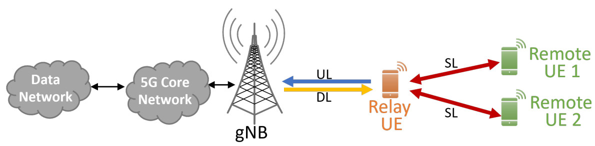


Fig. 1: Architecture of NR ProSe L3 U2N Relay.

Before traffic can be relayed, a series of procedures are performed by the relay and remote UEs in order to establish a direct link between them. The remote UE performs the relay discovery procedure to find an available relay UE in close proximity. Once a suitable relay UE is discovered and selected, a request to establish a direct connection for unicast communication is sent by the remote UE. If it is positively acknowledged by the relay UE and successfully received by the remote UE, then the SL direct link is established and can now be used to transfer data.

All the exchanges between the relay UE and the remote UE described above, including the traffic relaying, are done over the NR SL using a pool of shared radio resources. In the following text we describe the parameters affecting the dimensions of this pool and how resources are selected for traffic transmission.

B. Numerology

In 5G NR, the numerology (μ) refers to the physical waveform characteristics in terms of Subcarrier Spacing (SCS) and the timing of Time-Division Duplex (TDD) resources. Table I contains the supported numerologies for Band n47 (5.855 GHz to 5.925 GHz), which is designated for NR V2X usage, where devices are authorized to use the NR SL. As shown in the table, the higher the μ, the higher the SCS and the smaller the slot duration [16].

In the time domain, a radio frame, which has a duration of 10 ms, is divided equally into 10 subframes. In turn, each subframe, which has a duration of 1 ms, is divided into a variable number of slots, depending on the chosen numerology. In the frequency domain, the bandwidth is partitioned into sub-channels, each consisting of a fixed number of Physical Resource Blocks (PRBs). Each PRB is composed of 12 subcarriers, which, given the different SCSs, results in a variable frequency size depending on the numerology employed. This leads to variations in the number of sub-channels available for a given bandwidth across different numerologies.

TABLE I: Supported Numerology Values for Band n47.

Numerology (μ)	**Subcarrier Spacing (SCS)** ($2^{\mu} \times 15$ **kHz**)	**Slots per subframe** (2^{μ})	**Slot length** ($1/2^{\mu}$ **ms**)
0	15 kHz	1	1 ms
1	30 kHz	2	0.5 ms
2	60 kHz	4	0.25 ms

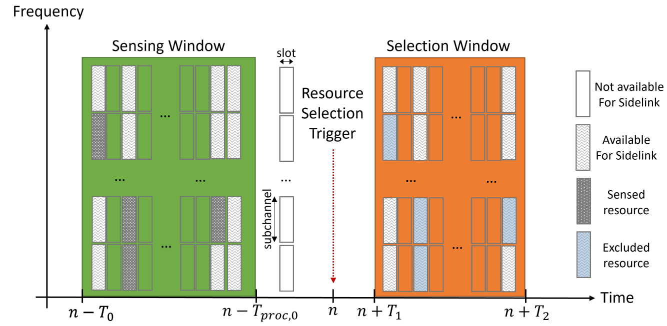


Fig. 2: Sensing and Selection Windows.

C. Resource Selection and Sensing

The resource pool allocated for the SL consists of time resources (i.e., slots) and frequency resources (i.e., sub-channels). An SL resource consists of a slot in time and a number of contiguous sub-channels in frequency, as illustrated in Fig. 2. The available slots for SL are indicated by the TDD pattern and the SL bitmap, while the number of sub-channels considered for a transmission are calculated based on the size of the payload to transfer, the associated control messages, as well as the selected Modulation and Coding Scheme (MCS).

As shown in Fig. 2, when a resource selection is triggered at time n, the UE should select resources for its transmissions within the resource pool corresponding to the selection window. If sensing is enabled, the resource selection phase may be preceded by the sensing-based exclusion phase, where resources to exclude from the available candidate resources are identified. A resource may be excluded if another transmitting UE picked it for a future transmission. Such reservations are detected during the sensing window. The receiving UE measures the Reference Signal Received Power (RSRP) associated with the control message that announce the future transmission(s) (the first-stage SL Control Information (SCI) message) and stores this information for use during the resource selection triggering [17].

If Semi-Persistent Scheduling (SPS) is used, the Resource Reservation Interval (RRI) parameter is also communicated in the first-stage SCI message. It establishes the periodicity of the SPS transmissions, and is used to project the transmissions detected during the sensing window in time so that the resources corresponding to projections within the sensing window are excluded from selection. The sensed resources are excluded only if the measured RSRP is above a predefined threshold. The UE may also exclude a projection of resources if it was transmitting during the sensing window. This is done to conservatively account for transmissions that may have not been detected due to the half-duplex mode of operation. There is a minimum percentage of resources that should remain for selection. If the percentage is not met, the exclusion is rolled-back and started again with a lower RSRP threshold. If the percentage cannot be met with any threshold, no resources are excluded.

As shown in Fig. 2, the sensing window is defined by T_0

and $T_{proc,0}$, while the selection window is defined by T_1 and T_2. The UE will use the information sensed during the period $[n - T_0, n - T_{proc,0}]$ to compute the resource exclusions. The selection window includes all available SL resources within the range of slots $[n + T_1, n + T_2]$, where both T_1 and T_2 are expressed in slots, and their equivalent place in time depends on the numerology. Parameters $T_{proc,0}$ and T_1 represent guard periods to account for any required processing. The 3GPP standard states that T_2 should be selected so that is less than or equal to the Packet Delay Budget (PDB) of the traffic to be served.

The resources selected within a selection window are randomly selected from the SL resources remaining after the sensing-based exclusion, if sensing is enabled. Following the selection of resources in a selection window, if SPS is enabled, the selected resources are projected into the future according to the RRI and reserved until a new selection is triggered.

D. Half-Duplex

UEs use the same transceiver to transmit and receive data over SL. This prevents them from transmitting and receiving data simultaneously due to self-interference. Thus, half-duplex operation refers to the mode of operation in which a UE can either transmit or receive in a given time slot, but not both [16]. Given the distributed nature of resource selection in the SL when using Mode 2, half-duplex operation is an important performance limiting factor when bidirectional communication is expected between UEs, as we will show in Section IV.

E. Scheduling Type

In 5G NR SL, two scheduling types are supported: SPS and Per-Protocol Data Unit (PDU), i.e., dynamic scheduling [18]. With SPS, a UE selects resources once and consecutively repeats this selection in the future for a number of times based on the reselection counter. The interval at which resources are reserved for a Transport Block (TB) is separated by the RRI in time. This type of scheduling is more suited for Constant Bitrate (CBR) traffic with a predefined packet inter-arrival time. However, when a UE operates using dynamic scheduling, it selects new SL resources for each TB and its associated blind or feedback-based retransmissions, if enabled. Dynamic mode is more appropriate for non-CBR traffic.

F. HARQ Retransmissions

In order to increase the TB reception probability, 5G NR SL supports both blind-based and feedback-based HARQ. Both types of HARQ are configurable and support up to 32 transmissions of the same TB (nTx) [19]. Unlike feedback-based retransmissions, blind-based retransmissions do not rely on any positive or negative acknowledgements (ACK/NACK). The transmitting UE keeps sending the same TB for a known number of repetitions, independently of whether or not the receiving UE successfully receives and decodes the transmitted TB. While feedback-based HARQ can suppress unnecessary retransmissions and prevent capacity drain, it may also lead to higher latency and increased resource utilization to accommodate feedback signaling.

IV. System-level evaluation

In this section, we describe the system-level evaluation we use to illustrate the performance trade-offs considering a scenario where the U2N relay functionality is enabled to provide network access to a squad of soldiers. We use a ProSe-ready, system-level simulator that builds on the model described in [7]. This simulator is based on ns-3 and provides a baseline NR SL implementation that now supports more advanced resource scheduling, a more standard-compliant sensing algorithm, and the NR ProSe layer due to our efforts. This includes an implementation of L3 U2N relays, which is used in this evaluation.

A. Scenario

We consider a squad composed of ten notional soldiers [20]. One soldier is stationed inside of the squad's vehicle while the other soldiers are deployed to scout the area. All soldiers carry ProSe-enabled devices and all devices have the same characteristics. We also assume that the vehicle is tactically located in an area, such that the UE of the soldier inside of the vehicle is in-coverage and directly connected to the network via the gNB over the UL and DL channels. However, the scout soldiers, regardless of whether or not they are within the gNB's coverage range, indirectly connect to the network via the relay UE over the SL channel. This is because the vehicle will remain stationary, and thus, assumed to always remain in-coverage, while the scout soldiers will be mobile and possibly move into areas that are out-of-coverage. Thus, the soldier in the vehicle enables the U2N relay functionality and that device acts as the U2N relay UE, providing connectivity to the scout soldiers, while the scout soldier devices then take on the role of remote UEs in the U2N relay architecture. In the simulations, we assume that the scout soldiers are randomly deployed within a circle that has a radius, r, that is centered on the squad's vehicle, as shown in Fig. 3. We then randomly select *nSoldiers* scout soldiers to be actively communicating throughout the simulation. In the next section, we show results for different values of r and *nSoldiers* within the ranges listed in Table II.

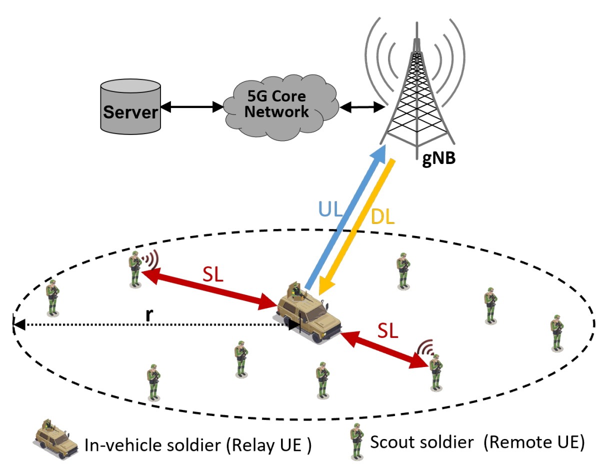


Fig. 3: Evaluation scenario example with nSoldiers = 2.

We use Band n47 and the simulator makes use of two Bandwidth Parts (BWPs): one is used for in-network communication (relay UE to/from gNB) and the other is used for SL communication (relay UE to/from remote UEs). We set up the simulation so that the relay connection with the gNB is sufficient to support the relayed traffic without any losses, and so that the scheduling process is the only delay that is incurred. This scheduling delay is less than 4 ms in the UL and less than 2 ms in the DL for this particular scenario.

Scout soldier devices may have one or two simultaneous traffic flows. This includes traffic from the device to a server in the network (UL traffic flow) and from the server to the device (DL traffic flow). When there are active traffic flows in both the UL and DL, we refer to this as bidirectional traffic. Each traffic flow is configured with the parameters listed in Table II, and we configure the system to be able to sustain this traffic with an SPS RRI equal to the packet inter-arrival rate (20 ms) and a selection window size (16 ms) smaller than the PDB (20 ms).

We use the following Key Performance Indicators (KPIs) to represent the system performance:

- Packet Delivery Ratio (PDR) – a ratio that represents the number of received packets over the number of transmitted packets at the application layer
- Average Packet Delay – the mean duration for all received packets from the moment a packet is transmitted to the moment it is received at the application layer

In the results, we specify which flows are used to calculate the KPIs: "UL" for all UL flows, "DL" for all DL flows, and "Sys" for all flows in the system when there is bidirectional traffic. It is worth mentioning here that random variables affect many aspects of our simulation, including soldier positions, traffic flow start times, packet reception events, and resource selection. While a single trial would yield variations during resource selection and packet receptions, initial conditions, such as soldier positions and flow start times would be fixed. Thus, we choose to average the data collected from a large number (200) of short (10 s), independent trials rather than a single, long trial since varying initial conditions can also affect performance. The results that we present include the mean and 95 % confidence interval for each KPI. Refer to Table II for other relevant simulation parameters.

TABLE II: Simulation Parameters.

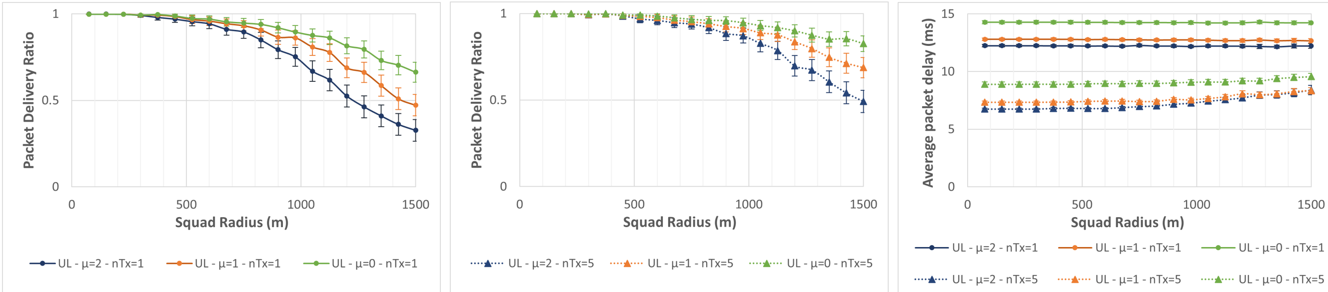
Deployment Parameters	
Squad radius (r)	[75 m , 1500 m]
Number of in-vehicle soldiers	1
Number of scouts soldiers	9
Number of communicating scouts soldiers (nSoldiers)	1, 5, 9
System Parameters	
Channel model	3GPP Spatial Channel, Rural Macro
Central frequency	5.89 GHz (band n47)
Bandwidth	80 MHz (40 MHz for UL/DL, 40 MHz for SL)
UE Transmit power	23 dBm[1]
UE antenna height	1.5 m
Numerology (μ)	0, 1, 2
Subchannel size	10 RBs
Number of subchannels	20 (μ=0), 10 (μ=1), 4 (μ=2)
TDD pattern	DL|DL|DL|F|UL|UL|UL|UL|UL|UL
SL parameters	
Sidelink bitmap	{1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1}
PSCCH MCS index	0 (fixed)
PSSCH MCS index	5 (fixed)
HARQ retransmission scheme	Blind-based retransmissions
Number of transmissions (nTx)	1, 2, 5
Sensing	Enabled, disabled
Sensing window (T_0)	100 ms
Min. percentage of resources	20%
Scheduling algorithm	SPS with RRI = 20 ms
Guard parameters	$T_{proc,0}$ = 2 slots, T_1 = 2 slots
Selection window size (time)	16 ms
Selection window size (slots) ($T_2 - T_1$)	16 (μ=0), 32 (μ=1), 64 (μ=2)
Traffic parameters	
Traffic pattern	Constant Bit Rate (CBR)
Packet Size	60 Bytes
Packet Inter arrival interval	20 ms
Data rate	24.0 kb/s
Packet delay budget	20 ms
Direction	UL, DL, Bidirectional

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

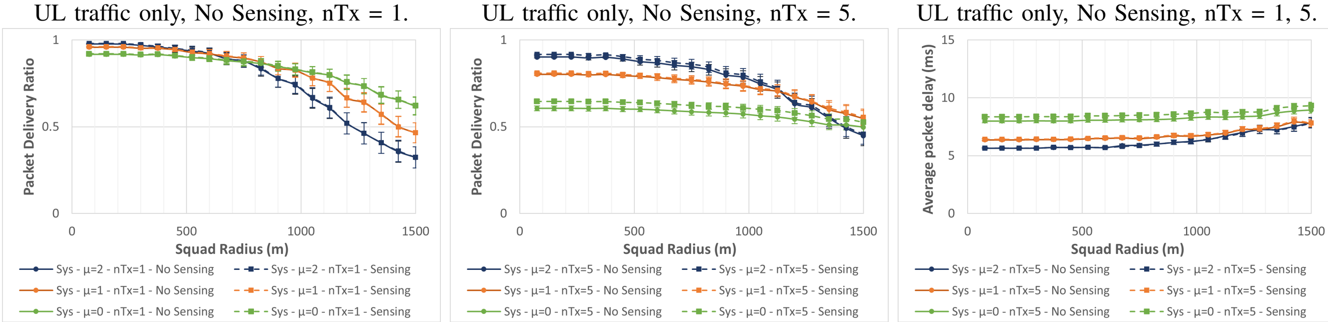
B. Results

Fig. 4 shows the simulation results for the case when only one scout soldier is actively communicating over the course of the simulation (nSoldier = 1) and Fig. 5 shows the results for when there are multiple active scout soldiers in the system (nSoldier = 5 and nSoldier = 9). Fig. 4a, 4b, and 4c show the results for the case when there is only UL traffic, but we observe the same trends for the case when there is only DL traffic with nSoldier = 1. Fig. 4a shows the PDR in the case where no HARQ retransmissions of the packets are performed (nTx = 1). We observe a PDR close to 1 for squad radii up to r = 450 m, however, the PDR decreases as r increases beyond that. For reference, a PDR of 1 would mean that all packets are received successfully, while a PDR of 0 would indicate that no packet is received successfully. In this single transmitter scenario, this degradation is the result of the signal attenuation over distance mainly due to path loss. We also observe different performance depending on the numerology, μ, with smaller values of μ yielding a better PDR for larger values of r. For example, for r = 1500 m, the PDR is about 0.32 with μ = 2, 0.47 with μ = 1, and 0.67 with μ = 0. A larger μ requires a larger SCS which means that the power is more spread out over the bandwidth during a transmission. This ultimately leads to a weaker signal and is why we see a decrease in performance for larger μ as r is increased.

Fig. 4b shows the PDR with nTx = 5 and we observe that enabling retransmissions to perform HARQ combining procedures leads to performance increases for all values of μ. Moreover, we observe improved performance for larger values of r compared to the results with nTx = 1. For example, for r = 1500 m, the PDR is about 0.49 with μ = 2 (16 % improvement), 0.68 with μ = 1 (21 % improvement), and 0.83 with μ = 0 (16 % improvement). However, the use of retransmissions comes with increased resource utilization, which, in this single transmitter case, does not affect the



(a) Packet delivery ratio. UL traffic only, No Sensing, nTx = 1.
(b) Packet delivery ratio. UL traffic only, No Sensing, nTx = 5.
(c) Average packet delay. UL traffic only, No Sensing, nTx = 1, 5.



(d) Packet delivery ratio. Bidirectional traffic, nTx = 1.
(e) Packet delivery ratio. Bidirectional traffic, nTx = 5.
(f) Average packet delay. Bidirectional traffic, nTx = 5.

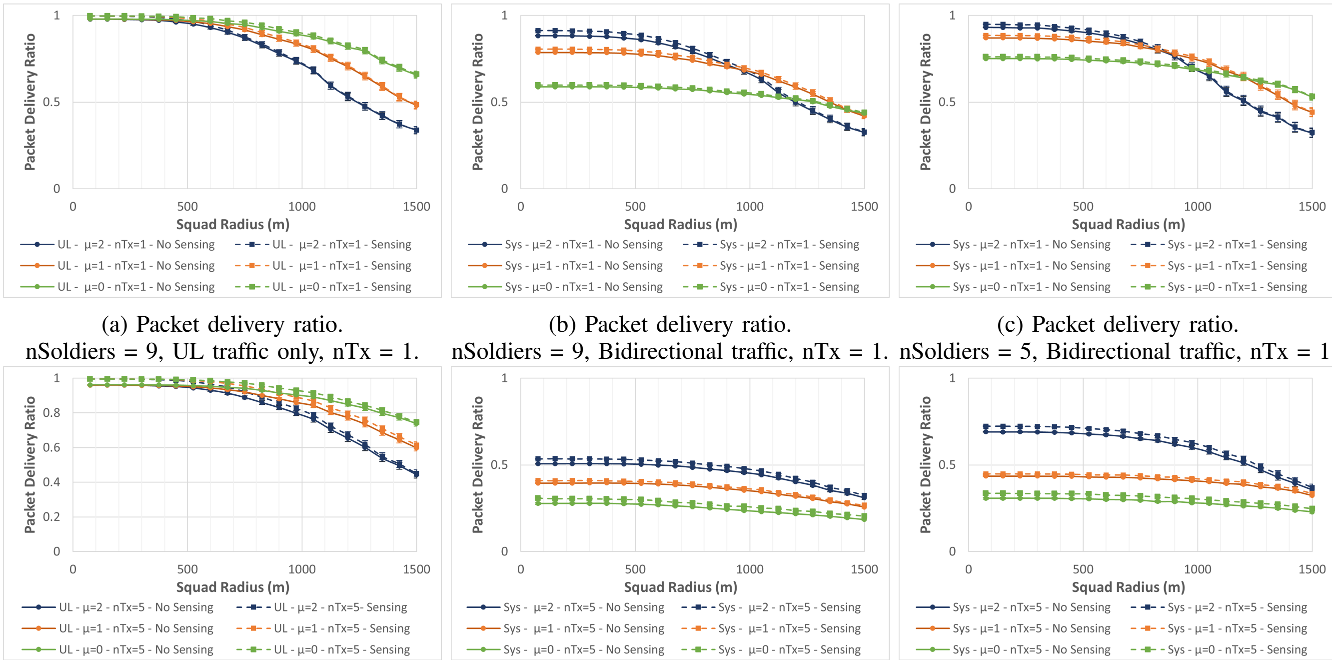
Fig. 4: System-level simulation results. One active soldier in the squad (nSoldiers = 1).

performance but becomes critical when the SL is shared, as we will see later in this section.

Fig. 4c shows that we have smaller average packet delays with larger values of μ. This is due to the different distribution of time resources (slots) in the resource selection window. A larger value of μ results in a shorter slot duration, which, in turn, results in an increased number of slots that are available within the selection window. It also decreases the amount of time that it takes for a resource to become available. For example, when $\mu = 0$, the first SL slot is not available until 3 ms after the first DL slot, but when $\mu = 2$ the first SL slot is available after 0.75 ms. Moreover, with a larger number of transmissions to select resources for (e.g., nTx = 5), the probability of the resource for the first transmission happening sooner also increases with a higher μ. When nTx = 5 and as r increases, decoding errors, due to signal attenuation, also increase. This leads to a need for more retransmissions to decode the packet and results in an increase in delay at the application layer. This can be observed with large values of r in Fig. 4c. For example, with μ = 2 and nTx = 5, the delay is around 6.81 ms when r= 450 m and goes up to 8.39 ms when r= 1500 m. These average packet delay trends persist throughout all of the considered scenarios regardless of the number of active users.

Fig. 4d, 4e, and 4f show results when nSoldier = 1 but with bidirectional traffic. The KPIs of both UL flows and DL flows show the same trend and the aggregated performance is what is depicted in the figures. In this case, both the relay UE and the active remote UE share the SL resources for their transmissions, and we show results with sensing-based resource selection enabled as well. In Fig. 4d and 4e, we can see that the highest PDR the system achieves is 0.97. This is mainly due to the half-duplex mode of operation. When the relay UE and the remote UE transmit at the same time, they miss each other's transmissions which results in packet loss. This resulting loss is either due to missing a data packet sent on the shared channel or missing the SCI message sent on the control channel. If the UE misses the SCI message on the control channel, then this will lead to future losses on the shared channel simply because the UE did not receive the control information that indicates when and where upcoming data in the shared channel will be transmitted. Moreover, SPS makes the half-duplex limitation persistent in time until a reselection is triggered. This loss in performance is accentuated with a lower μ, as the resource pool is more constrained in time resources. As r increases, we see the trend reverses as the signal attenuation adds to the drop in performance. For example, in Fig. 4d, we see that μ = 1 provides a higher PDR than μ = 2 starting at r = 825 m, and μ = 0 provides a higher PDR than μ = 1 starting at r = 1050 m. We also observe in Fig. 4e that enabling the sensing-based resource selection helps improve the performance, especially in the time-limited resource pools of lower μ values, where we see an improvement of up to 5 % in the PDR when μ = 0. However, this improvement may be accompanied by a larger delay as shown in Fig. 4f, where we observe up to 0.42 ms of delay increase with μ=0 and sensing.

Fig. 5a and 5d show the results for the scenario in which



(a) Packet delivery ratio. nSoldiers = 9, UL traffic only, nTx = 1.

(b) Packet delivery ratio. nSoldiers = 9, Bidirectional traffic, nTx = 1.

(c) Packet delivery ratio. nSoldiers = 5, Bidirectional traffic, nTx = 1.

(d) Packet delivery ratio. nSoldiers = 9, UL traffic only, nTx = 5.

(e) Packet delivery ratio. nSoldiers = 9, Bidirectional traffic, nTx = 5.

(f) Packet delivery ratio. nSoldiers = 5, Bidirectional traffic, nTx = 5.

Fig. 5: System-level simulation results. Multiple active soldiers in the squad (nSoldiers = 9, 5).

all scout soldiers have active traffic flows (nSoldier = 9) in the UL direction only. This means that the SL is shared between the nine remote UEs for their transmissions towards the relay UE. When sensing is not used, we see that the PDR is below 0.96 for all μ in the scenario with nTx = 5. This is due to collisions between the transmissions of different remote UEs using the same resources, which is accentuated by the increase in the total number of transmissions when nTx = 5. This causes interference at the receiver side and potential packet losses. However, sensing-based resource selection helps to mitigate this problem and we see better performance when it is used, with a PDR > 0.99 for values of r up to 375 m, 450 m, and 525 m when $\mu = 2$, $\mu = 1$, and $\mu = 0$, respectively. We also simulate the case with DL traffic only and we find the same trends as with nSoldier = 1, as the relay UE is the only transmitter and there are enough resources to satisfy all of the traffic that goes towards the nine remote UEs.

When we have bidirectional traffic for all scout soldiers in the squad, a maximum PDR of 0.89 is achieved, as shown in Fig. 5b with $\mu = 2$, nTx = 1, and r =75 m. We see the same crossover points in performance that we see with one active UE, where the effect of the half-duplex mode is more pronounced for smaller r with lower μ values until the signal attenuation becomes more prevalent with larger r and accelerates the drop for higher values of μ. The loss in performance is significant, as the relay UE needs to transmit the DL traffic to all remote UEs. Thus, due to the half-duplex mode, the relay UE will lose a significant number of transmissions from the remote UEs for the traffic flows in the UL direction. Fig. 5e shows that increasing the number of retransmissions further degrades the PDR with no benefit for any of the values of r that are considered. Fig. 5a and 5f show similar trends for the case when nSoldier = 5 with bidirectional traffic. However, since there is less demand on the system it does a better job of sustaining the traffic and network performance is improved.

C. Discussion

In the previous section, when there is only one active soldier (baseline case) with UL traffic, we see that increasing the numerology decreases the packet delay. However, the PDR decreases more rapidly as the radius increases since a higher numerology leads to a weaker signal. At the same time, increasing the number of blind HARQ retransmissions for each numerology can provide a boost in performance for larger radii since this increases message redundancy. However increased redundancy requires more resources. When there is traffic in both directions (UL and DL), we see that a larger numerology provides better performance for a smaller squad radius since the increase in numerology provides more resources (i.e., slots) in a capacity-limited situation. However, the performance of a larger numerology decreases more rapidly than that of a smaller numerology when the radius is increased, similar to what we see in the UL only case. Unlike the UL-only case, when there is traffic in both directions, we also observe a slight improvement in the PDR when sensing is enabled, since the relay and remote UEs share the same resource pool.

Expanding on the baseline case, when increasing the number of active soldiers to five and nine, these trends also hold true.

However, the overall performance shows that the increase in traffic results in a lower PDR for all squad radii. We also observe a slight improvement in the overall performance for each case with multiple active users when sensing is enabled.

A notable finding in our study is the significance of the half-duplex mode of operation. Whenever there is more than one UE using the same resource pool, due to the randomness of the resource selection algorithm in Mode 2, it is always possible for multiple UEs to transmit at the same time, and thus, not receive each other's messages. Furthermore, in each scenario with bidirectional traffic or multiple active users the negative effect of the half-duplex mode of operation is present. Hence, what we see is that the half-duplex effect occurs regularly and increases with resource contention in the time domain, regardless of whether or not sensing is enabled. Therefore, in the case of ProSe UE-based relays, one should always be mindful of the number of time resources available between the relay UE and all other UEs with the same resource pool. This is because no matter how much bandwidth is available to support UE traffic, the likeliness of a relay UE and any other UE selecting the same resource in time may have a significant impact on the system's performance. This is paramount when considering U2U relays for both single-hop and multi-hop use cases as the U2U relay(s) use the SL for both reception and forwarding of relayed traffic, and each hop of the relaying chain is impacted by the effects of the half-duplex mode of operation mentioned above.

V. Conclusion

In this paper, an evaluation of 5G NR SL features is performed to demonstrate their impact on the network performance of 3GPP's L3 U2N ProSe relay service when it is leveraged to enable network access to a military squad deployed with units that have limited coverage. This includes an overview of the 5G NR SL system and ProSe functionalities that we later use to evaluate key features that include sensing-based resource allocation, HARQ blind retransmissions, and numerology settings.

There are several avenues that we would like to explore to further this research. The two main goals of our future work include both extending coverage and increasing network performance. The most interesting path involves building upon our current simulation model to support U2U relay so that we can evaluate how a multi-hop deployment can extend coverage. A promising feature that we plan to explore in the future is feedback-based HARQ. Based on our analysis of blind retransmissions, feedback-based HARQ has the potential to reduce the overall traffic load since retransmissions should only occur when they are necessary. We would also like to investigate more complex deployments that consider several relays per squad, as well as multi-squad scenarios. This would allow us to gain further insights on the applicability of 5G NR SL features, understand what can be gained from emerging technologies in terms of network coverage and reliability, as well as provide useful insights that have the potential to directly impact the development of 3GPP standards.

Acknowledgment

The authors express their sincere gratitude to Thomas R. Henderson from the University of Washington for the insightful technical discussions and significant involvement in the development of the system-level simulator used in this study.

References

[1] 3GPP, "Feasibility study for Proximity Services (ProSe)," 3rd Generation Partnership Project (3GPP), Technical Report (TR) 22.803, 06 2013, version 12.2.0.

[2] ——, "Study on system enhancement for Proximity based Services (ProSe) in the 5G System (5GS)," 3rd Generation Partnership Project (3GPP), Technical Report (TR) 23.700-33, 12 2022, version 18.0.0.

[3] ——, "Study on Proximity-based Services in 5GS Phase 2," 3rd Generation Partnership Project (3GPP), Work Item Description SP-211653, 12 2021.

[4] ——, "Introducing additional mission-critical use cases for 5G Sidelink, technology gaps and proposed enhancements," 3rd Generation Partnership Project (3GPP), TDoc RP-223388, 12 2022.

[5] F. Cintron, D. Griffith, C. Liu *et al.*, "Study of 5G New Radio (NR) Support for Direct Mode Communications," *NIST, Tech. Rep.*, 2021.

[6] N. Patriciello, S. Lagen, L. Giupponi *et al.*, "5G New Radio Numerologies and their Impact on the End-To-End Latency," in *2018 IEEE 23rd International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD)*, 2018, pp. 1–6.

[7] Z. Ali, S. Lagén, L. Giupponi *et al.*, "3GPP NR V2X Mode 2: Overview, Models and System-Level Evaluation," *IEEE Access*, vol. 9, pp. 89 554–89 579, 2021.

[8] Z. Ali, S. Lagén, and L. Giupponi, "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.

[9] D. Wang, R. R. Sattiraju, A. Qiu *et al.*, "Effect of Retransmissions on the Performance of C-V2X Communication for 5G," in *2020 IEEE 92nd Vehicular Technology Conference (VTC2020-Fall)*, 2020, pp. 1–7.

[10] B. Zhang, Q. Lu, C. Shi *et al.*, "Sidelink Relay in NR and NR Advanced," in *2021 IEEE/CIC International Conference on Communications in China (ICCC Workshops)*, 2021, pp. 358–362.

[11] N. Chukhno, A. Orsino, J. Torsner *et al.*, "5G NR Sidelink Multi-Hop Transmission in Public Safety and Factory Automation Scenarios," in *IEEE Network*, 2023, pp. 1–7.

[12] S. Narayanan, O. Liberg, A. Höglund *et al.*, "Sidelink Optimizations for Layer-3-Based IoT Relaying in 5G NR," in *IEEE Internet of Things Magazine*, vol. 5, no. 2, 2022, pp. 140–145.

[13] D. Kawamatsu, K. Mizutani, and H. Harada, "Highly Efficient Relay Routing Method for 5G Cellular V2V Communications," in *2022 IEEE Conference on Standards for Communications and Networking (CSCN)*, 2022, pp. 218–223.

[14] 3GPP, "Study on system enhancement for Proximity based Services (ProSe) in the 5G System (5GS)," 3rd Generation Partnership Project (3GPP), Technical Report (TR) 23.752, 03 2021, version 17.0.0.

[15] ——, "Technical Specification Group Services and System Aspects; Proximity based Services (ProSe) in the 5G System (5GS)," 3rd Generation Partnership Project (3GPP), Technical Specification (TS) 23.304, 03 2023, version 18.1.0.

[16] ——, "Technical Specification Group Radio Access Network; NR; Physical channels and modulation," 3rd Generation Partnership Project (3GPP), Technical Specification (TS) 38.211, 01 2023, version 17.4.0.

[17] ——, "Technical Specification Group Radio Access Network; NR; Physical layer procedures for data," 3rd Generation Partnership Project (3GPP), Technical Specification (TS) 38.214, 03 2023, version 17.5.0.

[18] ——, "Technical Specification Group Radio Access Network; NR; Medium Access Control (MAC) protocol specification," 3rd Generation Partnership Project (3GPP), Technical Specification (TS) 38.321, 03 2023, version 17.4.0.

[19] ——, "Technical Specification Group Radio Access Network; NR; Radio Resource Control (RRC) protocol specification," 3rd Generation Partnership Project (3GPP), Technical Specification (TS) 38.331, 03 2023, version 17.4.0.

[20] Department of Defense (DoD), "Military units," https://www.defense.gov/Multimedia/Experience/Military-Units/Army/#army, 2023.