

Contention-Based Access for Ultra-Reliable Low Latency Uplink Transmissions

Bikramjit Singh[†], Olav Tirkkonen[†], Zexian Li[§] and Mikko A. Uusitalo[§]

[†]*Department of Communications and Networking, Aalto University, Espoo, Finland*
{bikramjit.singh, olav.tirkkonen}@aalto.fi

[§]*Nokia Bell Labs, Espoo, Finland*
{zexian.li, mikko.uusitalo}@nokia-bell-labs.com

Abstract—We consider sporadic ultra-reliable and low latency communications in the uplink 5G cellular systems. Reliable low latency access for randomly emerging packet transmission needs cannot be guaranteed in current wireless systems. To achieve the goal of low latency and high reliability simultaneously, we propose a contention-based transmission scheme aimed for the users with small payloads. We target to reduce collision probability by considering multiple transmissions for the same packet for reliable reception. We find the optimal number of consecutive multiple transmissions that reduces collisions and achieves target reliability within the latency window. By means of intended frame structure design for the 5G cellular systems, results are drawn and comparisons are made with default multi-channel slotted ALOHA access scheme.

Keywords—5G; ultra-reliable and low latency communications (URLLC); multi-channel slotted ALOHA; contention; collision; small packet

I. INTRODUCTION

The 5G communication systems will be introduced in the early 2020s, enabling expansion of International Mobile Telecommunications (IMT) going beyond IMT-2000 and IMT-Advanced services. *Ultra-Reliable and Low Latency Communications (URLLC)* is a target usage scenario that enables real-time control and automation of dynamic processes for vertical applications. Such services require very high reliability and often short latency down to the millisecond level [1], [2]. The most stringent reliability requirement on URLLC currently being standardized is 99.999 % under the radio latency bound of 1 ms [2]. For a small packet with 32 bytes, the packet error rate must not be higher than 10^{-5} , and the maximum allowable radio latency, including possible re-transmissions is 1 ms.

To shorten the latency, User Equipment (UE) can transmit uplink traffic in a contention-based grant-free manner without waiting for grant or resource allocation message. In contention-based access two or more UEs may attempt to transmit simultaneously over a shared channel and collisions may occur. ALOHA is a well known example of a contention protocol [3]. Slotted ALOHA increases throughput by dividing the channel into slots of equal duration and making the UEs contend for these slots [4]. In multi-channel slotted ALOHA protocols [5]–[7], the contention resource is divided into multiple channels, one of which is chosen by the accessing UE. In [5], multi-channel slotted ALOHA is analyzed for fixed spectrum for multi-channel satellite communication. In [6], the UEs access the channel with some probability, which is broadcasted by the network in order to prevent collisions. However, the discussed schemes do not follow tight latency requirements.

Packet success rate in [7]–[9] is improved for a contention-

based access in slotted ALOHA. The schemes consider repetition coding where the packet is transmitted multiple times to improve the success rate. Further, [7], [9] considered successive interference cancellation or iterative collision resolution where some of the collided packets are recovered by subtracting the packet replicas from the others. In [7], [9], the latency is increased for the collided packet resolution where significant waiting period is consumed in the reception of other UE's packet replica and the following interference cancellation. Complexity is also increased related to enabling packet recovery process, e.g., storing of packet replicas and collision information. Further, the packet replicas are required to be equipped with pointers to the slots containing them, or embed with UE-specific identification which is known by the network in advance. The schemes in [7], [9] improve the packet success rate, but latency is increased. Accordingly, they are not suitable from the perspective of 5G URLLC services. Scheduled access can be favored for high arrival traffic where UEs quite often have data to transmit [10]. However, for sporadic arrival rates, scheduling is not a good option anymore due to the signaling overhead, increased latency owing to resource request and allocation. Further, control information has to be delivered in a reliable manner as well.

In this letter, we consider uplink contention-based access for sporadic arrivals for the URLLC packets. Hence, UEs can transmit data in an arrive-and-go manner without sending scheduling request and receiving resource grant from the network. This reduces latency and therefore, it has a growing interests among notable industrial players for 5G URLLC services [11]. To further reduce latency, the latest 3GPP agreed TTI length 0.125 ms is adopted in our analysis, where each frame contains both control and data information [11]–[13]. With this, on average 4 Transmission Time Intervals (TTIs) are available for the transmission in one direction with 0.5 ms latency.

We consider diversity transmission (or repetition coding) similar to [8] to improve the reliability of multi-channel slotted ALOHA for URLLC services, the same data packet from one UE is sent multiple times in consecutive TTIs. The packet success rate is improved at the expense of transmission redundancy still within the tight URLLC latency window. In contrast to [8], here we deduce the *optimal number of repetitions* so that packet success rate follows the strict URLLC reliability constraint, e.g., 99.999 %. The protocol also considers scheduled allocation in case the same URLLC services can be provided with less resources than with contention-based access. Furthermore, we combine diversity transmission with multi-user detection, which further increases packet success rate. Multi-user detection uses the advanced coding schemes and/or transceiver designs, e.g.,

described in [14], [15], and improves packet decoding success by enabling the system to recover the packets from collision.

II. SYSTEM MODEL

We consider sporadic uplink transmission scenario where there is a population of N UEs that randomly use the radio resources for their data transmission. For each UE, the packet arrival follows Poisson process with exponentially distributed inter-arrival time [12]. The average number of random access events in an interval is denoted by λ . With a TTI length T , and an average packet inter-arrival time μ , we have $\lambda = T/\mu$. The N Poisson arrivals processes in the cellular network are assumed independent. 3GPP URLLC reliability requirement [2] considers small packet size of 32 bytes, and we assume that each packet may require an access slot of a resource block in frequency domain and a unit TTI in time domain. The maximum permitted End-to-End (E2E) latency for URLLC is L and the URLLC reliability target is $\mathbb{P}_{\text{rel}} = 0.99999$. The system bandwidth is divided into K resource blocks in one TTI. For the sake of simplicity, we focus on MAC-layer reliability against the collisions, and do not consider the transmission channel impact in the study. With lowest possible MCS scheme, we assume that a single transmission can achieve $\text{BLER} < 1 - \mathbb{P}_{\text{rel}}$ [16].

III. CONTENTION-BASED ACCESS PROCEDURE

Assume a certain amount of resource is reserved for URLLC services, if the UE population exceeds the amount of available resources, contention-based access can be exploited. This may result in collisions between the packets, and the reliability performance may suffer. Depending on the network scenario, e.g., UE population, traffic pattern, network may adopt various strategies to minimize or resolve collisions in the contention resource. For example, it can consider advance receivers with multi-user detection. The network may also ask UEs to transmit multiple times to ensure the packet's successful reception. The scheme can be termed as diversity transmission, where the diversity is introduced in the packet transmission.

A. Collision probability analysis

Given the modeling parameters described in Section II, first we provide collision probability analysis for a multi-channel slotted ALOHA. The analysis is re-drawn for the completeness and comparison with the extended schemes described in the later sections.

Probability that one UE to have one or more random transmission event(s) x (i.e. packets to transmit) in a TTI is

$$\mathbb{P}_{\text{ra}} = \mathbb{P}(x > 0) = 1 - e^{-\lambda}.$$

Now, there might be a possibility that the other $N - 1$ UEs try to access the same TTI of that of UE of interest. The probability that no other UE (from UE set $\{N - 1\}$) has random transmission event in the same TTI as the UE of interest is

$$\mathbb{P}_0 = (1 - \mathbb{P}_{\text{ra}})^{N-1}.$$

The probability that n UEs (from UE set $\{N - 1\}$) has random transmission events in the same TTI is

$$\mathbb{P}_n = \binom{N-1}{n} \mathbb{P}_{\text{ra}}^n (1 - \mathbb{P}_{\text{ra}})^{N-n-1}.$$

The probability that these n UEs do not access the same resource block as the UE of interest in the given TTI is

$$\mathbb{P}_{\text{acc}}(n, 0) = \left(\frac{K-1}{K}\right)^n. \quad (1)$$

However these n UEs do not collide with the UE of interest but may collide with themselves. Now the probability of no collision between the UE of interest and any other UE is

$$\mathbb{P}_{\Sigma} = \sum_{n=1}^{N-1} \mathbb{P}_n \mathbb{P}_{\text{acc}}(n, 0). \quad (2)$$

Then the collision probability of the UE of interest is

$$\mathbb{P}_c = 1 - \mathbb{P}_0 - \mathbb{P}_{\Sigma}, \quad (3)$$

$$= 1 - \left(\frac{e^{-\lambda} + K - 1}{K}\right)^{N-1}. \quad (4)$$

As an example, with the assumption of $K = 6$ resource blocks, average packet inter-arrival time $\mu = 10$ s, packet length equivalent to one resource block, TTI length $T = 0.125$ ms, URLLC target $\mathbb{P}_{\text{rel}} = 0.99999$, and the average arrival rate is $\lambda = 1.25 \times 10^{-5}$. Using Eq. (4), we find that the maximum number of UEs that can be supported with collision probability no more than $\mathbb{P}_c \leq 1 - \mathbb{P}_{\text{rel}} = 10^{-5}$ is $N = 5$. The collision probability curve is depicted in Fig. 1.

B. Collision reduction

The network can reduce collisions or decode collided packets by multi-user detection and diversity transmission.

1) *Multi-user detection*: Previously in Section III-A, we assumed that the probability of successful decoding of a collided packet is zero. However, the collision probability can be reduced if we consider advanced receivers with multi-user detection capability. To proceed with the analysis, let us denote the probability of detection of a packet of the UE of interest in case it collides with simultaneous packets of q other UEs $\mathbb{P}_{\text{det}_q}$. Now, the probability that out of n UEs, q UEs access the same resource block as the UE of interest is

$$\mathbb{P}_{\text{acc}}(n, q) = \binom{n}{q} \frac{(K-1)^{(n-q)}}{K^n}.$$

The probability of successful reception in case of collision with other q UEs from n UEs is then

$$\mathbb{P}_{\text{succ}}(n, q) = \mathbb{P}_{\text{acc}}(n, q) \mathbb{P}_{\text{det}_q}.$$

The sum probability of the successful packet reception of the UE of interest if there is a collision with $q = 1, \dots, n$ other UEs is

$$\mathbb{P}_{\text{succ}}(n) = \sum_{q=1}^n \mathbb{P}_{\text{succ}}(n, q) \mathbb{P}_{\text{det}_q}. \quad (5)$$

Now, the total probability of successful packet transmission of the UE of interest if other $n = 1, \dots, N-1$ UEs transmit in the same TTI is

$$\mathbb{P}_{\text{succ}} = \sum_{n=1}^{N-1} \mathbb{P}_n \mathbb{P}_{\text{succ}}(n). \quad (6)$$

Using Eq. (3), the collision probability is

$$\mathbb{P}_c^{\text{MUD}} = \mathbb{P}_c - \mathbb{P}_{\text{succ}}, \quad (7)$$

$$= 1 - \left(\frac{e^{-\lambda} + K - 1}{K}\right)^{N-1} \quad (8)$$

$$- \frac{(N-1)! \sum_{n=1}^{N-1} (e^{-\lambda} - 1)^n \sum_{q=1}^n (K-1)^{n-q}}{e^{\lambda(N-1)} \sum_{n=1}^{N-1} (N-n-1)! K^n \sum_{q=1}^n (n-q)! q!} \mathbb{P}_{\text{det}_q}.$$

The detection probabilities $\{\mathbb{P}_{\text{det}_q}\}$ depend on the scenario, e.g., traffic pattern, transmission power, receiver capability. According to [14], successful decoding of collided packet begins when one packet is 2.5 dB stronger than others, and the detection probability becomes almost 100 % when relative

signal strength is 5 dB or more.

To obtain the detection probabilities, we simulated a packet arrival scenario in a single node network based on the parameters described in [1]. Each UE chooses the resource block randomly when it has a packet to transfer. We draw Signal-to-Interference Ratio (SIR) distribution of the collided packet when it simultaneously transmits with other $q = 1, 2, 3, 4, \dots$ packet(s) on a same resource and obtain its relative SIR strength. For the sake of simplicity, we consider the detection probability to be the percentage of collisions given the other q packets where the UE of interest is at least 5 dB stronger than the others, as proposed in [14]. Hence the average detection probabilities for the given collision scenario are $\mathbb{P}_{\text{det}_1} = 0.393$, $\mathbb{P}_{\text{det}_2} = 0.207$, $\mathbb{P}_{\text{det}_3} = 0.115$, $\mathbb{P}_{\text{det}_4} = 0.109$, and so on. From Eq. (8) and the given parameters in Section III-A, we obtain the maximum number of UEs that can be supported in case of multi-user detection as $N = 8$, and Fig. 1 depicts the supported UE population against the collision probability.

2) *Diversity transmission*: The collision probability can be reduced by means of diversity transmission. In this case a UE transmits the same data packet in Γ subsequent TTIs. However, in every subsequent TTI, the resource block used is randomly chosen from the K available blocks. With this, some packets may collide but some may be received successfully. This brings down the overall collision probability. With Γ diversity transmission, each UE transmits the same data packet Γ times in Γ consecutive TTIs. We treat the K contention resources in Γ TTIs as a new contention pool, which then has K^Γ resources randomly chosen from. For the diversity transmission in contention pool, the traffic intensity is increased to $\Gamma\lambda$. Analogously to (4), the collision probability then becomes

$$\mathbb{P}_c^{\text{DT}} = 1 - \left(\frac{e^{-\Gamma\lambda} + K^\Gamma - 1}{K^\Gamma} \right)^{N-1}. \quad (9)$$

Let us find the maximum number of UEs that can be supported for the URLLC transmissions with reliability target \mathbb{P}_{rel} . The parameters are stated in Section III-A. Given the URLLC E2E latency constraint $L = 1$ ms, the maximum diversity level that can be supported for uplink transmissions is $\Gamma = L/2T = 4$, i.e., at most four consecutive transmissions are allowed. For collision probability $\mathbb{P}_c^{\text{DT}} \leq 10^{-5}$ and $K = 6$ resource blocks, the maximum number of UEs that can be supported is $N = 260$. Fig. 1 depicts the supported UE population against the collision probability for varying diversity degrees. Further, Table I details the diversity levels and the corresponding supported UE population sizes for different arrival rates. Hence, for a given load and resource, the network can choose an optimal number of repetitions that meets reliability and latency constraints.

TABLE I.
URLLC UE POPULATION SIZES FOR ARRIVAL RATES AND DIVERSITY DEGREES. CONTENTION POOL WITH $K = 6$ RESOURCE BLOCKS.

UE population (N)		Diversity (Γ)			
		1	2	3	4
Mean arrival rate (λ)	1.25×10^{-6}	49	145	577	2593
	1.25×10^{-5}	5 [†]	15	58	260
	1.25×10^{-4}	1 [†]	2 [†]	6 [†]	26

[†]The contention-based access is resource inefficient as $K \geq N$. Instead, scheduled access should be utilized where resource consumption will be $K = N$. See Section III-C for more details.

Scenarios with high population and/or high arrival rate may not benefit from diversity transmission. They are effective only in a relative abundance of resources, where they can be utilized to optimize the resources to deliver the targeted reliable services. However, for UEs with high arrival rate, scheduled access is preferable. The discussed model here presumes sporadic transmissions. In case of high arrival rates, collisions can happen between the replicas of different packets of a UE if access is random. In case of high arrival rates, there might be instances that the whole contention resource K may run into outage due to the large amount of packets and replicas generation for the individual UEs. Hence, for high arrival rates, accurate modeling for contention transmissions should consider self-collisions. In such scenarios, URLLC services cannot be established without scheduled access. Fig. 2 depicts the simulated and analytical behavior of the collision performance for different arrival rates, population sizes and diversity degrees, and $K = 6$ resource blocks. The analytical model breaks down at high arrival rate. The higher the diversity degree, the earlier the model breaks down. However, at the target reliability level, the model works well.

3) *Combined multi-user detection and diversity transmission*: The network exploits both multi-user detection and diversity transmission. From Eq. (8) and (9), the collision probability is

$$\mathbb{P}_c^{\text{MUD+DT}} = 1 - \left(\frac{e^{-\Gamma\lambda} + K^\Gamma - 1}{K^\Gamma} \right)^{N-1} - \left(\frac{(N-1)!}{e^{\Gamma\lambda(N-1)}} \sum_{n=1}^{N-1} \frac{(e^{\Gamma\lambda} - 1)^n}{(N-n-1)! K^{\Gamma n}} \times \sum_{q=1}^n \frac{(K^\Gamma - 1)^{n-q}}{(n-q)! q!} \mathbb{P}_{\text{det}_q} \right). \quad (10)$$

We consider the referenced parameters listed in Section III-A-1, detection probabilities $\{\mathbb{P}_{\text{det}_q}\}$ stated in Section III-B-1, and maximum diversity level $\Gamma = 4$ for the transmissions. For collision probability $\mathbb{P}_c^{\text{MUD+DT}} \leq 10^{-5}$, the maximum number of UEs supported is $N = 428$. In Fig. 1, the blue curve depicts the number of UEs that are supported against the target collision probability.

There is an increase in latency from multi-user detection in the schemes in Section III-B-1 and III-B-3, but we assume that the receiver has sufficient hardware to treat multi-user detection in a pipeline fashion, so that latencies are not substantially increased with multi-user detection.

C. Flexible Access

In some scenarios of Table I, contention-based access underperforms, as with $K = 6$ resource blocks, the maximum number of UEs supported with scheduled (orthogonal) allocation is $N = 6$ which has zero collision probability. The fact that the allocation in contention-based access is random and therefore limits the number of UEs subject to the target collision probability. If we consider a lower packet arrival rate, the contention-based access may outperform scheduled allocation, e.g., if the average arrival rate at $\lambda = 1.25 \times 10^{-5}$ is halved, i.e., $\lambda \rightarrow \lambda/2$, then the maximum number of supported UEs increased to $N = 10$ for same resource $K = 6$ resource blocks. The network can choose between using contention-based or scheduled access, depending on which mode needs less resources for given reliability and latency targets. The number of resources used by the system is then accordingly

$$K_{\text{contention}} = \arg \max_K \mathbb{P}_c^{\text{MUD+DT}}(K, \Gamma),$$

$$\text{s.t. } \mathbb{P}_c^{\text{MUD+DT}} \leq 1 - \mathbb{P}_{\text{rel}},$$

$$\Gamma \leq L/2T$$

and

$$K = \min(K_{\text{contention}}, N). \quad (11)$$

From Eq. (11), if $K_{\text{contention}} < N$, then the network chooses contention-based access with collision probability $\mathbb{P}_c^{\text{MUD+DT}} < 1 - \mathbb{P}_{\text{rel}}$. Otherwise, the network does scheduled allocation on a bandwidth $K = N$ carriers and has zero collisions. We see in Fig. 1, with flexible access, the collision performance improves at the bottom of CDF tail.

IV. CONCLUSION

In this letter, we analyzed uplink contention-based transmission scheme for the ultra-reliable and low latency communications in 5G, facilitating tight latency constraint for sporadic and small packet data transmissions. Contention-based transmissions enables low latency access to data channels. To improve the reliability against the collisions, diversity transmission can be utilized where packet replicas are sent multiple times to reduce the collision probability. It offers prominent gains and lower implementation complexity in comparison to multi-user detection. Further, it would be beneficial to have packet replicas equipped with pointers to the selected resource blocks in the other $\Gamma - 1$ diversity transmissions, as suggested in [7]. It may offer faster processing, e.g., rejection of redundant packets in case a packet had received successfully, or in deriving additional gain from diversity combining.

V. ACKNOWLEDGMENT

This work was supported in part by the Finnish Funding Agency for Innovation (TEKES) under the project ‘‘Wireless for Verticals (WIVE)’’. WIVE is a part of 5G Test Network Finland (5GTNF).

REFERENCES

[1] B. Singh *et al.*, ‘‘Ultra-reliable communication in a factory environment for 5G wireless networks: Link level and deployment study,’’ in *Proc. IEEE PIMRC Workshop*, pp. 1-5, Sep. 2016.

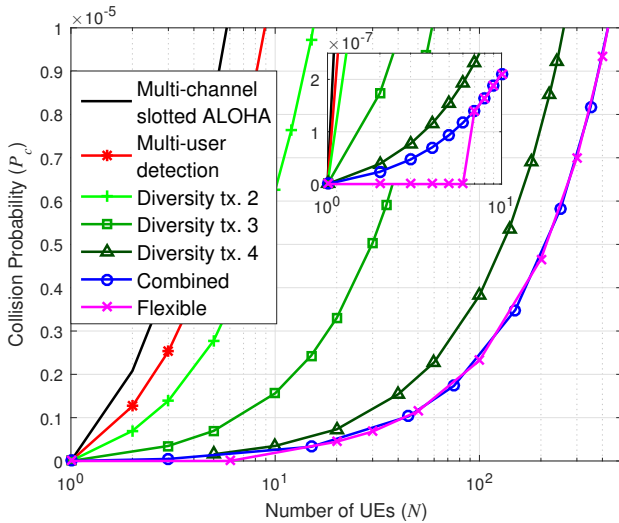


Fig. 1. URLLC UE population supported against the collision probability for various schemes, and $K = 6$ resource blocks, and mean arrival rate $\lambda = 1.25 \times 10^{-5}$. Zoom-in part depicts the collision probability at the bottom of CDF tail.

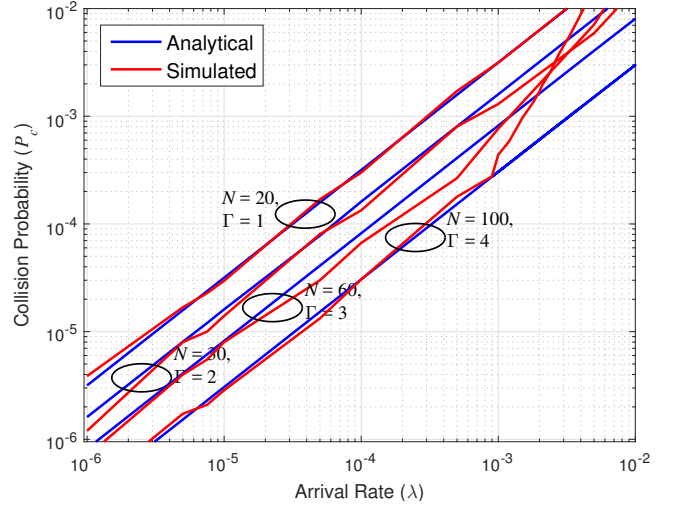


Fig. 2. Collision analysis for the contention transmissions for the different UE arrival rates, population sizes and diversity levels.

[2] 3GPP, ‘‘Study on scenarios and requirements for next generation access technologies,’’ 3GPP TR 38.913 V14.2.0, Tech. Rep., Mar. 2017.

[3] N. Abramson, ‘‘The ALOHA system - Another alternative for computer communications,’’ in *Proc. FJCC*, vol. 37, 1970.

[4] L. G. Roberts, ‘‘ALOHA packet system with and without slots and capture,’’ *ACM SIGCOMM Comput. Commun. Rev.*, vol. 5, no. 2, pp. 28-42, Apr. 1975.

[5] H. H. Tan and H. Wang, ‘‘Performance of multiple parallel slotted Aloha channels,’’ in *Proc. INFOCOM*, pp. 931-940, Mar. 1987.

[6] M. Tavana, V. S.-Mansouri and V. W. S. Wong, ‘‘Congestion control for bursty M2M traffic in LTE networks,’’ in *Proc. IEEE ICC*, pp. 5815-5820, Jun. 2015.

[7] E. Paolini *et al.*, ‘‘Coded random access: Applying codes on graphs to design random access protocols,’’ *IEEE Commun. Mag.*, vol. 53, no. 6, pp. 144-150, Jun. 2015.

[8] G. Choudhury and S. Rappaport, ‘‘Diversity ALOHA-A random access scheme for satellite communication,’’ *IEEE Trans. Commun.*, vol. 31, no. 3, pp. 450-457, Mar. 1983.

[9] M. Ghanbarinejad and C. Schlegel, ‘‘Irregular repetition slotted ALOHA with multiuser detection,’’ in *Proc. WONS*, pp. 201-205, Aug. 2013.

[10] B. Singh *et al.*, ‘‘Selective multi-hop relaying for ultra-reliable communication in a factory environment,’’ in *Proc. IEEE PIMRC*, pp. 1-6, Sep. 2016.

[11] Nokia and Alcatel-Lucent Shanghai Bell, ‘‘Contention-based grant-free transmission for URLLC,’’ 3GPP TSG-RAN WG1 Meeting NR#1, Jan. 2017.

[12] 3GPP, ‘‘Study on new radio access technology physical layer aspects,’’ 3GPP TR 38.802 V14.0.0, Tech. Rep., Mar. 2017.

[13] Nokia and Alcatel-Lucent Shanghai Bell, ‘‘Flexible numerology for 5G,’’ 3GPP TSG-RAN WG1 Meeting #85, May 2016.

[14] B. Firner *et al.*, ‘‘Multiple receiver strategies for minimizing packet loss in dense sensor networks,’’ in *Proc. ACM MobiHoc*, pp. 211-220, Sep. 2010.

[15] F. Monsees *et al.*, ‘‘Compressive sensing multi-user detection for multicarrier systems in sporadic machine type communication,’’ in *Proc. IEEE VTC Spring*, pp. 1-5, May 2015.

[16] M. Sybis *et al.*, ‘‘Channel coding for ultra-reliable low latency communications in 5G systems,’’ in *Proc. IEEE VTC Fall*, pp. 1-5, Sep. 2016.