Reducing Handover Outage for Autonomous Vehicles with LTE Hybrid Access

Lauridsen, Mads; Kolding, Troels; Pocovi Gerardino, Guillermo Andres; Mogensen, Preben Elgaard

Published in:
2018 IEEE International Conference on Communications, ICC 2018 - Proceedings

DOI (link to publication from Publisher):
10.1109/ICC.2018.8422737

Publication date:
2018

Document Version
Accepted author manuscript, peer reviewed version

Reducing Handover Outage for Autonomous Vehicles with LTE Hybrid Access

Mads Lauridsen[1], Troels Kolding[2], Guillermo Pocovi[2], Preben Mogensen[1,2]
[1]Department of Electronic Systems, Aalborg University, [2]Nokia Bell Labs
Aalborg, Denmark
{ml, pm}@es.aau.dk {troels.kolding, guillermo.pocovi}@nokia-bell-labs.com

***Abstract*—Autonomous vehicle applications require sub-100 ms message latency with a high success probability. While this will be supported by upcoming fifth generation networks, hybrid access technology using multiple LTE connections, may pave the way for the autonomous vehicle applications today. In current LTE networks, handovers often lead to long data interruption. Therefore, we study the handovers and the related data interruption, and how performance can be improved through hybrid access, using redundant transmissions over multiple connections.**

The study is based on drive tests, where four QualiPoc measurement smartphones are connected to different LTE networks simultaneously, performing synchronized ping latency measurements, and recording handover events. The results show that a handover event during a ping measurement, prolongs the average latency from 60-80 ms to more than 200 ms. However, the handovers do not occur simultaneously in the measured LTE networks. Therefore, hybrid access with two connections is shown to reduce the handover outage by a factor 60. Furthermore, the 99.9 %-tile latency is reduced 66 %, by using two simultaneous connections, as compared to the best single network measurement.

I. Introduction

Recently, wireless connectivity for autonomous vehicles (AVs) has received a significant amount of attention. The reason is that wireless transfer of information will enhance the capabilities of the cars, drones, robots, etc. as compared to solely relying on radar, lidar and other sensors mounted on each vehicle. The wireless exchange of information between vehicles and infrastructure must be reliable, and with low latency to provide an improvement on top of the build-in sensors. An example of user safety-oriented requirements is a round trip time (RTT) latency target below 100 ms, with a success probability of at least 99.99 % [1].

The future fifth generation (5G) radio technology targets to provide ultra-reliable, low-latency communication, and thus supports the AVs. However, the 5G technology is yet to be standardized. Therefore, the deployment of 5G, providing sufficient spatial availability to cover national roads, is also not imminent. Thus, it is attractive to re-use the existing LTE infrastructure for AVs. Unfortunately, recent measurements [2] show that commercial LTE deployments, even though they have an average RTT latency of 50-70 ms, will exceed the 100 ms target in 2-5 % of the transmissions. Within 3GPP there is ongoing work to enhance LTE to support reliable and low-latency AV applications, [3], but such deployment upgrades also take time on both the infrastructure and the device side. An alternative, and immediately available solution, is to combine multiple readily-available mobile networks, using multiple operators and multiple technologies - commonly referred to as *hybrid access*. Such solutions can provide lower cost and faster time-to-market, as compared to upgrades of LTE and future 5G.

The hybrid access (or multipath) concept has been studied extensively for more than 10 years for both link, IP, transport, and application layers [4]. A key focus area has been fixed-mobile convergence [5], where a device can leverage both wired and wireless connections. Moreover, there is significant interest in combining multiple wireless technologies, to provide enhanced connectivity for mobile devices. The work on vehicular communication includes UDP-based lab experiments on WiFi targeting lower latency [6], and application-based measurements on 3G and WiFi targeting throughput improvements [7]. However, the introduction of Multipath-TCP (MP-TCP) [8] has triggered multiple studies on hybrid access for AVs. The reason is that this extension to the Transmission Control Protocol (TCP), allows multiple paths to be utilized easily. The goal has mainly been to improve the throughput, e.g. by use of LTE and WiFi in a vehicle to infrastructure network [9], and by use of multiple simulated wireless interfaces for an unmanned aerial system [10].

In addition to increasing the data rates, the hybrid access work in [11] shows, how the latency performance is also improved due to handovers occurring at different times in different wireless systems. By transferring the data redundantly, (data duplication in 3GPP terms), in multiple wireless systems, the overall data transfer becomes more reliable, even though one of the systems may not be able to deliver any service at a given time. In this regard, [12] performs measurements using two LTE connections, on the same carrier, and a redundant MP-TCP scheduler on-board a train. The measured average latency is reduced by half, and packet drops are nearly uncorrelated. However, limited details on the handover events and the latency distribution are presented [12].

The objective and contribution of this work is an experimental study on how hybrid access, using multiple LTE operators and multiple phones connected to the same operator, impacts handover and latency performance. We measure ping latencies on QualiPoc measurement smartphones, which simultaneously collect handover event statistics. Since the phones do not support hybrid access, the obtained measurement traces are

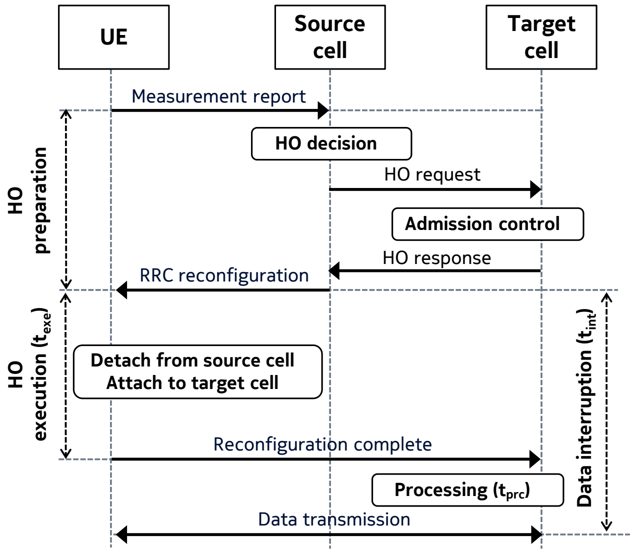

Fig. 1. Break-before-make intra-LTE handover procedure. Based on [13].

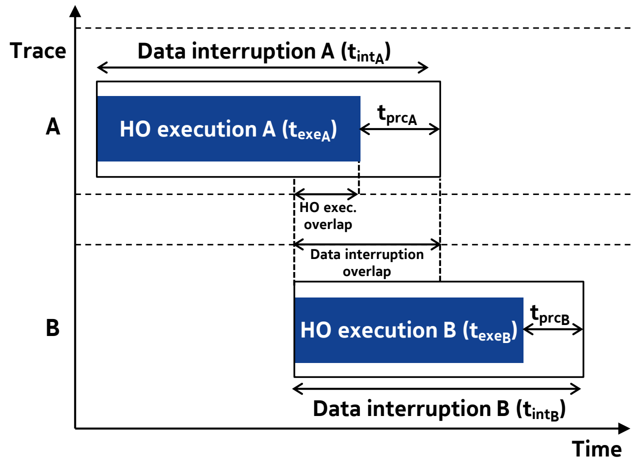

Fig. 2. Time trace of handover (HO) events for two LTE devices.

post-processed, in order to emulate the behavior of redundant data transmission over multiple wireless interfaces.

The paper is structured as follows: in Section II the LTE handover events and how they impact latency performance are described. Next we explain how the hybrid access approach can minimize the handover outage. In Section III the measurement equipment & scenario, and the post-processing methodologies for time synchronization and hybrid access emulation are described. We present our results in Section IV, which is followed by the discussion and conclusion in Sections V and VI, respectively.

II. LTE Handover Events and Hybrid Access

Handovers occur when the LTE device reconfigures its connection from the current serving cell to a new cell, e.g. due to changes in coverage or load conditions [13]. Since LTE employs break-before-make handovers, there is a data interruption when the device is detaching from the old, and attaching to the new cell [14]. Note this study is limited to intra-LTE handovers, although handovers between radio access technologies are also of interest.

As illustrated in Fig. 1, the data interruption starts with the handover execution phase. The handover execution is initiated, when the device receives the Radio Resource Control (RRC) Connection Reconfiguration message, which indicates the need for a handover from the current serving cell. The phase completes, when the device sends the RRC Connection Reconfiguration Complete upon successful random access to the new cell [13]. Due to message processing in the target cell, and transmission latencies [15], the data interruption continues after the handover execution has completed.

Hybrid access enables a device to duplicate its packets across all interfaces. As discussed in the introduction, the method can thus be used to improve the throughput and latency, by benefiting from the interfaces providing different performance at different points in time and space [9]–[12]. Furthermore, hybrid access may also reduce the probability that there is no data connection available at all. The reason, according to our hypothesis, is that multiple LTE interfaces are not likely to experience handovers at the same time. Therefore, LTE hybrid access may be a useful method to minimize the impact of the data interruption phase, e.g. for AVs.

This work evaluates the potential of LTE hybrid access with specific focus on mitigating the impact of handovers. In order to quantify the observations, a performance metric is defined for handover outage. Specifically, ρ describes the percentage of time a device experiences data interruption due to handover:

$$\rho = \frac{1}{t_{\text{tot}}} \cdot \sum_{n=1}^{N} t_{\text{int}}(n) \qquad [-], \qquad (1)$$

$$t_{\text{int}}(n) = t_{\text{exe}}(n) + t_{\text{prc}}(n) \qquad [\text{s}], \qquad (2)$$

where t_{tot} is the total duration of the measurement [s], N is the number of experienced handover events [-], $t_{\text{int}}(n)$ is the data interruption [s], $t_{\text{exe}}(n)$ is the observed execution time [s], and $t_{\text{prc}}(n)$ is the observed additional delay [s] for the nth handover event.

In the case of hybrid access handover outage, the term t_{int} in (1) is the time when all active interfaces are simultaneously interrupted. This concept is illustrated in Fig. 2, where a time trace of the handover events experienced by two different mobile interfaces (A & B) is sketched. The handover execution overlap between two or more devices is obtained by inspecting the RRC Connection messages. However, as observed in Fig. 2, the actual data interruption overlap is larger due to the processing time T_{prc}. The dependency on T_{prc} will be further analyzed in Section IV. Note the metric ρ in (1) is agnostic to the specific type of hybrid access in use.

In addition to the handover outage ρ, the RTT latency is a key performance indicator for this study. In particular, the system reliability, defined as the probability that the latency target is achieved [1], is of interest.

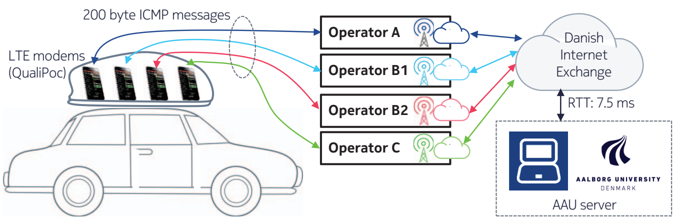

Fig. 3. Hybrid access handover and latency measurement setup.

III. Measurement & Processing Methodology

A. Equipment & Scenario

The measurements are performed using four Samsung Galaxy S5 QualiPoc measurement smartphones. Each phone sends a 172 bytes echo request packet, using the Internet Control Message Protocol (ICMP) with 200 ms interval. Including the ICMP and IP header, the payload is 200 Bytes. This traffic corresponds to typical message payloads and transmission intervals of AV applications [1]. The phones ping towards a server at Aalborg University, which is connected to the Danish Internet Exchange via a low-latency 10 Gb/s fiber [2]. According to [2] the round trip time between Aalborg University and the Internet Exchange is 7.5 ms. Even though we use 172 bytes pings, as compared to 128 bytes in [2], the estimate is deemed valid. The QualiPoc measurement phones include an application that enables scheduling of the aforementioned ping job. In addition, the application also records LTE RRC messages, System Information Blocks, and radio layer measurements. During the measurements, the phones are forced to the LTE technology, but they select the carrier frequency according to the operator's coverage and traffic steering policy. Besides the measurement software, the phones run a minimum number of Android applications and background processes. The hybrid access measurement setup is illustrated in Fig. 3.

Two of the phones are connected to the same operator (B1 & B2 in Fig. 3), to study the potential of device location diversity as in [12]. The physical separation of the two phones is $\sim$20 cm. The two other phones are connected to the remaining two main operators of Denmark (A & C), enabling the study of multi-operator diversity. Some operators partially employ site-sharing, but their traffic steering and mobility control is expected to result in different ping latencies and handover events. The measurements were performed back and forth on the highway between Aalborg and Frederikshavn, with a speed of approximately 100 km/h. The $\sim$130 km measurement route is illustrated in Fig. 4.

B. Measurement Synchronization & Processing

The QualiPoc phones provide timestamped measurement logs, including ping latency traces and RRC messages. The phones are time synchronized, using a broadcasted Bluetooth synchronization signal from a QualiPoc tablet, which acts as a master for the QualiPoc phones. The absolute time synchronization of each phone is in the order of 100 ms

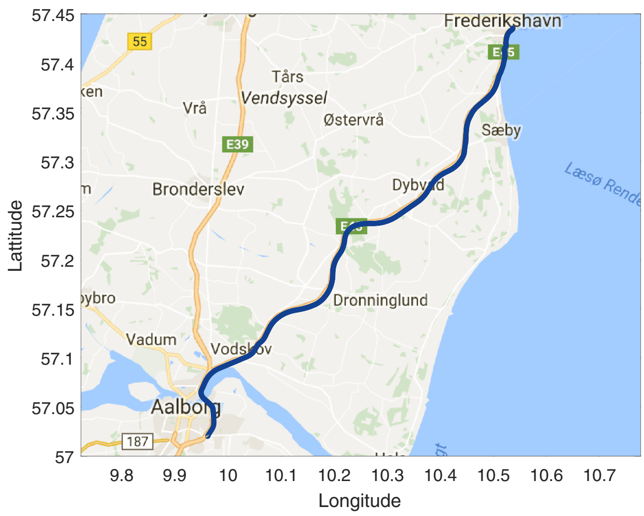

Fig. 4. Measurement route (in blue) in Northern Jutland, Denmark.

[16]. The resulting absolute time difference of up to 200 ms between two phones is critical, when comparing the handover events, because each event on average is 20-40 ms (see the measurement results in Fig. 5). Since the handover event is shorter in time than the absolute time offset, the occurrence of handover overlaps, defined in Fig. 2, may not be detected properly. To handle this measurement limitation, the phones are time synchronized through post-processing in Matlab. The starting point, is to search for overlapping handovers between pairs of phones i.e. cross-correlating two measurements. This is achieved by sweeping a time offset of -100 ms to 100 ms, with 1 ms granularity, in one of the phones' trace. For each time offset the number of overlapping handovers, and the accumulated handover execution, and data interruption times are logged. This procedure is repeated for all six phone pair combinations. The time offset for each phone is then fixed to the sweep time value, where the worst case performance, i.e. the highest accumulated handover execution time for the six combinations, is observed. The fixed absolute time offset per phone is then used for the ping latency processing.

To examine the impact on ping latency, a redundant hybrid access scheduler is implemented in Matlab. The emulator assumes that the same data (ping packet) is sent on all available interfaces, i.e. the four measurement phones. Having calibrated the absolute time offset between the phones, a common time vector, starting with the first measurement, amongst the phones is generated. The resolution of the time vector equals the ping interval of 200 ms. Next the four phones' traces are compared with the common time vector. Ping latency samples from each phone are linked to the common time vector, based on their time of occurrence, (i.e. the time where the phone sends the ICMP Echo Request). Using the redundant scheduler, the lowest latency sample across the four traces is reported in the following Section.

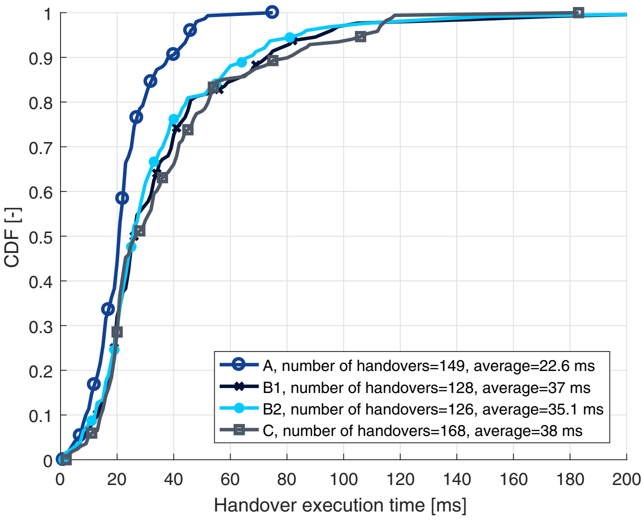

Fig. 5. Handover execution time per operator.

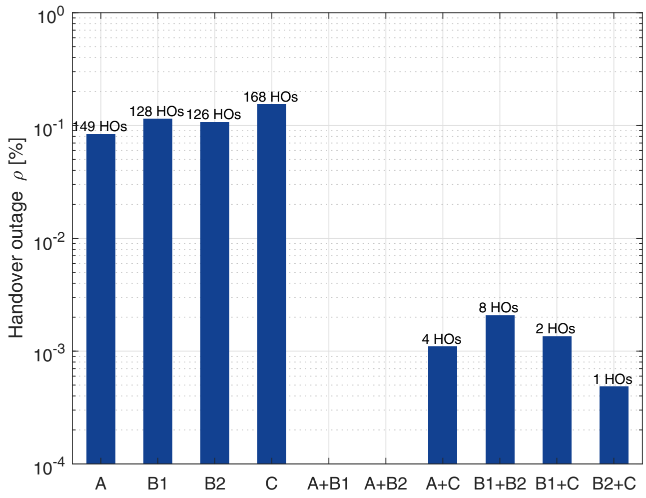

Fig. 6. Handover outage ρ and number of handovers (HOs) for $t_{\mathrm{prc}} = 0$ ms.

TABLE I
HANDOVER OUTAGE ρ FOR COMBINATIONS OF OPERATORS AND t_{prc}.

	Handover outage ρ [%]		
Processing time t_{prc}	0 ms	5 ms	10 ms
Single operator	0.0838-0.1540	0.1024-0.1742	0.1209-0.1945
Two different operators	0.0005-0.0014	0.0006-0.0016	0.0007-0.0021
Same two operators	0.0021	0.0030	0.0041
Three operators	-	-	-

IV. Results

This Section contains the measured handover execution times, the worst case estimate of overlapping handovers including the processing time, ping latencies per operator, ping latency with and without a handover event, and finally an estimate of the latency, when redundantly transmitting data over the multiple wireless interfaces. The operators' names are hidden, but known to the authors.

A. Handover Performance

The empirical cumulative distribution functions (CDFs) for the handover execution times of the four operators are shown in Fig. 5. During the 130 km drive the phones experienced 126-168 handovers and 0 handover/radio link failures. Operators B1, B2, and C on average require $\sim$37 ms to perform the handover procedure, while operator A on average completes it in 23 ms. These observations are in line with [2].

Even though operator A performs more than 20 additional handovers, as compared to operators B1 and B2, the overall handover outage ρ, defined in (1), is 20-25 % lower as illustrated in Fig. 6. Operator C has the longest average handover execution time and highest number of handovers, and thus the handover outage is 45 % higher than operator A. The handover outage ρ for all operators is in the order of 0.1 %. That prevents the operators from delivering reliable real-time communication, because the handover and the data transfer may coincide.

Fortunately, the use of redundant hybrid access entails that the data interruption can be mitigated. The reason is that it only occurs, if the data interruption events of the two connections overlap, as illustrated in Fig. 2. The handover outage for the operator combination pairs using $t_{\mathrm{prc}} = 0$ ms is given in Fig. 6. On average the handover execution times from two traces overlap for 13.7 ms. The total number of handovers, and the handover outage are greatly reduced, and for two out of six pairs completely avoided. Furthermore, Fig. 6 shows that the lack of site diversity for B1+B2, impacts the ability to avoid simultaneous handovers, resulting in a handover outage of 0.002 %. This may prevent hybrid access to a single operator (B) from providing reliability $\geq$99.99 %, but it is a factor 50 improvement compared to the 0.1 % outage level per phone.

As illustrated in Fig. 2, the data interruption depends on not only the handover execution time, but also the processing time after finishing the handover execution phase. In table I the handover outage ρ is estimated for processing time t_{prc} of 0, 5, and 10 ms. The 0 ms case corresponds to the pure handover execution phase shown in Fig. 6, while 10 ms is the commonly assumed performance in current LTE deployments [15]. The 5 ms result thus provides a view on what can be realized, if LTE or 5G is optimized.

In general, the single operator cases result in the largest handover outage. Combining the two connections from the same operator (B1+B2), reduces the handover outage by at least a factor 30. The combination of two different operators, provides close to a factor 60 improvement for all t_{prc} settings. Along the 130 km measurement route, there was no handover overlap across the three operators (A+B1+C).

B. Latency Statistics

Fig. 7 shows the CDF of the latency distribution per operator. The average values are in the range 56-113 ms, and in line with previous observations [2].

Operator B, (traces B1 and B2), provides $\leq$100 ms latency for $\sim$98 % of the samples, while operators A and C are 5 %-points lower. The number of samples experiencing latencies

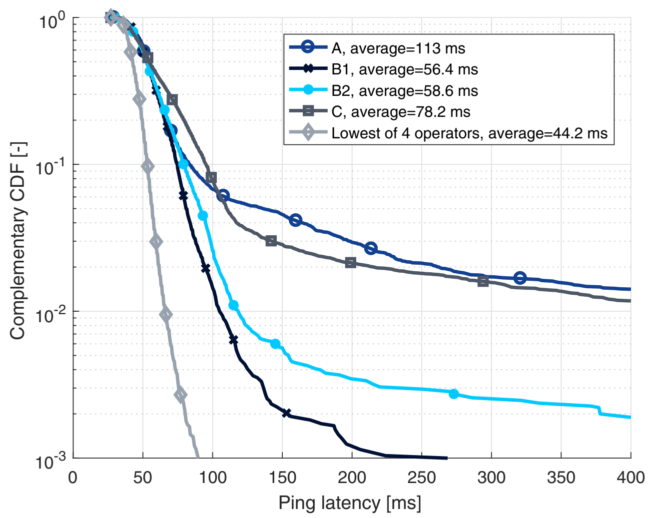

Fig. 7. Ping latency distribution per operator.

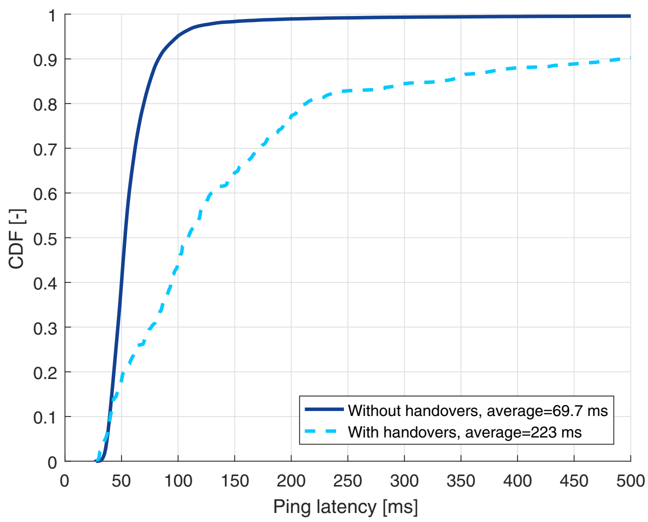

Fig. 8. Ping latency measurements with and without handover overlapping.

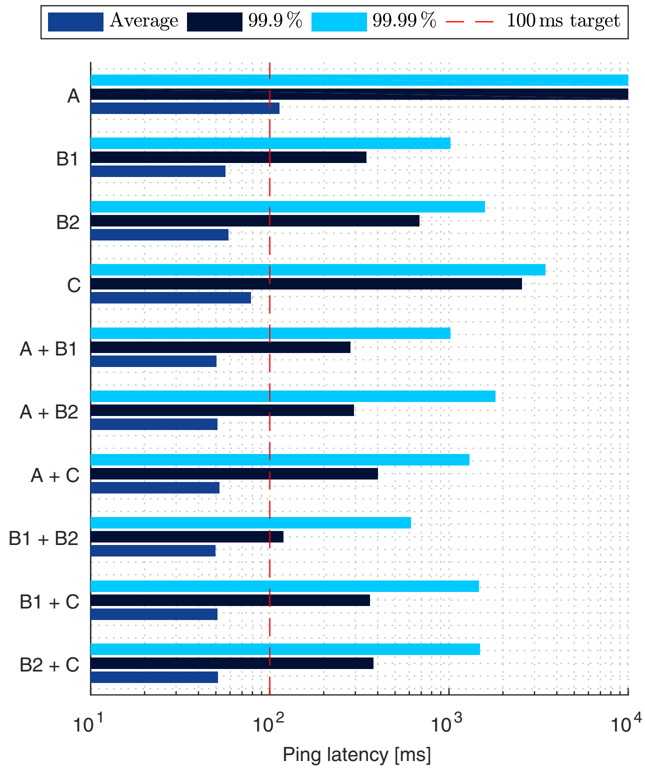

Fig. 9. Ping latency statistics for combined operator measurements.

in the 300-400 ms range is ten times higher for operators A and C as compared to B. The combination of B1 and B2, even though they originate from the same operator, may thus be the most favorable for providing reliable real-time communication. However, connecting to the same operator provides less site diversity, which affects the number of simultaneous handover events, as determined in Section IV-A.

The grey line with diamond markers in Fig. 7 is based on the lowest latency per time sample across the four traces. This results in an average latency of 44 ms, which is $\geq$20 % lower than the best individual operator (B1). Furthermore, more than 99.94 % of the samples experience $\leq$100 ms latency. However, it may not be feasible to utilize all four connections in a practical setup.

A key reason for the prolonged ping latency, is the handover events that result in data interruption. In Fig. 8 the latency results for all four traces are grouped, based on whether a handover overlapped with the ping latency measurement. Without handovers, the average latency is $\sim$70 ms, while 95 % of the samples are below 100 ms. However, when a handover event occurs simultaneously with the latency measurement, the average latency exceeds 200 ms. This underlines the advantage of reducing handover outage by use of hybrid access.

C. Combined Pings

As previously mentioned in Section III, samples were collected per operator along the 130 km measurement route. However, not every sample is valid due to ping timeouts, which are caused by e.g. handover, network load, and coverage holes. On average 69 % of the samples contains a ping measurement from all four operators. In 25 % of the cases, three operators provide a ping latency measurement, while 4 % and 0.7 % contain just two or one operator(s), respectively.

Fig. 9 shows the average, 99.9 %-tile, and 99.99 %-tile ping latency for the four operators, and emulated redundant combinations of operator pairs. The red dashed line is the 100 ms target. The average latency of operators B1, B2, and C fulfills the target (as also shown in Fig. 7), but the 99.9 %-tile does not. This is also the case for the redundant hybrid access combinations of operator pairs, but B1+B2 achieves 119 ms at the 99.9 %-tile, which is a 66 % reduction of the single operator results. However, there is still a significant gap to fill, as all combinations but B1+B2 exceed 1 s at the 99.99 %-tile. By combining three operators, the 99.9%-tile is 107-120 ms, while the 99.99 %-tile exceeds 350 ms. Utilizing all four operators, results in 99.9 % and 99.99 %-tiles at 90 ms and 135 ms, respectively.

V. DISCUSSION

The vision of our work is a hybrid access solution that combines legacy and future cellular radio technologies, multiple network operators, and potentially also other wireless technologies, such as WiFi and MulteFire. In this initial work, LTE was studied to evaluate the potential of operator and device location diversity. Given the observations in this study, it is determined that hybrid access using multiple LTE connections, can provide a significant latency reduction for both average and tail users, by limiting the impact of handovers. However, improvements in terms of optimized handover methods, and lower core network latencies are still required [2].

In 3GPP new handover techniques, such as synchronized handover, RACH-less handover, and multi-connectivity, are being studied for LTE-Advanced Pro and 5G [14]. Each of the techniques target to minimize the handover outage, which thus may become a less critical point for the future AV applications. However, in current LTE networks the hybrid access method proves useful for mitigating this issue. In addition, the results in Table I indicate the benefit of further reducing the processing time t_{prc}, defined in Fig. 1.

In this study a redundant scheduler was emulated. It provides low latency and high robustness towards handovers, but the cost of transferring twice the data is also high. In addition, the simultaneous use of multiple wireless radios, will impact the energy consumption [11]. Thus, it is important to evaluate, whether other schedulers can be developed to limit the number of transmissions, while maintaining the good latency performance observed in our measurement. Furthermore, it is important to consider, whether the currently used redundant scheduler, must be activated at both client and server side.

VI. CONCLUSION

Autonomous vehicle applications rely on wireless communication with low latency and high reliability. In this respect the widely deployed cellular LTE technology may be applicable. In this work, we measure 20-40 ms intra-LTE handover execution times in three commercial networks, using QualiPoc measurement smartphones. The ensuing data interruption time, is a critical factor for delivering reliable real-time communication. Using the QualiPoc phones for ping measurements, we observe that the average latency increases from 60-80 ms, when a handover is not present, to more than 200 ms, when a handover occurs during the ping measurement.

In a 130 km test drive in Denmark, the phones are without data connection 0.1 % of the time due to $\sim$150 handovers per network. However, we observe that the handovers do not occur simultaneously across the three LTE networks. This points to the potential of hybrid access, where data is redundantly transmitted through multiple wireless interfaces. Using post-processing we emulate hybrid access with pairwise combinations of the three networks, and note the handover outage is reduced 60 times, only resulting in 0-4 overlapping handovers. Two of the QualiPoc phones were connected to the same LTE network, and this hybrid access combination results in 8 overlapping handovers, and reduces the outage by a factor 30, as compared to the best single connection case.

The results also illustrate the benefit of reducing the message processing time in the handover target cell. Decreasing the time from 10 ms to 5 ms, reduces the overall handover outage 10-25 %, depending on the operators used.

Having observed the significant improvement in handover outage, the latency measurements are combined across the operators. While the best combination of two ping measurements only results in $\sim$10 % reduction of the average latency, the 99.9 %-tile is reduced by 66 %, and the 99.99 %-tile by 40 %. Thus hybrid access for LTE improves the communication latency and reliability for autonomous vehicle applications, by reducing the handover outage.

Future work includes measurements in diverse scenarios, using multiple transport layer protocols and traffic models.

ACKNOWLEDGMENT

Thanks to the Master of Science students Henrik H. Rasmussen, Kasper W. Mortensen, Rasmus S. Mogensen, and Christian Markmøller (Department of Electronic Systems, Aalborg University) for collecting the measurement data.

REFERENCES

[1] G. Pocovi, M. Lauridsen, B. Soret, K. I. Pedersen, and P. Mogensen, “Automation for on-road vehicles: Use cases and requirements for radio design,” in *IEEE VTC Fall*, Sept 2015.

[2] M. Lauridsen, L. C. Gimenez, I. Rodriguez, T. B. Sørensen, and P. Mogensen, “From LTE to 5G for Connected Mobility,” *IEEE Communications Magazine*, vol. 55, no. 3, pp. 156–162, March 2017.

[3] 3rd Generation Partnership Project, “Study on latency reduction techniques for LTE,” 36.881 Release 14.0.0, July 2016.

[4] M. Li, A. Lukyanenko, Z. Ou, A. Yl-Jski, S. Tarkoma, M. Coudron, and S. Secci, “Multipath Transmission for the Internet: A Survey,” *IEEE Communications Surveys Tutorials*, vol. 18, no. 4, pp. 2887–2925, June 2016.

[5] K. Samdanis, F. Leitao, S. Oechsner, J. R. I. Riu, R. D. C. Ros, and G. Fabregas, “From Interworking to Hybrid Access Systems and the Road toward the Next-Generation of Fixed-Mobile Convergence,” *IEEE Communications Standards Magazine*, vol. 1, no. 1, March 2017.

[6] N. Chiba, M. Ogura, R. Nakamura, and H. Hadama, “Dual transmission protocol for video signal transfer for real-time remote vehicle control,” in *Asia-Pacific Conference on Communication*, Oct 2014, pp. 315–320.

[7] P. Deshpande, X. Hou, and S. R. Das, “Hybrid Wireless Access for Vehicular Networking: Integrating 3G and Metro-Scale WiFi,” 2010.

[8] A. Ford, C. Raiciu, M. Handley, and O. Bonaventure, “RFC 6824: TCP Extensions for Multipath Operation with Multiple Addresses,” Internet Engineering Task Force, Jan 2013.

[9] J. Mena, P. Bankole, and M. Gerla, “Multipath TCP on a VANET: A Performance Study,” in *SIGMETRICS*. ACM, June 2017.

[10] R. M. N. Chirwa and A. P. Lauf, “Performance improvement of transmission in Unmanned Aerial Systems using multipath TCP,” in *IEEE ISSPIT*, Dec 2014.

[11] C. Paasch, G. Detal, F. Duchene, C. Raiciu, and O. Bonaventure, “Exploring Mobile/WiFi Handover with Multipath TCP,” in *SIGCOMM CellNet Workshop*. ACM, Aug 2012.

[12] A. Frommgen, T. Erbshuer, A. Buchmann, T. Zimmermann, and K. Wehrle, “ReMP TCP: Low latency multipath TCP,” in *IEEE International Conference on Communications*, May 2016.

[13] 3rd Generation Partnership Project, “Evolved Universal Terrestrial Radio Access and Evolved Universal Terrestrial Radio Access Network; Overall description; Stage 2,” 36.300 Release 14.4.0, Oct 2017.

[14] L. C. Gimenez, “Mobility Management for Cellular Networks: From LTE Towards 5G,” PhD thesis, Aalborg University, May 2017.

[15] T. Kolding, L. Gimenez, and K. I. Pedersen, “Optimizing Synchronous Handover in Cloud RAN,” in *IEEE VTC Fall*, Sep 2017.

[16] Rohde & Schwarz Customer Support, “Synchronization accuracy between FreeRider III slaves,” Private communication, Aug. 2017.