Adding the Next Nine: an Investigation of Mobile Broadband Networks Availability

Ahmed Elmokashfi, Dong Zhou, Džiugas Baltrūnas

Simula Research Laboratory, Fornebu, Norway

ahmed@simula.no, dong@simula.no, dziugas@simula.no

ABSTRACT

The near ubiquitous availability and success of mobile broadband networks has motivated verticals that range from public safety communication to intelligent transportation systems and beyond to consider choosing them as the communication mean of choice. Several of these verticals, however, expect high availability of multiple nines. This paper leverages end-to-end measurements to investigate the potential of current mobile broadband networks to support these expectations. We conduct a large-scale measurement study of network availability in four networks in Norway. This study is based on three years of measurements from hundreds of stationary measurement nodes and several months of measurements from four mobile nodes. We find that the mobile network centralized architecture and infrastructure sharing between operators are responsible for a non-trivial fraction of network failures. Most episodes of degraded availability, however, are uncorrelated. We also find that using two networks simultaneously can result in more than five nines of availability for stationary nodes and three nines of availability for mobile nodes. Our findings point to potential avenues for enhancing the availability of future mobile networks.

CCS CONCEPTS

• **Networks** → **Network measurement**; **Network reliability**; **Mobile networks**;

KEYWORDS

Interdependent networks; correlation networks; ultra-reliable communication

1 INTRODUCTION

Mobile Broadband (MBB) is becoming increasingly ubiquitous and an economic boon with mobile data traffic projected to increase by almost tenfold between 2015 and 2020 [11]. This success has encouraged industries that range from public safety communication to transportation systems to retail and beyond to use or consider using MBB for their communications. Many of these industries, however, expect carrier-grade availability of multiple nines, but present commercial MBB networks are hardly able to guarantee

more than two nines [4, 35]. The need for a high level of availability and reliability is likely to become even more central as services like smart grid communication, intelligent transportation systems, driver-less cars, and tactile Internet become mainstream [18, 20].

Today, use cases with strict availability requirements largely opt for operating their own communication networks. For example, most countries use specialized Terrestrial Trunked Radio (TETRA) and TETRAPOL networks to facilitate public safety communications [45, 46]. Well designed TETRA networks provide up to 99.95% availability [28] but are expensive to build, use specialized handsets, and are narrow band. These limitations have prompted governments around the world to seek alternatives. The US and Canada have decided to build dedicated national LTE networks [1, 19]. The UK has followed a more radical approach by deciding to depend on a commercial MBB operator starting from mid-2017 [21]. Building a specialized MBB network, while suitable in some cases, will not scale to cater for all services with tight availability budget. Also, many service providers will not be able to build such specialized networks, potentially stifling innovation. This has motivated industry and academia to put offering ultra high availability as one of 5G key targets [33, 37].

In this paper, we leverage end-to-end measurements to investigate the potential of current MBB networks to support services with high availability expectations, i.e., at least 99.9%. Our goal is threefold: (1) provide an empirical evaluation of the availability offered by today's commercial networks, which can inform vertical industries and developers; (2) highlight potential areas for improvement as an input to the ongoing discourse for defining 5G; and (3) contribute a methodology that can be leveraged for quantifying the availability of mobile networks.

Assessing MBB availability is a non-trivial task for two main reasons. First, there is no single metric that fully captures MBB availability and reliability. Hence, we need to use several metrics like data session stability, packet loss, and end-to-end delay to be able to characterize MBB availability. Second, MBB networks tend to share infrastructure, for historical and practical reasons, forming a complex temporal-spatial network of networks. This sharing can range from national roaming where an operator uses another operator's network in areas where it does not have base stations to the sharing of regional or long-haul transport infrastructure. The impact of these dependencies is often hard to gauge, since most of what is shared are physical links that can not be probed using traditional IP measurement tools like ping and traceroute. Furthermore, infrastructure sharing extends the task of assessing MBB availability beyond a single network.

We conduct a large-scale measurement study of network availability and show how it is influenced by infrastructure sharing.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

MobiCom '17, October 16–20, 2017, Snowbird, UT, USA

© 2017 Association for Computing Machinery.

ACM ISBN 978-1-4503-4916-1/17/10...\$15.00

<https://doi.org/10.1145/3117811.3117842>

We also explore the potential of using several operators simultaneously to increase the overall availability even when operators share infrastructure. We focus on two metrics for characterizing availability: the stability of user equipment (UE) data connectivity to the mobile network and packet loss. Measurements used in this study are conducted using the NorNet Edge (NNE) measurements infrastructure [30]. NNE consists of hundreds of nodes for measuring MBB networks in Norway. Each node is connected to up to five operators. These nodes are stationary and hosted in places like schools and government offices. They are distributed to be representative for (indoor and stationary) users in urban/suburban areas. Measurements used in this study span three years from 2014 to 2017. The size and spread of NNE allow us to correlate measurements from different locations and networks in order to gain insights into interdependent failures. In addition, we use a set of mobile nodes on trains to evaluate the potential of using several operators when users are moving.

Our findings confirm the viability of end-to-end measurements in assessing network availability. We find that the centralized architecture of MBB networks is a major cause of correlated service outages within a single operator. We also find that causes of correlated packet loss episodes often lie in the mobile core network and the Internet. Overall, shared infrastructure contributes minimally to unavailability in MBB networks, but it can also result in outages that impact a large geographical area. Consequently, we find that combining connections from different operators can greatly boost MBB availability. In summary, we make the following contributions:

- We present, to the best of our knowledge, the most comprehensive study of MBB availability both in terms of duration and number of measurement probes. Further, we present a first attempt on studying the impact of shared infrastructure on MBB availability.
- We show that correlating measurements from geographically diverse probes can be leveraged for inferring interdependencies between operators.
- We quantify the benefit of combining two operators for both stationary and moving cases.
- We suggest possible ways for enhancing MBB connections availability — a key target of 5G.

2 MBB AVAILABILITY

Mobile networks are stateful, meaning that a user equipment (UE) must establish a data session to the mobile core network before it engages in any data exchange. This data session is a virtual tunnel between the UE and the packet gateway in the core network and is called a Packet Data Protocol (PDP) context and Evolved Packet System (EPS) bearer in UMTS and LTE networks, respectively. A data session is established on demand in UMTS. LTE is, however, an all-IP network. Thus, an attached UE keeps a default EPS bearer always established. Accordingly, the most basic measure of mobile service availability is the ability to maintain a stable data session. Once the data session is established, the next measure is the ability to reliably send data over it. The latter can be extended further to evaluate whether the network is able to offer a predictable and acceptable level of performance for a specific application [26]. In

this paper, we focus on data session stability and packet loss as key metrics for assessing MBB reliability.

2.1 Causes of data session downtime and packet loss

A UE can experience packet loss or a complete loss of network connectivity for a number of reasons. Degraded wireless link performance may trigger the radio access network (RAN) to disconnect attached UEs. This can also happen when the UE and the network fail to successfully complete a horizontal handover (i.e. between different cells or basestations), inter-frequency handover (e.g. a handover from LTE-800MHz to LTE-2600MHz), or inter-RAT handover (e.g. a handover from UMTS to LTE). Failures of RAN and transport equipment as well as congestion in the RAN and transport can also lead to connection failures. Furthermore, outages that involve central components in the core network, like the packet or signaling gateways, can result in many users losing their connectivity.

Outages impact, in terms of affected users and duration, varies depending on underlying root causes. A poor wireless link impacts a few users. A failure at the Home Subscriber Server (HSS) can prevent many UEs from attaching to the network causing a large scale denial of service [48]. A fiber cut or a failure of an optical cross connect in the transport network can isolate a number of base stations. This can also affect customers of several MBB operators when the affected infrastructure is shared [12, 34]. The impact on user experience varies depending on outages duration. For example, a TCP session can recover by resending lost segments, if an outage lasts less than its retransmission timer (RTO). Note that the minimum RTO value is 1 second. However, longer outages result in TCP timeouts that can last up to 1 minute [40]

2.2 Metrics and approach

Data session downtime, which we refer to in the following as *downtime*, reports on MBB data session stability, while packet loss measures the usability of an established data connection. We combine these two metrics to quantify the *overall availability* (OA) of an MBB data connection. OA is defined as follows:

$$OA = 100 * (1 - c) * (1 - p) \quad (1)$$

Where c is the fraction of time the data connection is down and p is the fraction of lost packets. In this paper, we focus on downtime, packet loss, and OA as key metrics for assessing MBB availability. We base our assessment on three-years worth of measurements of four commercial Norwegian MBB operators. These measurements were collected by a large number of measurement nodes that are spread across Norway. We describe the measurement setup in the subsequent section.

3 MEASUREMENT INFRASTRUCTURE

This section describes the measurement setup and methodology.

3.1 NorNet Edge

NNE is a unique dedicated infrastructure for the measurement and experimentation with MBB networks. It comprises over a hundred stationary measurement nodes that are geographically spread in more than 100 municipalities all over Norway as well as set of

Figure 1: The measured networks.

backend servers. The nodes distribution reflects the population distribution with bias towards urban areas. More than half of the nodes are deployed in the three largest cities in Norway where a quarter of the population lives. This infrastructure has been operational since July 2013 ¹.

A NNE node is a custom-made single board computer with one Fast Ethernet port, and 7 on-board USB ports. It runs a standard Debian Linux and a set of customer tools to manage connectivity to MBB networks. Each node is connected to up to four UMTS/LTE operators and one CDMA2000 1x Ev-Do network ², using standard subscriptions. For the UMTS/LTE operators, the connections were, in the early days of NNE, through E3131 and E353 Huawei USB modems which support up to HSPA+. Later as LTE deployment picked up, NNE moved to using E392 Huawei USB modems which support LTE. Almost all NNE modems have supported LTE since the end of 2015. The modems are configured to prefer the best supported radio access technology (RAT). In addition, there are seven NNE nodes that are placed aboard NSB ³ regional and inter-city trains. The routes followed by these trains traverse a reasonable mix of urban and rural areas. The measurements collected aboard the trains closely mimic the user experience of the train passengers. We use the GPS location data from the train's fleet management system to identify the location of NNE measurement nodes and train speed during the measurements. The GPS locations are updated every 10 to 15 seconds in the fleet management system. In this paper, we only use measurements from trains to quantify multihoming benefits when UEs are moving (see Sec. 7). For more details about the measurement setup on trains please refer to our previous work [6]. The NNE backend contains the server side of the measurements. It also includes a set of servers for managing and monitoring the nodes, and for storing measurement data.

3.2 Measured networks

Nodes in the NNE platform are connected to up to four UMTS/LTE networks, which are Telenor, Telia, Tele2 and Network Norway. These networks are not completely independent, they rather have a set of intricate dependencies as depicted in Fig. 1. All four operators maintain independent core networks but vary in their RAN footprint and transport connectivity, i.e., the physical network that interconnects the RAN and the core network. Next, we describe the intricate dependencies that bind these networks.

- *Telenor* maintains a nation-wide radio access network (RAN), a set of regional transport networks, and a nation-wide IP core network that consists of two identical yet separate fiber rings [17]. The transport networks interconnects Telenor's base station with its core IP network which interconnects them with the mobile core network.
- *Telia* ⁴ maintains a nation-wide RAN, but uses Telenor's transport and IP core networks for interconnecting its RAN and mobile core network.
- Before Q1 2015 *Tele2* and *Network Norway* together owned and operated a RAN called *Mobile Norway*. This RAN did not have a nation-wide coverage. When outside their home network, Tele2 customers used Telia's RAN and Network Norway customers used Telenor's RAN. Mobile Norway mainly used a transport and core IP network called *Broadnet* for interconnecting its base stations and mobile core. Telia acquired Tele2 in Q1 2015 resulting in the decommissioning of Mobile Norway and Tele2 brand. Further, Network Norway switched to completely using Telia's RAN. These changes are illustrated by the dashed and solid lines in Fig. 1.

These complex relationships together with the multi-homed nature of NNE present a unique opportunity for assessing correlated failures and thus the impact of infrastructure sharing.

3.3 Measurement methodology and data set

A NNE node attempts to maintain an active data session at all times and immediately attempts to restore failed sessions. For measuring downtime, we record all timestamps where an active data session fails as well as all corresponding restoration timestamps. For measuring packet loss, an NNE node sends a 20-byte UDP packet every second to an echo server that is part of the NNE backend which echoes the packet back to the node. A packet is considered lost, if no reply is received within 60 seconds ⁵. We inspect the server side's system logs to remove all loss episodes that are caused by system artifacts, e.g., server reboots. We also disregard loss episodes that are likely caused by congestion and failures in the NNE backend upstream provider's network. More specifically, we remove all loss episodes that simultaneously impact 10% of connections or more to three different operators. We also remove all downtime events that are related to nodes maintenance activities or node-specific artifacts like reboots and hardware failures.

Software running on the nodes also collects various metadata attributes (e.g. RAT, signal strength, etc) from the modems and periodically sends them to the NNE backend. This metadata is important for contextualizing the collected measurements. Our measurement period spans three years from the 1st of January 2014 to the 31st of December 2016. Overall, we collect measurements from 277 distinct measurement nodes and 545 MBB connections.

Our analysis, unless mentioned otherwise, is based on measurements from stationary nodes. Unlike when nodes are moving, the stationary scenario allows us to focus on characterizing MBB availability under relatively static conditions. Not only does this help

¹A realtime visualization of NNE measurements is available at <http://robustennet.net/map>.

²We only focus on measurements from the UMTS/LTE operators, since we do not have the necessary logs to measure the CDMA2000 network downtime.

³NSB is a publicly-owned railway company operating most passenger trains in Norway.

⁴Formerly known as Netcom.

⁵NNE connections do not generally experience extreme delays. Measured RTTs to the backend servers are below 100 msec in 99.5% of the cases.

Figure 2: The distribution of downtime (left), packet loss (middle), and overall availability (right) in 2016. A considerable fraction of connections exhibits three nines of availability. characterizing basic MBB availability, but also resembles many actual application scenarios that require high availability and are stationary like M2M communication.

4 MEASURING AVAILABILITY

The plots in Fig. 2 show the distribution of downtime, packet loss, and OA for Telenor, Telia, and Network Norway (NNorway) stationary connections in 2016⁶. The line in the OA distribution marks an OA of 99.9%, that is unavailability of 86.4 seconds per day on average. Note that by connections here we refer to the measured subscriptions.

The measured networks exhibit a high level of availability as measured by different metrics. However, there are clear differences between operators. More than 80% of Telenor connections achieve an OA of three nines or more, while only 60% and 70% of Telia and NNorway connections, respectively, achieve the same feat. Most of the unavailability in Telia can be attributed to packet loss since there are no major differences between downtime distributions for Telia and Telenor. Differences between NNorway and the other operators in OA can be traced back to differences in both downtime and packet loss. Note that stationary connections availability is an upper bound on MBB availability. Moving connections are subject to repeated handovers resulting in downtime and packet loss that are several times higher than the stationary case [6, 47].

We compare received signal quality across operators to verify whether the observed differences are a byproduct of NNE nodes placement. In particular, we measure the reference signals received power (RSRP) for each connection every minute in the first two weeks of June 2016. Then calculate the average RSRP. We focus on RSRP because it is a suitable measure of signal quality in LTE [13]. Recall that most of the measured connections are LTE. Figure 3 shows the average RSRP for all measured connections from the three operators above. Across networks, over 80% of connections receive a reasonably strong signal; over -100 dBm. Further, there are no clear difference between the three operators. This shows that the observed differences in downtime and packet loss can not be attributed to differences in coverage quality.

Figure 3: Mean RSRP. Most connections receive a reasonable signal quality.

Figure 4: The fraction of connections with 99.9% OA over time.

4.1 Availability over time

Having seen that today's commercial MBB networks are characterized by a high level of OA, we now turn to investigating how OA has changed over time and causes of differences between operators. To this end, we measure the fraction of connections that have a 99.9% OA in each year in the study period, which we present in Fig. 4. All operators, except Tele2, have exhibited a positive trend in OA, with Telenor outperforming the other operators. Telenor connections with 99.9% OA have increased from less than a third in 2014 to over 80% in 2016.

We believe that the main cause of this positive trend is LTE rollout. LTE is known to outperform UMTS on several performance metrics [7, 24, 25]. There is less packet loss in LTE thanks to the more robust air interface, the minimal state machine and flatter network architecture [13]. To check whether the differences between the operators can be attributed to differences in LTE deployment, we quantify LTE deployments as seen by the measurement nodes at

⁶Tele2 is not present on these graphs since it was decommissioned in Q2 2015.

Table 1: LTE deployment as seen by NNE nodes

	Telenor	Telia	NNorway
2014	28% (4.3%)	31% (13.4%)	25% (17.5%)
2015	80% (3.2%)	55% (28%)	72% (10.6%)
2016	93% (3.4%)	70% (17.8%)	89% (5.2%)

the end of each studied year. Using the metadata collected by NNE nodes, we breakdown the life time of each connection according to time spent on different RATs. To monitor the change in LTE deployment over time, we perform this breakdown for two weeks every year one in mid-June and the other in mid-December. We classify a connection as predominantly LTE, if it has spent 90% of the measured week on LTE. Further, we label a connection as mixed-LTE, if it reports less than 90% LTE coverage.

Table 1 shows, for each operator, the percentage of predominantly LTE connections as well as mixed-LTE connections in December of every year⁷. The latter is inside the parentheses. When summing these numbers, we find that each operator's LTE coverage has doubled over the measurement period. We, however, note that Telia, compared to the other operators, has less predominantly LTE connections and more mixed-LTE connections. This hints that differences in time spent on LTE can be a plausible explanation for the differences in packet loss between Telia and the other operators in Fig. 2. To check this hypothesis, we perform the following simple test. For all Telenor and Telia connections, we identify all time periods, in December 2016, where these connection were exclusively on LTE⁸. We then compare packet loss for these periods. If the difference between the two operators persists then it can not be explained by the time spent on LTE. Figure 5 shows the measured packet loss distribution. Unlike in Fig. 2, Telia exhibits a lower packet loss which supports our hypothesis.

A closer look at Telia connections' metadata shows that, unlike other operators, Telia does not implement a seamless inter-RAT handover from UMTS to LTE. Once in UMTS, a Telia connection must become idle or in a paging radio state before going back to LTE. NNE nodes send data every second which does not allow these conditions to be realized. We, hence, find that Telia connections return to LTE only when they detach and reattach to the network. Note that the lack of such implementation may not adversely affect smart phones, since they tend to become idle when users are not using them. It, however, affects users with active data sessions while on UMTS. This highlights the importance of properly implementing inter-RAT handover from UMTS to LTE to ensure that UEs are always attached to the best available RAT.

Takeaways. The overall availability of stationary MBB connections has significantly improved over time, but it is still far from the carrier-grade availability of five-nines or more. The improved availability can be largely attributed to the stability brought by LTE. The observed differences between operators highlights the importance of independent long-term performance benchmarking of mobile networks.

⁷June results are consistent with the trend in the table.

⁸Other months show similar results.

Figure 5: Comparing loss while on LTE. Telia loss is clearly lower.

5 CORRELATION OF OUTAGES

As Sec. 2.1 explains, there are numerous causes for unavailability (outages). Some of them are hard to pinpoint, especially those triggered by failures in shared infrastructure. In this section, we investigate whether end-to-end measurements can accurately discern infrastructure sharing. To this end, we employ a simple correlation-based approach that leverages the scale of NNE and the geographic spread of its nodes. Then gauge the suitability of end-to-end measurements by cross-checking the observed correlations with the known dependencies between operators (see Fig. 1). If end-to-end measurements prove successful, we can divide outages into correlated and uncorrelated and thus be able to assess the role of shared infrastructure. This will allow to quantify contributions of different classes of outages thus pointing to possible directions for increasing the OA.

5.1 Correlating downtime and loss

We divide each connection's lifetime into a time series of five-minute bins. We then build a binary downtime time series, where we set the time series value to one for bins where connections experience downtime and to zero otherwise. If a downtime spans multiple consecutive bins, we set all of them to one. We also construct packet loss time series by calculating the packet loss rate in each bin. We divide our time series into quarterly time series (i.e. 3 months long) giving us a total of up to 12 quarterly time series per connection. This division is important for observing changes in correlations over time. We then calculate the Pearson's correlation coefficient for every pair of downtime time series and for every pair of packet loss time series [9]. Pearson's correlation is a measure of linear dependence (i.e. whether the two time series exhibit similar and simultaneous changes). We include only pairs of time series that overlap for at least one day. On average, we have 52 days overlap between time series pairs.

To avoid spurious correlations due to short and excessively long downtimes, we only consider downtime outages that last for between 1 minute and 1 hour. We also only consider packet loss that is greater than 3%, i.e., losing 9 packets in a five-minute bin, to avoid spurious correlation due to random loss. If loss is lower than 3%, we set it to zero. Note that packet loss is measured only when the connection is up. Correlating downtime and packet loss may lead to different conclusions, as these metrics can give insights into different types of root causes. Downtime involves root causes that lie between the measurement node and the mobile core network.

Figure 6: The average correlation coefficient between connections from the same network (intra-operator), connections from different networks (inter-operator), and connections to different networks that terminate at the same measurement node for downtime (top) and packet loss (bottom).

Packet loss, however, may involve root causes that lie in the Internet between the mobile core and NNE servers.

We divide the connection pairs into three groups. 1) *Intra-operator*, for measuring correlations within a single operator. These correlations can be caused by shared transport as well as centralized components. 2) *Inter-operator*, for measuring correlations between different operators. These correlations are caused by shared infrastructures and thus are expected to be higher for pairs that are geographically close to each other. 3) *Diagonal pairs*, which are inter-operator pairs that are connected to the same measurement node, i.e., present at the same location. These pairs correspond to the highest geographic proximity. Hence, we expect diagonal pairs to exhibit much higher correlations compared to the average inter-operator pair. The six panels in Fig. 6 show the average correlation coefficient $\bar{\rho}$ over time for the different groups for both downtime and packet loss.

Intra-operator correlations are erratic, varying widely from a quarter to another. This indicates that the underlying causes do not follow a specific trend, e.g., correlation did not decrease as LTE deployment picked up.

Inter-operator correlation averages are about an order of magnitude lower than the *intra-operator* ones, but they are characterized by evident trends and changes. The difference in absolute values of $\bar{\rho}$ hints that inter-operator failures due to shared infrastructures are less frequent than their intra-operator counterparts. As expected, operators that use the same RAN (e.g. Telenor and NNorway; and Telia and Tele2 before Q1'2015) are more correlated. The observed

correlations *match well* the changes in the relationship between the measured networks which we described in Sec. 3. For example, the correlation between Telenor and NNorway has dropped since early 2015, while the correlation between Telia and NNorway has picked up since then. Operators that only share transport, i.e., Telenor and Telia exhibit lower correlation. This correlation, however, jumps in some quarters, e.g., downtime Q4'15 and packet loss in Q2'14 and Q4'15, to take values comparable to operators that share RAN. Furthermore, Operators without significant infrastructure sharing like Telenor and Tele2 exhibit extremely low correlation.

The correlation between operators that share infrastructure is very evident when looking at *diagonal pairs*; $\bar{\rho}$ is an order of magnitude higher compared to all pairs. This high difference suggests that access and regional transport networks, and perhaps co-located cell sites, are key contributors to correlated failures. Note that we observe no difference in $\bar{\rho}$, when looking at diagonal pairs, for operators without significant sharing, e.g., Telenor and Tele2. This confirms that the observed high diagonal correlations are not caused by artifacts at the measurement nodes.

5.2 Correlation and geography

In order to gain insights into causes of observed correlations, we next investigate whether the presence of correlation is related to the geographic distance between nodes. For each pair of connections with non-zero correlation, we compute the p-value for the measured correlation using a permutation test [22]. Our goal is to test the null hypothesis that the pair are not correlated. We fix

Figure 7: CDF of distances between node pairs for downtime (left) and packet loss (right). Solid lines (correlated) and dash lines (uncorrelated).

one of the time series and shuffle the other time series 1000 times, calculate the correlation coefficient for each of these 1000 pairs, and use these values to construct an empirical distribution of shuffled data correlation coefficient. Then, we compute the p-value for the measured correlation as its probability according to the shuffled data distribution. We reject the null hypothesis if the p-value < 0.05 . In other words, a pair is labeled as correlated, if the chance that the measured correlation is due to randomness is less than 0.05. Note that we do not use a fixed correlation threshold, e.g., greater than 0.5, to label a pair as strongly correlated. Instead, we determine whether an observed non-zero correlation is down to randomness; if not we label a pair as correlated.

Figure 7 shows the distribution of distances between correlated pairs (solid lines) and uncorrelated pairs (dashed lines) for both downtime and loss and a set of operator combinations. These distributions are based on correlation measurements from all quarters. Inter-operator pairs, that are correlated with respect to downtime, are generally separated by a shorter distance compared to uncorrelated pairs⁹. Half of correlated Telenor-Telia and Telia-NNorway pairs are separated by 100km or less compared to 300km for uncorrelated pairs. Further, apart from intra-Telenor, curves for correlated intra-operator downtime pairs collapse onto uncorrelated pairs. All packet loss curves, apart from a slight difference by Telenor-Telia pairs, collapse onto each other.

The expected distance between any pair of connections is 300km, which is a byproduct of the way NNE nodes are placed, i.e., mostly in the three major cities in Norway. Hence, for uncorrelated pairs, the mode at 300km in Fig. 7 just reflects nodes' placement. For correlated pairs, however, it means that these correlations are independent of geography and are thus likely caused by shared central components. Further, the prevalence of shorter distances between correlated inter-operator downtime pairs points to causes in nodes' geographic proximity. These include components in the access and

⁹This distinction is less evident for Telenor-NNorway since these two shared infrastructure for only the third of the study period.

regional networks as well as co-located sites. The absence of a strong correspondence between packet loss correlations and distance suggests that these correlations are dominated by packet loss that occurs in the core network and beyond. This indicates that whereas the RAN and transport are well provisioned, Internet paths from the mobile core are less so.

5.3 Correlation observability and sensitivity

This subsection assesses the robustness and sensitivity of the measured correlations. These correlations can be influenced by a number of factors including the thresholds we use for defining failures, the number and placement of measurement nodes, and the nature of the underlying failures.

Impact of thresholds. We start by investigating the thresholds we use in Sec 5.1 to construct the binary downtime and packet loss time series. Recall that in Sec 5.1, we only correlate connection downtime events that last between 1 min and 1 hour and packet loss that is at least 3%. We lower the minimum downtime duration to 30 seconds and recompute $\bar{\rho}$ for all quarters. We also measure $\bar{\rho}$ for packet loss for two new minimum packet loss thresholds which are 1% and 5%. We then compute the percentage change in $\bar{\rho}$. Packet loss correlations show a slight decrease when lowering the threshold and a slight increase when increasing it; both are less than 5%. In the worst case, downtime correlation decreases by 46% when lowering the minimum duration, which means that failures shorter than 1 minutes are mostly uncorrelated. There are also much fewer correlated downtime instances compared to packet loss which makes downtime correlation susceptible to perturbations brought by adding uncorrelated instances. These findings indicate that choosing conservative minimum thresholds results in avoiding spurious correlations and consequently observing otherwise difficult to spot correlations. This supports our decision to focus on non-trivial minimum downtime and packet loss values when constructing the respective time series.

Impact of measurements nodes placement. We now turn to analyzing the impact of measurements nodes placement on observed correlations. In other words, our goal is to check whether these correlations and corresponding trends can be attributed to a small subset of connection pairs, namely highly correlated pairs. We investigate the impact of nodes placement by removing all diagonal inter-operator pairs, i.e., nodes that have connection to both operators. Recall that the average correlation between diagonal pairs is an order of magnitude higher than the average inter-operator pair. We repeat this test for all quarters. Across quarters, removing diagonal pairs halves the downtime correlations but has no impact on packet loss. The impact of diagonal pairs is in agreement with the previous observations that inter-operator downtime correlations seem to stem from failures in access and regional transport infrastructure and that packet loss correlations seem to be related to failures in the mobile core and the Internet. However, the removal of these pairs does not produce a qualitative difference in observed correlation trends. This stability can be explained by the presence of many non-diagonal connections in the geographic proximity of each other. For instance, in Q4 2016 one fifth of Telenor-Telia pairs, excluding diagonal pairs, are within 100km from each other. As

Fig. 7 shows the presence of such nodes helps spotting correlations due to shared access and regional networks.

Takeaways. Correlating end-to-end measurements from geographically spread nodes proves powerful in capturing interdependencies between operators. We find that the causes of correlations in packet loss and intra-operator downtime seem to mostly lie in the mobile core and beyond. This involves the mobile core and the portion of the Internet between the mobile core and NNE servers for the former and the mobile core for the latter. Correlations in inter-operator downtime is caused by shared infrastructure, close to the edge. These findings suggest that operators need to follow different strategies to reduce correlations, 1) minimize infrastructure sharing for reducing inter-operator downtime correlations 2) diversify the mobile core for reducing intra-operator correlations 3) focus on the performance of paths from the mobile core to various destinations in the Internet for reducing packet loss correlations.

6 CHARACTERISTICS OF OUTAGES

Having established that end-to-end measurements capture interdependencies between operators reasonably well, we now take a closer look at the characteristics of correlated outages. We are interested in quantifying outages' magnitude and the extent and role of shared infrastructure.

6.1 Outages' magnitude

An outage size in terms of simultaneously impacted measurement nodes can give insights into root causes of outages. For instance, a large number of nodes at diverse geographic locations are likely to be impacted when a central network element fails. Outage size statistics help assessing operators' vulnerability to large outages, i.e., the extent of centralization within an operator's network. For characterizing outages, we first need to group individual connections' outages into correlated failure events.

Grouping outages into events. We assume that all connections which experience downtime (packet loss) in the same five-minute bin are impacted by a common underlying outage. While this approach is admittedly simplistic, we believe it is adequate for our analysis since we are interested in building a qualitative understanding of outages' magnitude. Next, we refine this grouping by verifying that the involved connections are significantly correlated. A pair is significantly correlated if the measured correlation is positive and is associated with a p-value ≤ 0.05 (see Sec. 5).

Constructing intra-operator events. After identifying all connections that experience an outage in the same time bin, we go through all these pairs and divide them to one or more subgroups of significantly correlated pairs. The rationale here is to only group connections that have proved significantly correlated over the entire quarter. These subgroups are constructed as follows. We start with an empty graph that includes all initially identified connections but no links. Then we add a link between all significantly correlated connection pairs. Next, we pick an arbitrary connection N and build a directed acyclic graph (DAG) [8], that is rooted at N , following a depth-first search approach. We essentially identify all connections that N is significantly correlated with and add them to the DAG, then identify their neighbors, i.e., all connections that they are significantly correlated with, and add them to the DAG

Figure 8: Intra-operator outages' size distribution for connection downtime (top) and packet loss (bottom) in Telenor.

and so on until there are no more connections to add. We remove the constructed DAG and repeat the same process until no connection is left. This process may result in multiple DAGs, which we consider as groups of nodes impacted by simultaneously occurring but separate events. Note that, by construction, any connection belongs to only one of the constructed DAGs. The outage size is the number of connections in each DAG divided by all connections that are available in the time bin of the outage.

Constructing inter-operator events. We identify all connections, from the two operators in question, that experience an outage in the same time bin and place them in two sets; one per operator. Then we start constructing a bipartite graph [8] by adding a link between pairs with significant correlation. Note that by a pair here, we mean two connections from two distinct operators. Finally, we traverse the bipartite graph and identify all connected bipartite sub-graphs which we consider as groups of nodes impacted by separate events. Note that, by construction, any connection belongs to only one of the constructed bipartite sub-graphs.

Intra-operator outage sizes. The panels in Fig. 8 show the CCDF of downtime and packet loss outage sizes for Telenor for different quarters. *Downtime* size distributions span a couple orders of magnitude and mostly collapse onto each other. These distributions appear to be consistent with a power-law distribution; a simple ordinary least square fitting yields a power law exponent $\alpha \approx 2 \pm 0.2$ indicating an undefined second moment, i.e., very high variability. The largest observed event impacted $\approx 40\%$ of connections. Other operators show qualitatively similar results. *Packet loss* outage size distributions span also two orders of magnitude, but large outages that impact more than 10% of the connections are more likely. The distribution for Q1'2014 differs markedly from the other

five quarters. In the first half of 2014, Telenor was suffering a mis-configuration that resulted in frequent packet loss events¹⁰. This misconfiguration resulted in a large extent of spurious correlation which increased the fraction of outages with large sizes. This high spurious correlation is due to the fact that the mis-configuration affected all connections. Hence, this spurious correlation is introduced by legitimate external factors and does not reflect a methodological weakness. To sum up, the long tail nature of outage sizes distributions confirms that the centralized MBB architecture is a major cause of correlated control and data plane outages.

Inter-operator outage sizes. The majority of correlated inter-operator downtime events have small impact. There is only four events that impact more than 5% of Telenor and Telia connections simultaneously. There are, however, more packet loss events that impact more than 5% of Telenor and Telia connections. Excluding outages that happened in the first half of 2014, there are 48 such events. These outages can be related to congestion and dynamics in the transport network and on shared Internet paths towards NNE servers. The former may not necessarily result in downtime if it does not last long enough for respective connection management timers to expire, which explains the difference between downtime and packet loss above. Looking at other inter-operator combinations, we note that for both metrics, the likelihood and extent of large simultaneous outages increase as the degree of infrastructure sharing increases. Operators that share the entire RAN are ≈ 8 to 10 times likelier to experience large simultaneous outages compared to operators that only share the core transport network.

Case studies. Next, we present and describe three outages demonstrating that the spread and scale of NNE yield a decent explanatory power.

1. Intra-Telenor downtime. At noon local time on 23/11/2015, 23 Telenor connections went down; this amounts to 27% of all active connections at the time. The first impacted and the last impacted connections experienced the outage within 43 seconds from each other. This is expected since EPS bearer tearing down is controlled by inactivity timers which are not synchronized across connections. The outage lasted for about 2 minutes. Except for three connections, all involved connections were hosted in and within 100 km from Oslo. Hence, the likely cause of this failure is either a failure of a network function in the CN that manages connections around Oslo, or the failure in the interconnect between Telenor's regional transport and its IP core. Note that it is unlikely that the base stations serving these connections share the same transport links, since they are located in different directions around Oslo¹¹.

2. Telenor-Telia downtime. At 02:19 AM on 26/11/2015, eight, eleven, and three Telenor, Telia, and NNorway connections respectively went down. Six affected nodes included both Telenor and Telia connections, which went down. The outage lasted about one minute. All the impacted connections are in Oslo or within 100 km from it. Hence, as in above the likely cause is the failure at the interconnect between Telenor's regional transport and its IP core. This outage highlights a possible weakness in the transport topology.

¹⁰3G connections were erroneously forced to an idle state even when they were actively exchanging data. Please refer to [5, 15] for more details.

¹¹Fiber in Norway is mostly laid next to railroad, highways, and power lines.

Figure 9: Correlated data plane failure in northern Norway. Geo-locations of impacted connections (left) and outage timeline (right).

3. Telenor-Telia packet loss. At 10:34 AM on 27/01/2016, eleven, ten, and two Telenor, Telia, and NNorway connections respectively experienced 100% packet loss for between 5 and 20 seconds depending on the operator. These connections are spread across the northern part of the country as illustrated on the map in Fig. 9. The same figure also shows how the number of available connections and those experiencing packet loss have changed in a 10-min period that includes the outage. The outage hit the two operators at about the same time. Soon, however, half of Telia connections disconnected, while Telenor recovered. This highlights that operators are using different inactivity timer values for managing the life time of data sessions. We also noticed that almost all connections suffered a sharp increase in RTT, up to seven times, in the two seconds preceding the outage. This hints that the underlying cause can be a reroute or a sudden congestion event at a central component in the transport network. Telenor runs a ring of two fiber backbones to cover the north of Norway. This ring starts in Tromsø and ends in Trondheim (See page 6 in [17]). Based on the location of affected connections, we expect them to be connected to different sides of the ring. Hence, the root cause most likely lies at one of the meeting points. Although this outage was short, it highlights a major weakness in the resilience of the MBB ecosystem in northern Norway.

Validating inferred outages. We validate all inferred intra-operator outages, that took place since Q3 2015, through direct communications with the affected operators. We have also leveraged publicly available fiber topology maps [17, 23, 36] together with the location of our nodes to verify whether a pair of nodes are likely to use the same fiber segments. This lends extra information to contextualize the identified outages (see case study 3 above). We plan to conduct a thorough validation in our future work.

6.2 Correlation graph

The measured correlations can be used for constructing a graph, i.e., a correlation network, by adding a link between correlated pairs. The primarily goal of constructing a correlation graph is to identify areas with high interdependence. For this to yield meaningful insights, observed correlations need to imply causation. To this end, we construct a simple hypothesis test.

The first idea behind our test is that correlation is a decreasing function of distance between measurement node pairs. The farther the two nodes are the less infrastructure they share. The second idea

is that a non spurious correlation between remote nodes has to be caused by outages closer to the mobile core network or systematic network-wide outages. These outages will also impact other remote nodes because of the hierarchical structure of the network.

We start by assuming that correlation between nodes that are spaced by less than X km implies causation. For other significantly correlated node pairs, which are spaced by more than X km, we check how many other nodes each pair is commonly correlated with. We then accept that the correlation between a node pair implies causation if they are commonly correlated with at least one node. This low threshold is reasonable given that we use a two-stage verification process, i.e., the permutation test in Sec. 5 and this test.

We use the constructed graph to assess how tightly coupled the investigated networks by computing the maximum correlation distance η , which is the largest great circle distance between causally correlated pairs. We experiment with a few values of X in the range from 25 to 100 km. The choice of X has a little impact on intra-operator correlation network. X 's value influences η for inter-operator downtime correlation network, because many inter-operator pairs do not have nodes in common. This sparseness is a consequence to the fact that NNE nodes have a mixed configuration when it comes to the attached networks. We plan to study the interplay between measurement infrastructure layout and the ability to discern causation in our future work.

The map in Fig. 10 gives an example of the inferred causally correlated networks for Telenor-Telia downtime pairs in 2015. It shows that nodes are grouped into distinct clusters that follow geography. This confirms that the observed correlation between Telenor and Telia is mainly caused by outages in access and regional networks. These dependencies can extend to tens of kilometers though, indicating the presence of a non-trivial level of centralization in the transport network. Similar maps for intra-operator downtime pairs and inter/intra-operator packet loss mostly consist of a single giant component denoting a country-wide correlation.

The other plots in Fig. 10 show η for different combinations of operators when setting X to 25 km. Intra-operator pairs demonstrate η values around 1000 Km. Inter-operator downtime pairs are characterized by a relatively shorter η . These, however, can reach up to 100 km, which correspond to the diameter of a large metropolitan. Inter-operator loss pairs exhibit higher η values matching our previous observations. We also note that η follows closely interdependencies between operators. η for Telenor-NNorway packet loss pairs dropped in Q2 2015, which corresponds to the time frame that NNorway stopped roaming on Telenor. We also observe an increase in η for Telia-NNorway packet loss pairs starting from Q1 2015, which corresponds to the time that NNorway started roaming on Telia.

Takeaways. We have shown that correlating outages from a geographically diverse set of vantage points can give insights into the extent of underlying events. Furthermore, combining the correlations with knowledge about placement of the measurement nodes can help identifying regions with shared dependencies. Our analysis also shows that shared infrastructure can result in outages with non-trivial magnitude in terms of geographical impact, e.g., case study 3.

Figure 10: The identified correlation network when correlating Telenor and Telia in 2015 (left), the correlation distance η for downtime (top-right), and for packet loss (bottom-right).

7 CORRELATED OUTAGES IMPACT AND OPPORTUNITIES

Having seen that downtime and packet loss outages can sometimes simultaneously impact a non-trivial fraction of connections, we now turn to assess their contribution to the overall downtime and packet loss that is experienced by individual connections. This is important, since it can point to concrete avenues for boosting availability.

7.1 How bad are large and correlated failures?

We start by identifying for each connection all downtime and loss outages that correspond to the following: 1) *Large failures*, which we define as outages that affect at least 5% of all connections simultaneously, and 2) *Correlated failures*, which are failures that are simultaneously experienced with at least one significantly correlated connection; this set includes also large failures. These two groups of outages correspond to large scale failures and limited failures, respectively. Then, we examine the distributions of downtime and packet loss when excluding outages related that belong to these two categories. Figure 11 shows these distributions for intra-Telenor and Telenor-Telia downtime in 2016¹². Downtime decreases when removing intra-operator large and correlated outages. Connections with low downtime benefit the most from removing correlated failures. For instance, across operators the fraction of connections with 5 nines availability, i.e., $downtime \leq 0.001\%$ at least doubles. Removing large and correlated *inter-operator outages* has a negligible effect. However, removing these events lead to a visible increase in uptime for operator-pairs that share the RAN. Packet loss exhibits qualitatively similar results. According to the aforementioned observations, improving the availability of the majority of connections requires reducing uncorrelated failures.

Pinpointing the exact causes of uncorrelated failures requires collecting additional logs from the UEs (e.g. physical and MAC layer messages) and the network-side. Instead we perform a high level investigation of differences between connections that achieve 4 nines uptime and those do not, using measurements from the 1st week of June and the 1st week of December of 2016 for both Telia and Telenor. We compare the number of handovers and the

¹²We omit plots for other combinations and years due to space limitation.

Figure 11: Breaking downtime into correlated and uncorrelated for intra-Telenor (left) and inter-Telenor-Telia (right)

Figure 12: Increasing availability through multi-connectivity.

average RSRP. These two metrics evaluate the radio stability and signal quality respectively. The median RSRP for connections with $4 \text{ nines} \leq \text{uptime}$ is higher by between 3 to 7 dBm depending on the operator. Also, connections with $4 \text{ nines} \leq \text{uptime}$ experience a median of 10 handovers per week, an order of magnitude lower than connection with $4 \text{ nines} \geq \text{uptime}$. These hint that better coverage as well as better radio planning and UE mobility management are key to increasing availability.

7.2 Leveraging multi-connectivity

Motivated by the observations above about the contribution of inter-operator correlated failures to overall downtime and packet loss, we next quantify the potential availability boost when combining connections from two different operators. We perform our evaluations for both stationary and moving connections.

7.3 Stationary connections

We consider all connection pairs that are present at each measurement node. Note that each measurement node includes up to four connections and at most one connection per operator. This gives a total of 101 pairs in 2016. We then perform a simple intersection operation between downtime time series of the two connections that constitute each pair to identify periods where both connections were experiencing an outage. To repeat the same for packet loss, we check whether packets that are sent in the same second are echoed back or not. Recall that the packet loss measurements are based on sending a small UDP packet every second over each connection.

Figure 12 shows the distribution of the OA when combining three operators; two at a time. As in Sec. 4, the overall uptime is calculated

Table 2: Multi-connectivity on the move

Node	Telenor	Telia	Combined
1	99.88%	99.95%	99.99%
2	97.94%	98.55%	99.47%
3	99.73%	99.65%	99.91%
4	99.76%	99.82%	99.94%

by multiplying the connection uptime by the fraction of successfully echoed packets. Multi-connectivity boosts the OA by an order of magnitude. Almost all connections are at least 4 nines available (i.e. a max of 8.64 seconds of unavailability per day) compared to 2 nines for individual connections (see Fig. 2). Furthermore, about half of Telenor-Telia and Telenor-NNorway combinations are at least 5 nines available. The benefit is less when combining operators that share the RAN, i.e., Telia and NNorway. Breaking the overall uptime to its constituents shows that it is mainly decided by the correlations in downtime since almost all packet loss pairs achieve five nines of availability or more. This essentially means that almost all packet loss less than 3% (i.e. less than the qualification threshold in Sec. 5) is uncorrelated. In other words, its root causes must lie on the air interface or individual operator's core network. Hence, to increase the fraction of 5-nines connections, we need to reduce correlations in downtime. In practical terms, this translates to enhancing the reliability of access and regional transport networks.

7.4 Moving connections

To quantify whether multi-connectivity benefits extend to the mobile case, we use measurements from four nodes aboard regional trains. We, however, only focus on measurements that are collected within Oslo metropolitan area because railroad stretches beyond this area suffer patchy coverage. Each node is connected to both Telenor and Telia. Our measurements span the period from the 1st of September to the 8th of December 2015, and comprise a total of 139 days worth of train driving time. We calculate the uptime experienced by each connection on these nodes and the overall uptime when combining them. Table 2 shows these values. Multi-connectivity has the potential to greatly improve MBB uptime on the move. Looking closer at the outages experienced by the moving connections, we find that they are mostly related to coverage holes and unsuccessful handovers. The huge benefit of connectivity is because operators coverage holes are not identical. Node 2 is placed on a train serving the southeastern part of Oslo and runs through a track with several tunnels and coverage holes.

Takeaways. Reducing correlated and large outages can boost the availability of already stable connections. Improving the availability of the majority of connections, however, requires addressing uncorrelated outages. Simultaneously using connections from two operators, that do not share the RAN, gives 5 nines availability in half of the measured cases. It also boosts moving connections availability to 3 nines.

8 RELATED WORK

Measurements of mobile networks. Recently there has been a growing interest in measuring and quantifying the performance and reliability of MBB networks. Several studies investigated the extent and causes of packet loss in MBB networks. Many of them

identified RRC state transitions as a major cause of packet loss, e.g., [5, 38, 39]. Chen et al. performed a RAN level analysis of delays and packet loss in UMTS networks and attributed most periods with degraded performance to congestion and overloaded base-stations [10]. A study by Huang et al. showed that physical layer retransmissions are behind most transport layer packet loss [25]. Other studies investigated packet loss under mobility and established that handovers are the main culprit [6, 31]. Shafiq et al. [41] characterized the operational MBB performance during two high profile crowded events in 2012 using network-side traces. To the best of our knowledge, this paper is the first to use end-to-end measurements to investigate how shared infrastructure impacts MBB networks availability.

Ultra reliable cellular communication. The growing interest in supporting machine type communication (MTC) with stringent reliability requirements have motivated many to investigate possible approaches for achieving this. Popovski [37] has identified five major impairments for wireless reliability that are centered around channel quality, availability of resources, and equipment failure.

Several proposals have focused on leveraging transmission diversity either by using network coding or through double association with a macro and micro cell, e.g., [29, 43]. Other efforts have focused on optimizing cellular scheduling for latency and loss sensitive applications, e.g., [49]. Our work compliments these efforts by providing an empirical evaluation of today's networks and showing that multi-homing can boost availability significantly.

Infrastructure sharing. Investigating the reliability of the Internet physical infrastructure has been the subject of several previous works. Durairajan et al. [14] investigated the impact of infrastructure sharing in the US long-haul fiber optic network. Other articles investigated the vulnerability of the Internet to physical failures as well as the resiliency of optical networks to localized failures [2, 50]. This paper takes a first step towards investigating the impact of physical infrastructure sharing on MBB networks.

9 DISCUSSION AND CONCLUSIONS

In this paper we use end-to-end measurements to study the availability of today's MBB networks. Building on longitudinal data set from a geographically spread measurements infrastructure, we investigate the impact of infrastructure sharing on MBB reliability. In the following we discuss our findings, their implications, and the limitations of this paper.

MBB availability. This study shows that MBB availability and reliability are improving over time. The fraction of connections with three nines of availability has almost tripled over the study period. MBB networks, however, remain far from providing predictable and consistent availability. Most of degraded periods are uncorrelated which reinforces the common wisdom about the role of the mobile air interface. Accordingly, today's networks are far from being able to support services that require ultra-reliable communication, key to many 5G-vision [3, 33]. Supporting such services will thus require enhancing the reliability of the air interface. Existing proposals include increasing the spatial diversity, downlink and uplink decoupling, and coordinated multipoint transmission [16, 27, 42, 44]. In addition, we believe that there is a need for radically rethinking connectivity by devising solutions to seamlessly and simultaneously combining multiple operators.

Interdependencies between operators. We have shown that end-to-end measurements can help illuminating interdependencies between operators. Outages due to infrastructure sharing contribute minimally to the overall downtime experienced by individual connections. We, however, find that failures of shared components can isolate large geographic areas. While these outages are generally rare, their specter is alarming given the rising worries about targeted cyber-attacks. We believe that the regulatory bodies need to draft policies that regulate infrastructure sharing. Governments have already started recognizing vulnerabilities due to infrastructure sharing as a major threat [32].

Deriving additional metrics for reliability. Correlating outages seen by a diverse set of vantage points can help devising a set of new metrics to assess MBB reliability. More specifically, we have shown that two relevant metrics can be extracted from the measured correlations. The extent of centralization, which is measured by estimating the distribution of intra-operator events sizes. Ideally, operators should aim for engineering their networks to produce narrower outage size distributions. We have also proposed a simple hypothesis test to check whether measured correlations imply causation. We then construct a correlation network that interconnects all pairs with causal relationships. This graph can be used to measure, among other things, the extent of centralization and impact radius of infrastructure-sharing.

Leveraging multihoming. We have shown that combining connections from multiple operators can enhance MBB availability to a carrier-grade level. Implementing this, however, remains challenging. Existing business models and technology do not allow simultaneously using multiple operators via a single SIM card, which necessitates using two radios. Further, multi-path transport protocols like CMT-SCTP and MPTCP are by definition closed-loop protocols meaning that re-transmitting a lost packet may take one round trip time or more. This can potentially be solved by open-loop schedulers that leverage cross-layer communication, e.g., using MAC messages to detect link-layer retransmissions. Also, efficient network coding schemes can help reconstructing lost data segments.

Limitations. This study has two main limitations. First, packet loss is measured against a single server, which may lead to spurious correlations. We have applied a set of filters to remove loss because of failures in the NNE backend and within NNE upstream provider, Uninett. Outages that are caused by failures in the interconnect between individual operators and Uninett will translate into network-wide events. Going forward, measuring packet loss to other destinations can reduce correlations due to interconnect failures. Augmenting packet loss measurements with traceroute tests may help locating failures. Second, NNE nodes deployment is relatively sparse which may lead to overestimating spurious correlations. Having a denser deployment would further help identifying causal links.

ACKNOWLEDGMENTS

This work was supported by the Research Council of Norway grants 240850 and 209954. We thank our shepherd, Bozidar Radunovic, and the anonymous reviewers for their thoughtful feedback; and Amund Kvalbein and Olav Lysne for their feedback on earlier drafts.

REFERENCES

- [1] Action700. 700 MHz Broadband for Mission Critical Public Safety Data. <http://www.citig.ca/action700.aspx>, 2015.
- [2] P. K. Agarwal, A. Efrat, S. K. Ganjugunte, D. Hay, S. Sankararaman, and G. Zussman. The resilience of wdm networks to probabilistic geographical failures. *IEEE/ACM Transactions on Networking (TON)*, 21(5):1525–1538, 2013.
- [3] N. Alliance. NGMN 5G white paper. *White paper*, Feb, 2015.
- [4] D. Baltrunas, A. Elmokashfi, and A. Kvalbein. Measuring the reliability of mobile broadband networks. In *Proceedings of the 2014 Conference on Internet Measurement Conference, IMC '14*, pages 45–58, New York, NY, USA, 2014. ACM. ISBN 978-1-4503-3213-2. doi: 10.1145/2663716.2663725. URL <http://doi.acm.org/10.1145/2663716.2663725>.
- [5] D. Baltrunas, A. Elmokashfi, and A. Kvalbein. Dissecting Packet Loss in Mobile Broadband Networks from the Edge. In *Proc. of INFOCOM*, 2015.
- [6] D. Baltrunas, A. Elmokashfi, A. Kvalbein, and Ö. Alay. Investigating Packet Loss in Mobile Broadband Networks under Mobility. In *Proceedings of IFIP Networking conference*, May 2016.
- [7] N. Becker, A. Rizk, and M. Fidler. A measurement study on the application-level performance of LTE. In *Networking Conference, 2014 IFIP*, pages 1–9. IEEE, 2014.
- [8] G. Chartrand and P. Zhang. *A first course in graph theory*. Courier Corporation, 2012.
- [9] C. Chatfield. *The Analysis of Time Series: An Introduction, Sixth Edition*. Chapman & Hall/CRC Texts in Statistical Science. CRC Press, 2016. ISBN 9780203491683. URL <https://books.google.no/books?id=qKzyAbdaDFAC>.
- [10] Y. Chen, N. Duffield, P. Haffner, W. ling Hsu, G. Jacobson, Y. Jin, S. Sen, S. Venkataraman, and Z. li Zhang. Understanding the Complexity of 3G UMTS Network Performance. In *IFIP Networking Conference*, 2013.
- [11] Cisco. *visual networking index Mobile Forecast, 2015 - 2020*. Cisco Systems, Inc., February 2016.
- [12] CNN. *Why a massive cell phone outage hit the Southeast*, 2015. <https://goo.gl/lp4Cr1>.
- [13] C. Cox. *An Introduction to LTE*. John Wiley & Sons Ltd., 1st edition, 2012. ISBN 9781119970385.
- [14] R. Durairajan, P. Barford, J. Sommers, and W. Willinger. InterTubes: A study of the US long-haul fiber-optic infrastructure. In *ACM SIGCOMM Computer Communication Review*, volume 45, pages 565–578. ACM, 2015.
- [15] A. Elmokashfi, A. Kvalbein, D. Baltrunas, J. Werme, and E. Arge. Robusthet i norske mobilnett tilstandsrapport 2014. Technical report, Oslo, 04/2015 2015.
- [16] H. Elshaer, F. Boccardi, M. Dohler, and R. Irmer. Downlink and uplink decoupling: A disruptive architectural design for 5g networks. In *Global Communications Conference (GLOBECOM), 2014 IEEE*, pages 1798–1803. IEEE, 2014.
- [17] T. Faremo. Telenor's Landsdekkende nett. In *Fiberoptikk 2015*, Trondheim, Norway, November 2015.
- [18] G. P. Fettweis. The Tactile Internet: Applications and challenges. *IEEE Vehicular Technology Magazine*, 9(1):64–70, mar 2014. ISSN 1556-6072.
- [19] FirstNet. <http://www.firstnet.gov/>.
- [20] S. Forge, R. Horvitz, and C. Blackman. Is Commercial Cellular Suitable for Mission Critical Broadband? Technical report, European Commission, 2014.
- [21] M. Garwood. EE wins lucrative emergency services contract. <http://www.mobilenewscwp.co.uk/2015/12/10/ee-wins-7-billion-emergency-services-contract/>, December 2015.
- [22] P. I. Good. *Permutation, Parametric, and Bootstrap Tests of Hypotheses (Springer Series in Statistics)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2004. ISBN 038720279X.
- [23] G. A. Hansen. Telenett i nordområdene - spesielle utfordringer. In *Telecom World 2013*, Longyearbyen, Norway, February 2013.
- [24] J. Huang, F. Qian, A. Gerber, Z. M. Mao, S. Sen, and O. Spatscheck. A close examination of performance and power characteristics of 4G LTE networks. In *Proceedings of the 10th international conference on Mobile systems, applications, and services*, pages 225–238. ACM, 2012.
- [25] J. Huang, F. Qian, Y. Guo, Y. Zhou, Q. Xu, Z. M. Mao, S. Sen, and O. Spatscheck. An In-depth Study of LTE: Effect of Network Protocol and Application Behavior on Performance. In *Proc. of SIGCOMM*, 2013.
- [26] ICT-317669 METIS project. Deliverable D1.1 Scenarios, requirements and KPIs for 5G mobile and wireless system? <https://www.metis2020.com/documents/deliverables/>, 2013.
- [27] R. Irmer, H. Droste, P. Marsch, M. Grieger, G. Fettweis, S. Brueck, H.-P. Mayer, L. Thiele, and V. Jungnickel. Coordinated multipoint: Concepts, performance, and field trial results. *IEEE Communications Magazine*, 49(2):102–111, 2011.
- [28] P. Kelly. Mobile Broadband & the Future of Public Safety Networks from a Tetra Operator Perspective, April 2013. URL <http://m.eu2013.ie/media/eupresidency/content/documents/Presentation-8-15.50-Mobile-Broadband1.pptx>.
- [29] D. M. Kim and P. Popovski. Reliable uplink communication through double association in wireless heterogeneous networks. *IEEE Wireless Communications Letters*, 5(3):312–315, 2016.
- [30] A. Kvalbein, D. Baltrunas, J. Xiang, K. R. Evensen, A. Elmokashfi, and S. Ferlin-Oliveira. The Nornet Edge platform for mobile broadband measurements. *Elsevier Computer Networks special issue on Future Internet Testbeds*, 2014.
- [31] L. Li, K. Xu, D. Wang, C. Peng, Q. Xiao, and R. Mijumbi. A Measurement Study on TCP Behaviors in HSPA+ Networks on High-speed Rails. In *Proc. of INFOCOM*, 2015.
- [32] Lysneutvalget. Digital sårbarhet-sikkert samfunn, 2015.
- [33] Nokia. 5G use cases and requirements, 2014.
- [34] NRK. *Halve Norge var uten mobiltelefon*, 2011. <https://goo.gl/E0WfKp>.
- [35] OpenSignal. The state of LTE. <http://opensignal.com/reports/2015/09/state-of-lte-q3-2015/>, September 2015.
- [36] Osloecconomics. Konsekvensutredning - Alternativer for styrket robusthet i landsdekkende kjernenett. <https://www.regjeringen.no/contentassets/fe88e9ea8a354bd1b63bc0022469f644/no/sved/3.pdf>, August 2015.
- [37] P. Popovski. Ultra-Reliable Communication in 5G Wireless Systems. In M. Latva-aho, R. Tafazolli, N. Rajatheva, L. M. Correia, and T. Rasheed, editors, *5GU*, pages 146–151. ICST, 2014. ISBN 978-1-63190-055-6. URL <http://dblp.uni-trier.de/db/conf/5gu/5gu2014.html>.
- [38] F. Qian, Z. Wang, A. Gerber, Z. M. Mao, S. Sen, and O. Spatscheck. Characterizing Radio Resource Allocation for 3G Networks. In *Proc. of IMC*, 2010.
- [39] S. Rosen, H. Luo, Q. A. Chen, Z. M. Mao, J. Hui, A. Drake, and K. Lau. Discovering Fine-grained RRC State Dynamics and Performance Impacts in Cellular Networks. In *Proc. of Mobicom*, 2014.
- [40] S. Schütz, L. Eggert, S. Schmid, and M. Brunner. Protocol enhancements for intermittently connected hosts. *ACM SIGCOMM Computer Communication Review*, 35(3):5–18, 2005.
- [41] M. Z. Shafiq, L. Ji, A. X. Liu, J. Pang, S. Venkataraman, and J. Wang. A first look at cellular network performance during crowded events. In *Proc. of SIGMETRICS*, 2013.
- [42] B. Singh, Z. Li, O. Tirkkonen, M. A. Uusitalo, and P. Mogensen. Ultra-reliable communication in a factory environment for 5G wireless networks: link level and deployment study. In *Personal, Indoor, and Mobile Radio Communications (PIMRC), 2016 IEEE 27th Annual International Symposium on*, pages 1–5. IEEE, 2016.
- [43] V. N. Swamy, P. Rigge, G. Ranade, A. Sahai, and B. Nikolić. Network coding for high-reliability low-latency wireless control. In *Wireless Communications and Networking Conference (WCNC), 2016 IEEE*, pages 1–7. IEEE, 2016.
- [44] W. Tärneberg, M. Karaca, A. Robertsson, F. Tufvesson, and M. Kihl. Utilizing Massive MIMO for the Tactile Internet: Advantages and Trade-offs. In *IEEE SECON Workshops-Robotic Wireless Networks*. IEEE-Institute of Electrical and Electronics Engineers Inc., 2017.
- [45] TETRA. <https://tandcca.com/>.
- [46] TETRAPOL. <http://www.tetrapol.com/>.
- [47] F. P. Tso, J. Teng, W. Jia, and D. Xuan. Mobility: A Double-Edged Sword for HSPA Networks. In *Proc. of MobiHoc*, 2010.
- [48] Wireless2e. *Another Outage / More Lessons: Geo-redundancy Couldn't Prevent Verizon LTE Service Disruption*, 2011. <https://goo.gl/Yv0ljH>.
- [49] Y. Xu and C. Fischione. Real-time scheduling in LTE for smart grids. In *Communications Control and Signal Processing (ISCCSP), 2012 5th International Symposium on*, pages 1–6. IEEE, 2012.
- [50] L. Zhou. Vulnerability analysis of the physical part of the internet. *International journal of critical infrastructures*, 6(4):402–420, 2010.