Availability Indication as Key Enabler for Ultra-Reliable Communication in 5G

Hans D. Schotten and Raja Sattiraju
University of Kaiserslautern
Kaiserslautern, Germany
{schotten, sattiraju}@eit.uni-kl.de

David Gozalvez Serrano, Zhe Ren and Peter Fertl
BMW Group Research and Technology
Munich, Germany
{david.gozalvez-serrano, zhe.ren, peter.fertl}@bmw.de

***Abstract*—Due to its flexibility, cost-efficiency, and the ability to support mobility, wireless connectivity is seen today as a key enabler for a wide range of applications beyond classical mobile communications. A significant part of these applications depends on the capability of the wireless communication system to provide reliable connectivity. However, due to the randomness of the wireless propagation channel, reliability is still a critical issue in these systems. Some applications, such as vehicular and industrial applications, demand a level of reliability that wireless communication systems typically are not able to guarantee. This paper provides a framework that enables these applications to make use of wireless connectivity only if the transmission conditions are favorable enough. The concept is based on the idea that - despite the fact that it is practically impossible to ensure error-free wireless communication - it is feasible to derive boundary conditions for the transmission success. To this end, the paper introduces a novel metric for Ultra-Reliable Communication (URC) referred to as "Availability", that determines the expected presence or absence of link reliability at the time of transmission. The availability is signaled by means of an Availability Indicator (AI) to the applications. Moreover, we develop the system model for computing the AI and illustrate the potential benefits of the new reliability metric by means of a possible implementation for automotive scenarios.**

I. Introduction

During the recent years, there has been a tremendous surge in the demand for mobile and wireless connectivity, which is forecasted to grow exponentially within the foreseeable future. However, the future Fifth-Generation (5G) mobile communication system [1] that is currently under discussion will not only have to cope with an increasing demand of traffic volume, but also provide a broader range of applications with new and not yet supported requirements in terms of reliability, availability and efficiency.

Besides integrating new radio access concepts such as Device-to-Device (D2D) communications, Massive Machine Type Communications (MMC), Ultra Dense Networks (UDN), or Moving Networks (MN), the support of Ultra-Reliable Communication (URC) is seen as a key enabler for 5G mobile communication systems [2]. Fast growing application domains requiring URC are for example road safety systems, automatic train control systems, industrial applications, and some e-health services, where reliability is identified as an instrumental property [3].

For example, road safety applications based on wireless communications require that a very high percentage of safety messages are successfully delivered to other traffic participants within a certain deadline [4] [5]. The failure to comply with these requirements renders the road safety service completely useless and even harmful to the users relying on it. An analysis of the above mentioned applications indicates that "reliability" needs to be defined on a link as well as on a system level. Link reliability is the ability of a radio link to transmit and receive a certain amount of data successfully within a predefined deadline. System reliability is the ability of a system to accurately indicate the absence of link reliability to the application, and at the same time, to ensure the presence of link reliability as often as possible when required by the application.

In this paper, we introduce a system concept for URC based on a novel metric referred to as "Availability", which indicates the presence or absence of link reliability. URC services such as road safety applications require very high and predictable success rates within low deadlines (e.g., 100 ms). Due to the sensitive nature of these applications, it is of paramount importance to warn the application about the absence of link reliability according to the application-specific requirements. Wireless communication systems are typically not designed to provide reliability at all times and in every reception scenario, as this would results in an overdesigned system with a very inefficient air interface in terms of data rate and power consumption. Such an approach would harm the acceptance of URC services and restrict their usage. In the following, we define the URC concept in a universal manner, i.e., without relying on details of any air interface specification or signaling scheme. This approach is motivated by the fact, that a transport-agnostic definition from the applications point of view is required in order to allow URC services to be deployed in a wide range of scenarios. From this perspective, the implementation details related to the wireless communication system are not included in the proposed URC concept.

The rest of the paper is organized as follows. Section II introduces the URC system concept and defines the main metrics used for the computation of the availability metric.

[1]Part of this work has been performed in the framework of the FP7 project ICT-317669 METIS, which is partly funded by the European Union. The authors would like to acknowledge the contributions of their colleagues in METIS, although the views expressed are those of the authors and do not necessarily represent the project.

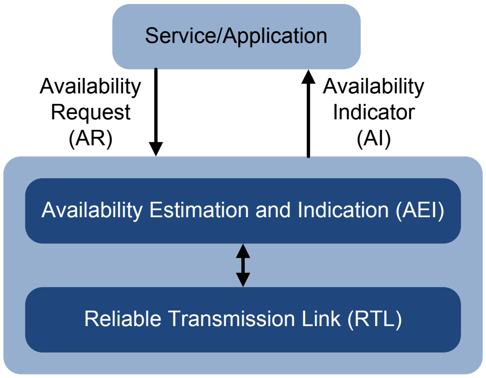

Fig. 1. System concept of URC.

Section III presents a possible implementation of the URC concept in the context of automotive scenarios. Section IV illustrates the benefits of the proposed concept by means of a realization based on the predicted SINR. Section V gives some concluding remarks.

II. System Concept and Definition

As illustrated in Fig. 1, the system concept for URC is based on a "Reliable Transmission Link" (RTL) that is optimized to transmit packets successfully and within a predefined deadline, and an "Availability Estimation and Indication" (AEI) mechanism that is able to reliably predict the availability of the RTL under given conditions. In addition, a novel link control indicator called Availability Indicator (AI) signals the outcome of the AEI to the application. In this context, an application requests an RTL by sending an Availability Request (AR) to the AEI. Depending on the implementation details, the AR contains information such as the packet size, the maximum acceptable delay until successful reception or the maximum tolerable error probability. The AEI is designed to indicate to the application the availability of the RTL for the forthcoming transmissions given the AR requirements. For the availability estimation, the AEI needs to monitor the channel conditions, e.g., by evaluating the Signal-to-Noise and Interference Ratio (SINR) and/or the ACK/NACK statistics of the retransmission protocols used at link level. Typically, the AI is a binary value, i.e., either RTL available (AI=1) or unavailable (AI=0). After indicating the RTL availability, the application will be able to use it by transmitting data packets over the RTL (not shown in Fig 1).

A. Mathematical Interpretation

In the following, we formulate mathematically the URC concept by adopting a simple time-slotted model, in which each time slot, τ, corresponds to the time interval $[t, t+\mathrm{D}_{\mathrm{AR}})$, where D_{AR} is the maximum delay tolerated by the application. Under this definition we define

$$\mathrm{RTL}(\tau) = \begin{cases} 1, & \text{transmission is successful} \\ & \text{for time slot } \tau; \\ 0, & \text{transmission is not successful} \\ & \text{for time slot } \tau. \end{cases}$$

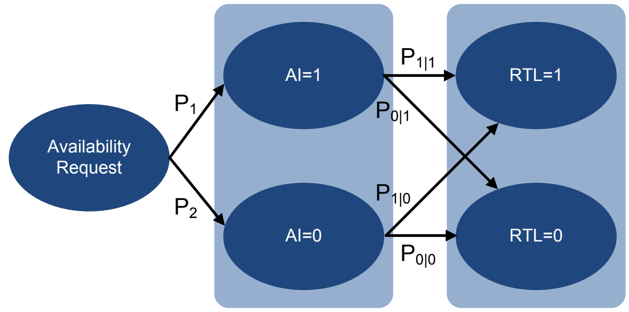

Fig. 2. URC State Transition Probabilities.

For the availability indication, a simple binary signaling format per time slot is used, where

$$\mathrm{AI}(\tau) = \begin{cases} 1, & \text{AI indicates the availability of RTL} \\ & \text{for time slot } \tau; \\ 0, & \text{AI indicates the non-availability of RTL} \\ & \text{for time slot } \tau. \end{cases}$$

Fig. 2 describes the URC state transition probabilities divided into two steps. The essence of the URC concept is that an application should rely on the wireless communication only in those instances in which the link reliability is guaranteed with a certain probability. In other words, once the AI indicates the availability of a RTL for time slot τ to the application, the probability of transmission success for the time slot τ must be above a certain value, P_{UR}, according to the application requirements. We refer to this criteria as the ultra-reliable requirement, which can be understood as the conditional probability

$$\mathrm{P}_{1|1} = \Pr(\mathrm{RTL}(\tau) = 1 | \mathrm{AI}(\tau) = 1), \tag{1}$$

where $\mathrm{P}_{1|1} \geq \mathrm{P}_{\mathrm{UR}}$. In this context, the goal of the URC system design is to maximize the availability of the RTL under the previously defined requirement:

$$\begin{aligned} \max \quad & \mathrm{P}_1 = \Pr(\mathrm{AI}(\tau) = 1) \\ \text{s.t.} \quad & \mathrm{P}_{1|1} \geq \mathrm{P}_{\mathrm{UR}}. \end{aligned} \tag{2}$$

From Fig. 2 and according to Bayes' rule, we have

$$\begin{aligned} \mathrm{P}_1 &= \Pr(\mathrm{AI}(\tau) = 1) \\ &= \frac{\Pr(\mathrm{RTL}(\tau) = 1)\Pr(\mathrm{AI}(\tau) = 1 | \mathrm{RTL}(\tau) = 1)}{\mathrm{P}_{1|1}}. \end{aligned} \tag{3}$$

It can be concluded that in order to optimize the URC concept, there are in general two possibilities: either by improving the transmission scheme ($\Pr(\mathrm{RTL}(\tau) = 1)$) or by improving the estimation of the AI ($\Pr(\mathrm{AI}(\tau) = 1 | \mathrm{RTL}(\tau) = 1)$). The first possibility can be achieved by means of more robust modulation and coding techniques, whereas the second possibility can be accomplished through more accurate channel estimation and prediction techniques. On the other hand, it is interesting to note that the two probabilities $\mathrm{P}_{0|1}$ and $\mathrm{P}_{1|0}$ in Fig. 2 correspond to the Type I and Type II errors [6] in statistical

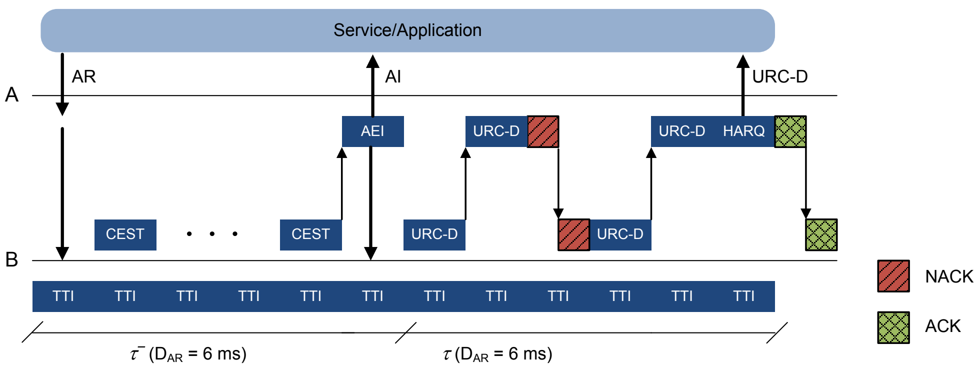

Fig. 3. URC Timing Structure.

probability analysis:

$$\mathrm{P}_{0|1} = \Pr(\mathrm{RTL}(\tau) = 0|\mathrm{AI}(\tau) = 1) \quad (4)$$
$$\mathrm{P}_{1|0} = \Pr(\mathrm{RTL}(\tau) = 1|\mathrm{AI}(\tau) = 0). \quad (5)$$

B. URC Example based on SINR

We now explain the URC concept using a simple example based on the predicted SINR. Let us assume that the SINR for time slot τ is $\Gamma(\tau)$. We define that the transmission is successful for time slot τ if the actual SINR is larger than or equal to a given threshold Γ_l ($\Gamma(\tau) \geq \Gamma_l$), which is given by the selected modulation and coding scheme including the use of retransmissions. On the other hand, the AEI signals to the application AI=1 if the predicted SINR for time slot τ, $\Gamma_p(\tau)$, is larger than or equal to the threshold Γ_l ($\Gamma_p(\tau) \geq \Gamma_l$). Under these definitions, we can formulate the optimization problem of the URC concept as

$$\begin{aligned} \max \quad & \mathrm{P}_1 = \Pr(\Gamma(\tau) \geq \Gamma_l) \\ \text{s.t.} \quad & \mathrm{P}_{1|1} = \Pr(\Gamma(\tau) \geq \Gamma_l | \Gamma_p(\tau) \geq \Gamma_l), \end{aligned}$$

whereas the type I and Type II error can be formulated as follows

$$\begin{aligned} \mathrm{P}_{0|1} &= \Pr(\Gamma(\tau) < \Gamma_l | \Gamma_p(\tau) \geq \Gamma_l) \\ \mathrm{P}_{1|0} &= \Pr(\Gamma(\tau) \geq \Gamma_l | \Gamma_p(\tau) < \Gamma_l). \end{aligned}$$

In this example, it would be possible to optimize the URC concept by decreasing Γ_l by means of more robust modulation and coding schemes, or by improving the computation of $\Gamma_p(\tau)$ in order to predict more accurately the availability of the RTL.

III. URC Implementation in Automotive Scenarios

Figure 3 describes a possible implementation of the URC system concept. According to this implementation, an application requests in time slot τ^- the use of an RTL for the transmission of data from “B” to “A”. This could be the case of a road safety application awaiting information from nearby traffic participants in order to alert the driver or to take control of the vehicle so that a collision could be avoided. The request is being done by means of an AR, which is first generated in “A” and then forwarded to “B”. Please note that the AR could alternatively be generated by the application in B. In response to the AR, “B” starts the transmission of a pilot signal for channel estimation (CEST), which is used by the AEI in “A” to predict the availability of the RTL in the following time slot τ according to the maximum delay that is tolerated by the application. In the case of Fig. 3, each time slot extends across six Transmission Time Intervals (TTIs) ($\mathrm{D}_{\mathrm{AR}} = 6ms$).

Based on this prediction, the AI will be signaled to the application requesting the RTL. In case of RTL availability (AI=1), the application informs both “A” and “B” about the start of the RTL usage. After that, the application starts sending the URC data (URC-D) from “B” to “A” through the RTL. Here a processing delay of 1 ms is assumed. “A” will then try to decode the URC-D. If successful, an ACK is sent; if not successful, a NACK is sent instead. Fig. 3 shows the case where the first decoding attempt is not successful, and a NACK is sent from “A”, which triggers “B” to send a retransmission. After receiving the retransmission, the HARQ recombining process in “A” decodes successfully the data and delivers it to the application.

As seen in the example, the AEI is in charge of predicting the availability of the RTL in future time slots based on the application requirements carried by the AR, and also on channel information such as SINR measurements. The AR comprises information about the application requirements that might impact the availability of the RTL, such as the data packet size or the maximum delay until successful reception. For example, in the case of a road safety application for intersection collision warning, data packets of approximately 200 bytes have to be delivered successfully to other traffic participants within 100 ms. The computation of the AI in automotive scenarios can also benefit significantly from information regarding the variability of the channel in the time domain, especially in wireless communication systems that allow the use of retransmissions and HARQ. Fast varying channels generally result in a higher time diversity that improves the performance gain of retransmissions and HARQ techniques. In this sense, the vehicle speed can be used to estimate the time variability of the channel and predict the availability of the RTL in a more precise manner. Furthermore, the AEI mechanism can also use related vehicular sensor and context information for the

TABLE I
SIMULATION ASSUMPTIONS

System	Parameter	Value
Cellular	Layout	Hexagonal
	Antennas	Omnidirectional
	Transmit Power	46 dBm
	Bandwidth	5 Mhz
	Carrier Freq	2 GHz
	Path Loss	L = L_0 + 37.6 Log_{10} (d)
V2V Link	Doppler	[10.0, 33.3, 100, 200] Hz
	Coherence Time	[42.3, 12.7, 4.2, 2.1] ms
	Path Loss	LOS
	MCS	QPSK & 64 QAM (HSDPA) [7]
	TTI	1 ms
	Retransmission	HARQ
	Receiver	Perfect MMSE
Application	Packet Size	1600 Bytes
	Max Deadline	6 TTI's

computation of the AI, such as route and trajectory information given by the navigation system of the vehicle in combination with pre-recorded radio maps, which might contain, among others, information related to the SINR or the packet error rate.

IV. PERFORMANCE EVALUATION

Based on the example in Section II-B and the implementation in Section III, we analyze numerically the URC concept for a V2V link underlying a cellular network. The system parameters used for the evaluation are summarized in Table I. A cellular network with 19 hexagonal layout cells and an inter site distance of 1000 meters is assumed as the underlying communication system. The V2V transmission takes place in the central cell with two-tiers of surrounding cells to provide a realistic interference scenario. We assume a use case in which a packet size of 1600 bytes (12800 bits) must be transmitted within 6 ms (D_{AR} = 6 ms), which corresponds to six TTIs of 1 ms each. Due to the short duration of the transmission, shadowing is not considered to be present in the V2V link, whereas fast fading for four different Doppler frequencies is assumed. On the other hand, the the ultra-reliable requirement, P_{UR}, is assumed to be 99.999%.

A. RTL Model

We have considered two different configurations of the RTL in order to deliver the 1600 bytes of application data within six TTIs:

1) Case 1 - Without Retransmissions: in this case, the 1600 bytes are allocated over six TTIs, which corresponds to 2134 bits per TTI using the MCS scheme 18 (QPSK, coding rate 3/4). For this configuration, the minimum SINR required to ensure a successful transmission is 14.3 dB ($\Gamma_{l_1} =$ 14.3 dB). In this case, all six transmissions need to be successfully received in order to guarantee a successful delivery of the whole packet. Assume Γ_i to be the received SINR of the i-th TTI, the condition for a successful transmission for time slot τ (RTL(τ) = 1) is

$$\min_{1\leq i\leq 6} \Gamma_i(\tau) \geq \Gamma_{l_1}.$$

2) Case 2 - With Retransmissions: in this case, the 1600 bytes are distributed over three TTIs whereas the remaining three TTIs are used for retransmissions. This corresponds to 4267 bits per TTI using the MCS scheme 23 (64 QAM, coding rate 5/6). For this configuration, the minimum SINR required to ensure a successful transmission is 19.1 dB ($\Gamma_{l_2} =$ 19.1 dB). Considering the use of one retransmission, the effective SINR for time slot τ after combining two transmissions [8] is given as:

$$\Gamma_i^c(\tau) = \frac{1}{2}(\Gamma_i(\tau) + \Gamma_{i'}(\tau)) + \min(\Gamma_i(\tau), \Gamma_{i'}(\tau)),$$

where Γ_i is the i-th transmission and $\Gamma_{i'}$ refers to its first retransmission. In this case, all three combinations must be correctly decoded at the receiver to fulfill the application requirements, i.e., the transmission for time slot τ is successful (RTL(τ) = 1) if

$$\min_{1\leq i\leq 3} \Gamma_i^c(\tau) \geq \Gamma_{l_2}.$$

B. AEI Model

The AEI mechanism determines AI by comparing the predicted value of the SINR for time slot τ, $\Gamma_p(\tau)$, with the thresholds Γ_{l1} and Γ_{l2} for case 1 and case 2 respectively:

$$\text{AI} = \begin{cases} 1, & \Gamma_p(\tau) \geq \Gamma_{l_k} \quad k = 1, 2 \\ 0, & \text{otherwise.} \end{cases}$$

In the simulation, we consider and compare three different approaches for the SINR prediction performed by the AEI, where the perdicted values for time slot τ are denoted as $\Gamma_p^B(\tau)$, $\Gamma_p^S(\tau)$ and $\Gamma_p^M(\tau)$, respectively.

1) B (Benchmark): A scheme that knows in advance the experienced SINRs of the six transmission TTIs. This scheme represents an upper bound of the achievable availibility as it considers perfect channel estimation. The computation of the predicted SINR for time slot τ can be formulated as

$$\Gamma_p^B(\tau) = \begin{cases} \min\limits_{1\leq i\leq 6} \Gamma_i(\tau), & \text{for Case 1} \\ \min\limits_{1\leq i\leq 3} \Gamma_i^c(\tau), & \text{for Case 2.} \end{cases}$$

2) S (Simple): A simple case that predicts the SINR by taking the minimum SINR of the six TTIs in the previous time slot τ^-. It is important to highlight, that this simple prediction approach does not consider the channel variations between time slots due to fast fading, and therefore, is expected to decrease the performance in fast varying channels. The computation of the predicted SINR for time slot τ in this case is the same independent of the use of retransmissions:

$$\Gamma_p^S(\tau) = \min_{1\leq i\leq 6} \Gamma_i(\tau^-).$$

3) M (Simple with Margin): In order to compensate for prediction errors of the simple prediction approach, which can result in a violation of the ultra-reliable requirement, a certain margin Γ_m is added to the decision threshold, such that:

$$\Gamma_p^M(\tau) = \min_{1\leq i\leq 6} \Gamma_i(\tau^-) + \Gamma_m.$$

TABLE II
PREDICTION METHOD - PROBABILITIES

Freq & Prob		Without Retransmissions			With Retransmissions		
Doppler	Probability	B	S	M	B	S	M
10 Hz	P_1	53.3	55.2	27.9	44.3	26	20.7
	$P_{1\|1}$	100	95.6	99.999	100	99.96	99.999
	$P_{0\|1}$	1.4	4.2	37.4	0.8	24.3	29.5
	Margin	0	0	4.5	0	0	1
33.3 Hz	P_1	48.3	52	1.5	39.7	23.3	5.7
	$P_{1\|1}$	100	84.5	100	100	94.5	99.9999
	$P_{0\|1}$	1.6	10.5	48.5	0.8	23.5	39
	Margin	0	0	14	0	0	6
100 Hz	P_1	38.2	46.9	0.8	37.79	19.3	3.14
	$P_{1\|1}$	100	71.09	99.999	100	93.3	99.999
	$P_{0\|1}$	1.47	10.9	39	0.83	24.4	35
	Margin	0	0	15	0	0	6
200 Hz	P_1	35	43	0.7	34.1	18.7	3.1
	$P_{1\|1}$	100	67	100	100	91	100
	$P_{0\|1}$	1.5	11.5	42	0.81	25.2	37
	Margin	0	0	15	0	0	6

Note that the minimum required Γ_m is found a posteriori based on the results obtained by the simple prediction method and according to the optimization framework defined in Section II, so that the ultra-reliable requirement, $P_{1|1} \geq P_{UR}$, is always satisfied.

Table II contains the relevant probabilities for both transmission schemes (with and without retransmissions) and the three prediction methods. The results show that, independent of the Doppler, it is possible to satisfy the ultra-reliable requirement ($P_{1|1} \geq 99.999\%$) as long as the propagation channel (i.e. the SINR) can be predicted perfectly (prediction method B). On the contrary, the ultra-reliable requirement cannot be satisfied with the prediction method S due to the very simple channel prediction that is performed in this case. The degradation increases significantly for higher values of Doppler as a result of the faster variations of the propagation channel. In particular, $P_{1|1}$ can deteriorate to values below 70% for Doppler frequencies above 200 Hz if retransmissions are not used.

The results also illustrate that the imperfect channel prediction of method S can be compensated by adding a certain margin to the predicted SINR (prediction method M) so that the ultra-reliable requirement is always satisfied. Nevertheless, this is achieved at the cost of lowering the availability of the RTL (P_1). For example, the availability in the case of Doppler frequencies above 200 Hz is below 1% if retransmissions are not used. This behaviour illustrates the trade-off between reliability and availability of the proposed URC concept. In this manner, it would be possible to increase the availability values illustrated in Table II by decreasing the reliability requirements of the application. For example, by extending the maximum acceptable deadline beyond 6 ms or by decreasing the packet size below 1600 bytes, so that more robust modulation and coding schemes and a higher number of retransmissions can be employed.

It is also important to highlight, that a smaller margin for the prediction approach S is required for the case with retransmissions, since the fast channel variations in high Doppler scenarios can be partially compensated by means of improved time diversity. Regarding the availability, it can be seen that for the same amount of resources (six TTIs) the use of retransmisions achieves better performance for high Doppler frequencies, whereas the configuration without retransmissions is preferable for lower values of Doppler. This illustrates the benefits of configuring the RTL according to variability of the propagation channel, which in turn is conditioned by the relative velocity between communicating nodes.

V. CONCLUSION

In this paper, we have proposed a system concept for Ultra-Reliable Communication (URC) based on the combination of a Reliable Transmission Link (RTL) that is optimized to transmit packets successfully within a predefined deadline, and an Availability Estimation and Indication (AEI) mechanism designed to reliable predict the availability of the RTL under given conditions. This concept is motivated by the strict reliability requirements of some applications such as vehicular and industrial applications, and by the fact that wireless communication systems cannot be designed to provide reliability at all times. In this context, the URC concept enables reliability-based applications by accurately indicating the non-availability of the RTL to the application by means of an Availability Indicator (AI), and by enabling the availability of the RTL as often as possible. In summary, the proposed concept lays the foundations for the support of URC in future 5G mobile communication systems, which is considered as key for the introduction of new services such as road safety and industrial applications. Although the preliminary evaluation presented in this paper already shows promising gains, future research regarding the computation of the AI and the configuration of the RTL, as well as practical implementation aspects, are expected to improve further the performance of the concept and possibilitate its integration in 5G systems.

REFERENCES

[1] "METIS - Mobile and wireless communications Enablers for the Twenty-twenty Information Society." [Online]. Available: https://www.metis2020.com/

[2] A. Osseiran, V. Braun, T. Hidekazu, P. Marsch, H. Schotten, H. Tullberg, M. Uusitalo, and M. Schellman, "The Foundation of the Mobile and Wireless Communications System for 2020 and Beyond: Challenges, Enablers and Technology Solutions," in *IEEE Vehicular Technology Conference (VTC Spring)*, 2013, pp. 1–5.

[3] F. Bai, H. Krishnan, and V. Sadekar, "Towards Characterizing and Classifying Communication-based Automotive Applications from a Wireless Networking Perspective," in *Proceedings of IEEE Workshop on Automotive Networking and Applications (AutoNet)*, 2006, pp. 1–25.

[4] F. B. F. Bai and H. Krishnan, "Reliability Analysis of DSRC Wireless Communication for Vehicle Safety Applications," *2006 IEEE Intelligent Transportation Systems Conference*, 2006.

[5] A. Birolini, *Reliability Engineering: Theory and Practice*, 2010.

[6] S. Kay, *Fundamentals of Statistical Signal Processing: Detection theory*, ser. Prentice Hall Signal Processing Series. Prentice-Hall PTR, 1998. [Online]. Available: http://books.google.de/books?id=vA9LAQAAIAAJ

[7] M. Rupp, C. Mehlführer, and M. Wrulich, *Evaluation of HSDPA and LTE: from testbed measurements to system level performance*. John Wiley & Sons, 2011.

[8] M. Gidlund, "Combined packet retransmission diversity and power adjustment scheme for high speed wireless networks." *Journal of Communications*, vol. 3, no. 1, 2008.