Profiling Mobile Broadband Coverage

Andra Lutu, Yuba Raj Siwakoti, Özgü Alay, Džiugas Baltrūnas and Ahmed Elmokashfi
Simula Research Laboratory
Email: {andra, yubars, ozgu, dziugas, ahmed}@simula.no

***Abstract*—Pervasive coverage and continuous connectivity of Mobile Broadband (MBB) networks are common goals for regulators and operators. Given the increasing heterogeneity of technologies in the last mile of MBB networks, further support for seamless connectivity across multiple network types relies on understanding the prevalent network coverage profiles that capture different available technologies in an area. Correlating these coverage profiles with network performance metrics is of great importance in order to forestall disturbances for applications running on top of MBB networks. In this paper, we aim to profile MBB coverage and its performance implications from the end-user's perspective along critical transport infrastructure (i.e., railways in Norway). For this, we deploy custom measurement nodes on-board five Norwegian inter-city trains and we collect a unique geo-tagged dataset along the train routes. We then build a coverage mosaic, where we divide the routes into segments and analyze the coverage of individual operators in each segment. We propose and evaluate the use of hierarchical clustering to describe prevalent coverage profiles of MBB networks along the train routes and classify each segment accordingly. We further analyze the areas we classify with each profile and assess the packet-loss performance of the networks in those areas.**

I. Introduction

Mobile Broadband (MBB) access to the Internet enables operators to join mobility and communications towards the common goal of offering subscribers performance and efficiency in highly dynamic mobile scenarios. However, Internet access under mobility brings a number of challenges, including high probability of service interruptions. A popular example of such scenarios is the case of travelers regularly commuting on public transport infrastructures, such as inter-city trains. In this context, tens or hundreds of passengers try to access the Internet simultaneously for entertainment, communication and work-related tasks, all while moving at high speeds. During the last years, railway operators throughout the world have been testing and providing commercial Internet connectivity solutions aimed at enabling on-board Internet services to train passengers. Various types of communication solutions have been advanced [1], [2], including cellular solutions, WLAN-based solutions or hybrid terrestrial/satellite solutions.

The performance of cellular-based solutions for on-board connectivity highly depends on the MBB coverage around the railway lines. MBB operators are the main providers of coverage maps for other stakeholders, including regulators, subscribers or businesses such as public transport operators. These coverage maps usually define the best radio access technology (RAT) in a region for a MBB operator. However, they do not offer information on how the distribution of different RATs in the same geographical region translates to the end-users experience. Given the increasing heterogeneity of technologies in the last mile of MBB networks, user experience highly depends on support for seamless handovers across multiple network types. Therefore, identifying the *network coverage profiles* that capture the distribution of all available technologies in the same area from the end-user experience is of great importance.

In this paper, we focus on profiling the MBB coverage along the critical railway infrastructure in Norway. Then, our goal is to build a **coverage mosaic**, where we classify and characterize railway route segments based on the distribution of RATs in that segment. For this, we use a vast dataset that we collect through periodic measurements from custom devices that we strategically place on-board several passenger trains. The dataset is pestered by numerous challenges, including high volume, the mixture of spatio-temporal coordinates and the presence of qualitative variables (i.e., the RAT value). Furthermore, depending on the deployment of base stations along the railway routes, the distribution of different RATs highly varies from one segment to another. Some operators rely on thresholds they impose of statistical descriptors to characterize coverage and classify different regions in coverage categories. Even when envisioning a simple intuitive classification such as "good" coverage (where we have dominant 3G and 4G) or "bad" coverage (where we have dominant 2G or no service), there is no consensus on what should be the threshold in terms of percentage of 3G/4G presence in a certain area in order to label that area with good coverage. Thus, it is challenging and cumbersome to use statistical descriptors to define good and bad coverage. Therefore, in this paper, we leverage ideas from machine learning to help us overcome limitations of using classification rules based on statistical descriptors to characterize coverage. More specifically, we evaluate the use of *hierarchical clustering* to characterize the distribution of different RATs from individual MBB providers along the train routes. Clustering enables us to maneuver easier this dataset and determine the salient coverage profiles, characterize them and then assign segments to the proper profile.

The coverage mosaic we produce successfully captures the mixture of available RATs as experienced by the end-user inside the train. Two main coverage profiles emerge from our analysis, one where 3G is dominant (which we further title *"profile A"*) and another where no service is dominant (which we further title *"profile B"*). This validates the intuition within the community regarding the *"good"* and *"bad"* coverage.

Moreover, through the stability analysis of these coverage profiles, we demonstrate the need for repetitive measurements (at least 5-10 measurement runs) in order to profile the coverage of a certain area. We further analyze the network performance within each coverage profile and pinpoint the areas with bad coverage profile as troubles zones with poor performance. Though we expect this for bad coverage, even in the areas with good coverage we find high variations in the network performance, suggesting that frequent handovers in these areas are a challenge for the operators and impact the end-user experience.

II. Measurement Setup and Dataset

In this section, we present the measurement infrastructure that we use in this paper, the measurements we deploy and the dataset we collect.

A. Measurement Infrastructure

We use the NorNet Edge [3] (NNE) dedicated mobile broadband measurement platform. We expand the NNE testbed to include 6 custom NNE measurement devices (i.e., NNE nodes) active on the NSB[1] regional trains in Norway. NNE nodes are single board computers that run a standard Linux distribution and connect to multiple MBB operators. The node connects via Huawei E392-u12 modems supporting 4G/LTE connectivity.

To measure network performance and capture basic quality of service (QoS) metrics (e.g., packet loss, latency), we send a 20-byte UDP packet every second over each connection to an echo server that is part of NNE backend and then record a reply packet from the server. A packet is considered lost if we do not receive a reply from the server within one minute. We transfer all the data collected on the node to a server we host in the back-end and then import it into a database. Along with the measurement results, each node also provides context information (i.e. metadata) that is very valuable during the analysis. Furthermore, we access the train GPS location from the NSB system. For our coverage analysis, we use the combination of metadata and GPS data results, which we explain in detail in the following section.

B. Geo-tagged Dataset

The regional trains that host our measurement nodes run periodically on 4 different national *routes*[2], covering over 2,500 km. We define each one-way train trip on a certain route as a *run*. We collect data from these nodes for the two largest MBB operators in Norway, Telenor and Netcom for 5 months (from November 2014 to March 2015). The number of runs we have for each operator on each route ranges from 60 to 150 runs and depends on the public schedule of the trains hosting the nodes.

[1]NSB is a government-owned railway company operating most passenger trains in Norway.

[2]The train routes are: Oslo-Voss, Oslo-Stavanger, Oslo-Trondheim and Trondheim-Bodø.

For our analysis, we specifically require geo-localization of the coverage information from the modems. The trains update their GPS locations every 10 to 15 seconds in the NSB fleet management system. For each GPS point, we find the corresponding RAT value (e.g. *No Service*, 2G, 3G or 4G) from the metadata readings. We define a *geo-tagged data point* as the GPS point where there is a corresponding metadata reading[3]. These geo-tagged data points forms our *geo-tagged dataset*.

III. Identifying Coverage Profiles

Our goal is to build a *coverage mosaic*, where we segment the routes and classify the coverage of each segment into coverage profiles that capture the distribution of RATs as the end-user would experience it. The purpose of building such a coverage mosaic is to enable further analysis in terms of network performance characterization and reliability. In this section, we propose and evaluate the use of hierarchical clustering to characterize coverage patterns in space-time. However, this comes with a number of challenges, which we formulate below.

A. Motivation and Problem Formulation

Investigating coverage patterns in terms of distribution of different technologies in the same area over time is challenging for three reasons. First, the RAT distribution varies greatly from one segment to another based on the deployment of base stations in an area. This information is usually not available from an objective source. Additionally, connectivity upgrades are common and operators do not make their strategies public. Second, the geo-tagged dataset is difficult to work with because of its large dimensionality and spatio-temporal inconsistency (i.e., data point's location and time of reading differ from run to run over the same route). Third, the measurement data is noisy because of a number of factors, including specific geography of the area, variable train speed, number of passengers in the train or congestion in the network. For example, all the end-users active in an area at a moment in time might not simultaneously use the fastest available RAT. All these reasons make efficient characterization of the network coverage using statistical descriptors challenging and cumbersome.

To tackle these challenges, we design, implement and evaluate a machine learning methodology that can help us characterize coverage patterns in the areas of interest. Our methodology contains two separate parts: *data morphing* and *clustering*. First, in the data morphing, we propose the segmentation of the region of interest in smaller areas and spatially group the geo-tagged data points in these segments. Second, we focus on capturing the prevalent coverage profiles and classify each route segment to a profile accordingly.

Although good or bad coverage may seam like a straightforward classification, due to the challenges we present above,

[3]Note that there are cases when we do not have metadata information for the GPS reading (e.g., when the modem is down, or the IP address is lost). We discard from our dataset the GPS points with missing metadata information.

it is surprisingly hard to give predetermined quantitative definitions of what is good or bad coverage when focusing on the combination of different RATs in the same area (e.g., is continuous 2G bad coverage? is it worse having intermittent 3G coverage? how can one set the thresholds on the RAT distributions to classify good or bad coverage?). We expand on this issue in the following Section III-C. We propose the use of unsupervised clustering to help us with the classification. The spatio-temporal heterogeneity of the geo-tagged dataset we collect makes the clustering algorithm a good fit for identifying patterns in the coverage offered by MBB providers. In particular, we choose the hierarchical clustering algorithm [4], [5], which clusters together data instances based on their similarity [6], thus highlighting the prevalent coverage profiles in the region.

B. Data Morphing

Our geo-tagged dataset from repetitive runs consists of numerous time-stamped instances of network-specific variables at different geographic coordinates along the railway routes. We identify the objects in our dataset as categorical variables (i.e., the RAT value) with dynamic location (i.e., the results from the measurement device does not always come in the same point in space). The interaction between the spatio-temporal dimensions of the dataset dictates the complexity and challenges in moving from acquiring the data to drawing knowledge through data analytics. In order to address these challenges, we begin by organizing the dataset into instances that we can easily compare.

Spatial binning. We first divide the railway routes into smaller segments using a fix grid of 2km $\times$ 2km that we superimpose on Norway's map. Each square grid block that overlaps on the train routes contains a segment of the route. The resulting *segments* are disjoint and uniquely identified by the fix spatial coordinates of the square *grid blocks* that contain them. We then partition the geo-tagged dataset by grouping the data points that fall along the same route segment.

In order to make an informed decision on the size of the grid we use for spatial binning, we investigate the trade-off between the speed distribution of trains, the granularity of the geo-tagged data points and the amount of performance measurements we need to assess the MBB performance within a single route segment. The granularity of the GPS points (which further dictates the granularity of the geo-tagged data points) is 10-15 seconds. In order to ensure that we collect significant network performance measurements (i.e., 100 different UDP pings) to characterize a geo-unit, we need at least 4 GPS readings within the same bin. Also, we observe that 98% of the speeds we register from the fleet management system are below 120 kph, with majority falling between 50-100 kph. Considering all the above observations, we decide to use a grid block of size *2km $\times$ 2km*. As expected, due to variation in the speed, the number of GPS points varies. Withal, we observe that the majority of the route segments have multiple GPS readings, and approximately 75% of the grid blocks per route have more than 4 GPS readings. We leave for future work a detailed analysis of the impact that the size of the grid we use for spatial binning has on the coverage patterns we observe.

Coverage Chart Time Series. After the spatial binning, the route segment with fix spatial coordinates becomes the object that we further characterize in terms of mobile coverage. A route segment is characterized by a variable number of RAT readings, corresponding to the geo-tagged data points from every run that the 2km $\times$ 2km area encloses. In this second step, we transform the categorical variable representing the RAT at each geo-tagged data point into a set of continuous variables that show the distribution of each of the RAT values (i.e., 4G, 3G, 2G or noS) over the set of data points along a segment of the route. We define the distribution of RAT over one run as the segment's **coverage chart**. For example, if a segment contains 5 different geo-tagged data points [noS, 2G, 3G, 3G, 4G], then we can derive the coverage chart of the segment for the measurement run: 2G: 20%, 3G: 40%, 4G: 20%, noS: 20%.

We merge the runs independently of the train trip direction along a route to generate a *coverage chart time series*. To further ensure that we compare route segments for which we have similar coverage chart time series and to correct any artifacts in the GPS readings, we further analyze the route segments for which we collect coverage charts from a minimum of 75% of the total number of runs.

C. The Clustering Approach

After morphing the dataset, we reduce the problem to the matter of quantifying the similarity between the segments' coverage charts. We show in Figure 1 the spatial variation of the distributions of different RATs for Telenor on the Oslo-Stavanger route. The x-axis represents the route segment ID and the y-axis represents the percentage of each RAT for that particular segment using median, first and third quartiles. The points on the figure are the median/quartile values while the lines shows the fitted curves to these points. As illustrated in the figure, the distribution of RATs greatly varies in the spatial domain, supporting our claim that defining thresholds on some statistical descriptors of the RATs distributions to profile coverage is difficult and nonadaptive.

In this section, we present the similarity metric we choose, the clustering method we follow and the approach we use to determine the optimal number of coverage clusters. We perform the clustering of segments on a per-route basis, thus applying the same methodology on datasets we collect along the 4 different routes .

Similarity metric. In order to calculate the similarity between two segments, we organize the coverage time series into vectors of coverage charts we measure at each run over a route. The length of the vector is fix and equal to the number of measurement runs we register on a route. In the case where the coverage chart for a segment is missing (either due to hardware issues during the run or due to sifting the data based on the minimum required number of data points), we artificially populate the coverage chart with null values

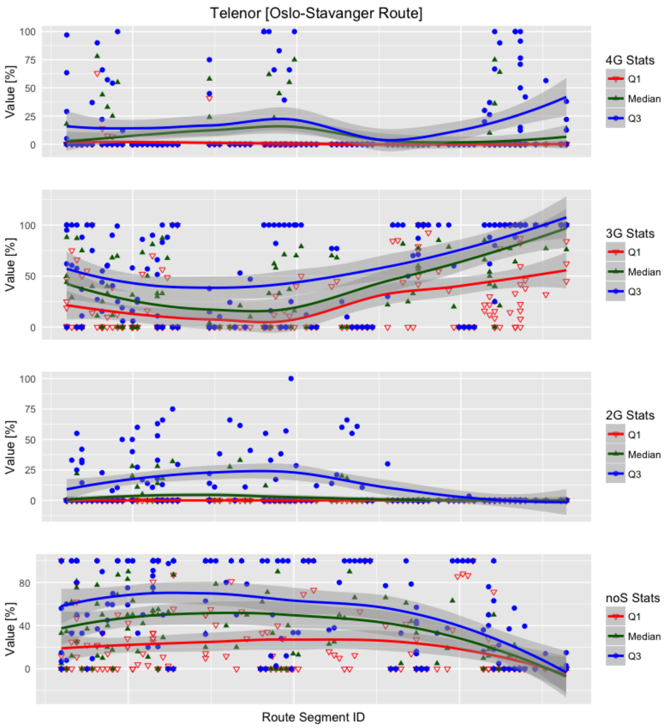


Fig. 1: *The distribution of RAT per route segment for Telenor along Oslo-Stavanger train route.*

for the corresponding coverage mode variables. We then use the extended Jaccard measure [7] to evaluate the similarity between two objects.

Clustering method. The clustering method we select is an average-link hierarchical clustering method, which organizes the data objects into a multi-level structure, based on the similarity between objects. Such methods consider the distance between two clusters to be equal to the average distance from any member of one cluster to any member of another cluster. More specifically, we employ here Ward's minimum variance method for clustering [8], which aims at finding compact, spherical clusters.

Optimal Number of Clusters. Hierarchical clustering leaves to the user the task of detecting the optimal number of clusters. To this end, we evaluate a *validity index* for different number of clusters [9]. The number of clusters that generates the best value for the index is then chosen as optimal. There is no general consensus on which validity index should be used. In this paper, we consider the Silhouette index [10] that represents an average, over all the clusters, of how similar the data in each cluster are.

IV. Clustering Results

In this section, we run the proposed methodology to characterize coverage along the four Norwegian train routes and then we analyze the results.

A. Coverage Clusters Analysis

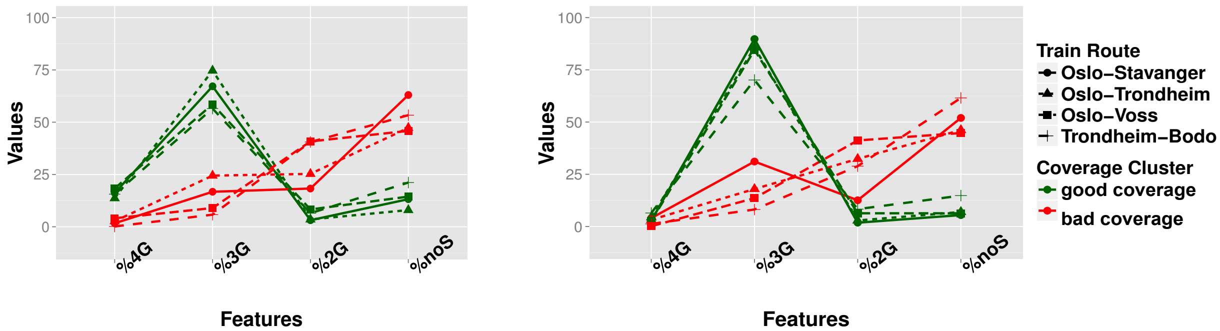
We apply the hierarchical clustering method to the coverage chart time-series of the segments covering the train routes. In other words, we calculate the similarity between segments' coverage chart time series and we group together the segments with similar coverage patterns. We then determine the optimal number of coverage clusters using the Silhouette index. For all the four routes and both operators, the Silhouette index gives 2 clusters, which we detail next.

Prevalent coverage profiles. The optimal clustering of route segments defines the dominant *coverage profiles* for the two MBB operators. We observe that the clustering algorithm identifies two main coverage profiles, which we generically label as *good* and *bad* coverage profiles. In Figure 2, we show the characteristics of these two coverage profiles, which consist of the average RAT distributions [2G, 3G, 4G, noS] over all runs for all the segments in the same cluster.

When analyzing the coverage profiles, we note that, in the areas with *Profile A*, Telenor has around 70% 3G accompanied with 15% 4G, while Netcom compensates with higher 3G availability (85% 3G) for its slower 4G deployment (5% 4G). This clearly shows the different deployment strategies of the operators. Furthermore, we observe slight differences in the profiles of different routes. For example, the Trondheim-Bod route in the *Profile A* areas of Netcom clearly stands out because it has a lower 3G and higher noS percentage compared to the other routes. For the *Profile B* coverage areas, we observe a high degree of *No Service* for both operators, which combines with a considerable amount of 3G and 2G. The distribution of 2G and 3G varies considerably among different routes. For both operators, the segments with *Profile B* coverage correspond to a large part of the critical transport infrastructure and concentrate in the rural inter-city area.

Coverage profiles on route. We next exemplify the coverage profiling and show the results we obtain for the Oslo-Stavanger route. In Figure 3 we depict the results for the hierarchical clustering we obtain for Oslo-Stavanger, both for Telenor and Netcom. Each subplot contains the dendrogram of the hierarchical clustering grouping according to the similarity measures we choose. We group the segments into two main coverage clusters, namely the *Profile A* and *Profile B* coverage clusters. We note that for both operators, the two clusters are well distanced. Additionally, the distances between elements within the same cluster are relatively small. Thus, the dendrogram allow us to clearly observe the separation between the clusters according to the similarity measure, validating the result of the validity index.

Though, using the Silhouette index, we systematically discover two prevalent coverage clusters that dictate the coverage profiles, we can further divide these clusters into smaller sub-clusters with more homogeneous profiles. In some cases, other indexes such as the Dunn or DB indexes indicate a larger number of clusters, because of the heterogeneity between the coverage time series of each of two major clusters. We see in Figure 3 that the coverage clusters contain several coverage sub-profiles that highlight the predominance of one RAT or the mixture of several RATs. For example, in the case of Telenor, we identify 4 different sub-clusters in the *Profile A* "good" coverage cluster, underlining the increasing heterogeneity of technologies in the MBB networks. For future work, we plan to collect MBB metadata information with a 1-sec granularity



(a) Telenor (b) Netcom

Fig. 2: *Coverage profiles along the four measured routes in Norway, for (a) Telenor and (b) Netcom. The coverage profile consists of four features that show the distribution of the four RATs over the segments in the same cluster.*

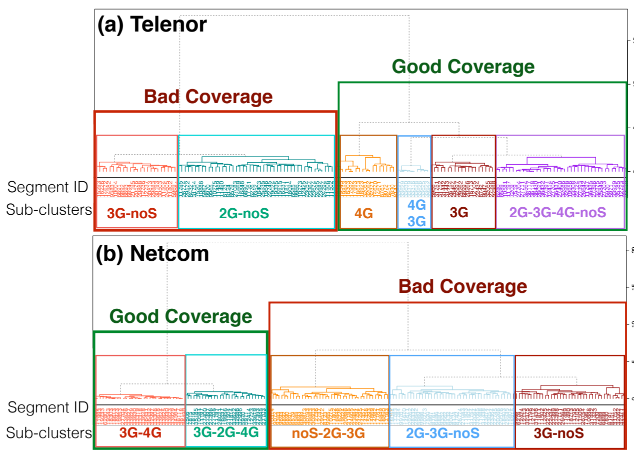


Fig. 3: *Clustering dendrograms showing the prevalent coverage profiles (i.e., Profile A - "good" coverage and Profile B - "bad" coverage) and the coverage sub-profiles along Oslo-Stavanger route, for (a) Telenor and (b) Netcom. We generate the dendrogram of the hierarchical clustering according to the similarity measures we show on the y axis. On the x axis we show the ID for each route segment we cluster.*

and further analyze the impact of sampling on the clustering results.

B. Coverage Clustering Stability

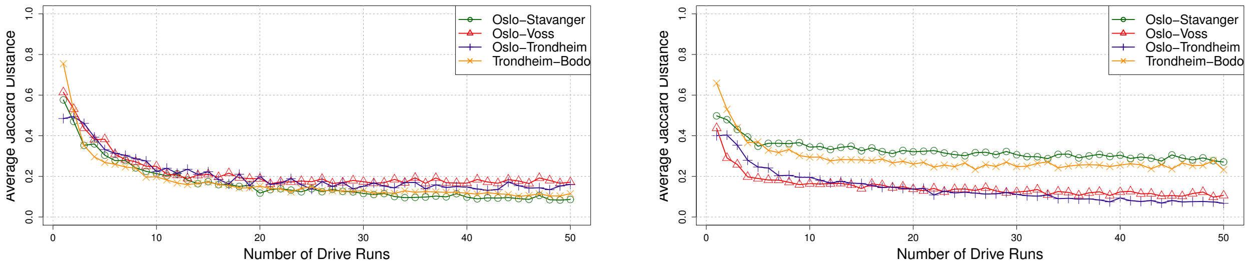
In this section, we focus on coverage cluster stability and investigate the minimum number of runs that is sufficient to classify a segment in one of the coverage profiles. Our goal is to quantify how much additional information regarding the coverage can each run bring to the clustering problem.

To this end, we run the clustering approach we explain in Section III for varying number of runs n between 1 and 50. Using the resulting coverage chart time series from n measurement runs, we cluster the segments and separate them in *Profile A* and *Profile B* coverage clusters. For example, in order to find the clusters of segments using only 2 measurement runs ($n = 2$), we select any pair of runs among all runs and apply the clustering algorithm. We thus obtain an assignation of coverage profile per route segment, for all routes and both operators. For each n, we repeat this exercise for 100 different combinations of n runs out of the ones available on each route. We obtain 100 different assignation of the coverage profiles per segment for each size n of the set of runs we use as input. In order to gauge the differences between the 100 coverage profile assignments for every n input measurement runs, we calculate the similarity between these 100 coverage profile assignments. We use the Jaccard distance [7], which is well-defined for binary vectors (for each segment, we assign 1 for *Profile A* and 0 for *Profile B*).

In Figure 4 we show the average Jaccard distance between the coverage profile assignments as a function of the number of runs we use as input. We show this for each operator and every route we measure. We conclude that individual runs are highly dissimilar, generating highly variable assignations of coverage profiles for the analyzed grid blocks. This is reflected in Figure 4 by the average distance corresponding to the number of runs equal to 1. However, we observe that, in order to obtain a stable coverage profile assignment to each grid block, the minimum number of drive runs required is between 5 and 10. This result is consistent over all the routes, for both operators.

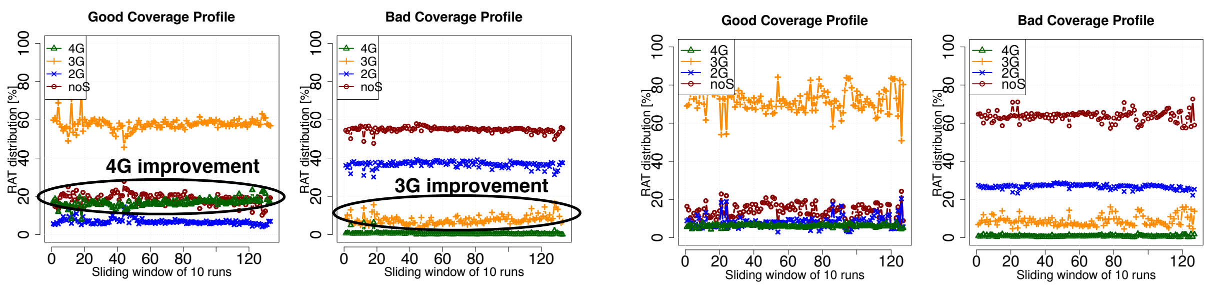
C. Coverage Profile Adaptability

Previously, we determined that 5-10 measurement runs bring enough to decide whether a route segment has *Profile A* or *Profile B* coverage. In this section, we aim to capture the evolution over the measurement period for these two coverage profiles in terms of the average distribution of RATs over the route segments that fall within each coverage cluster. We use a sliding window of 10 measurement runs over the available geo-referenced dataset and run the clustering approach we propose in Section III to assign each segment to the good or bad coverage cluster. This analysis allows us to capture the technology upgrades over the period of 5 months we measure. In Figure 5, we exemplify this analysis for the Trondheim-Bodø route, which is the one where we collect the highest number of measurement runs. We note that the coverage profiles are overall stable for both operators. However, for Telenor we observe a slight increase in the 4G distribution in the areas with *Profile A* coverage. Also, there is a small improvement in the 3G distribution in the areas with *Profile*



(a) Telenor (b) Netcom

Fig. 4: *Stability of coverage profile assignment in function of the number of measurement runs we use to build the coverage chart time series of the route segments.*



(a) Telenor (b) Netcom

Fig. 5: *Evolution of the coverage profiles when using a sliding window of 10 measurement runs to derive them.*

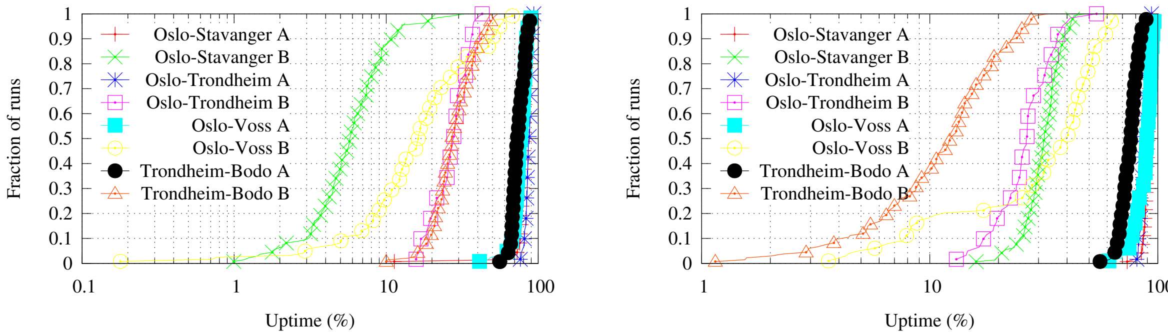
B coverage along the Trondheim-Bodø. This shows that our methodology is capable of capturing technology upgrades in the operators' networks.

V. Coverage Implications: Reliability and Performance

In this section, we focus on how coverage profiles correlate with the performance of the networks from the end-user point of view. This analysis opens the future possibility of using the resulting coverage mosaic with coverage profiles as an indicator for network performance. Furthermore, this will enable the design of contex-aware algorithms to improve the application quality of experience for end-users. We now turn to investigating per-profile MBB performance in terms of downtime and packet loss as key performance measures.

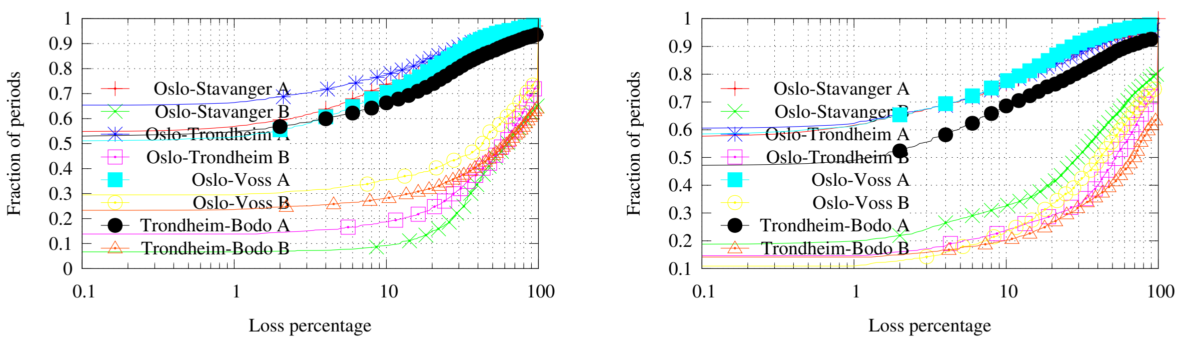
Uptime. To calculate connection uptime in a time window T, we divide the number of sent packet in T by the length of T. In our case, T represents the time that the NNE node spends inside a grid block, and has a minimum value of 30-45 sec. Figures 6a and 6b show the CDF of the fraction of uptime for each route and cluster combination. As expected there are clear differences between areas with different coverage profiles, with the areas with *Profile B* coverage ("bad" coverage) profile exhibiting very low uptime caused by the high percentage of no service. Furthermore, there are clear spatial differences between operators. For instance, the worst performing route in the *Profile B* coverage cluster is Oslo-Stavanger for Telenor and Trondheim-Bødo for Netcom. Trondheim-Bødo is the best performing route with *Profile B* coverage for Telenor. These differences indicate that multi-connectivity, i.e. the use of several operators simultaneously, can improve users experience along the same route. Analyzing differences in uptime between operators and between routes in conjunction with routes coverage profiles (see Fig. 2), indicates a tight coupling between coverage profiles and uptime.

Packet loss. Figures 7a and 7b show the CDF of packet loss in a grid block. We measure a significant difference in the extent of loss between *Profile A* and *Profile B* coverage clusters. Loss, however, remains high in the *Profile A* ("good" coverage) coverage cluster; between 30% and 50% of our samples depending on the operator and route exhibit more than 1% packet loss. This can be explained by the complex interplay of different RATs and the need for frequent handovers in the areas with *Profile A* coverage, which could be otherwise perceived as "good" coverage. Further, we observe that the ranking of good coverage route segments in terms of packet loss matches their ranking in terms of uptime. There is, however, less similarity between the *Profile B* coverage routes loss and uptime rankings; the worst route is always similar. We believe this similarity in ranking is because packet loss is usually experienced in areas with challenging coverage conditions (i.e., larger percentage of no service) which can also lead to a connectivity loss. We also measure how loss in a grid block varies in different runs and find that irrespective of the operator and route, the standard deviation of packet loss is twice as much the mean for at least 50% of the grid blocks. Interestingly, this variability is higher for areas with "Profile A" coverage (which can be perceived as "good" coverage), which underlines the fact that one-off measurements are not enough to make conclusions about performance under mobil-



(a) Telenor (b) Netcom

Fig. 6: *Uptime per route, run and cluster.*



(a) Telenor (b) Netcom

Fig. 7: *The distribution of packet loss in a grid block for different routes.*

ity.

VI. Related Work

Building accurate and reliable coverage maps has been in the attention of the community and a magnitude of work exists in this area. Drive tests are widely used by MBB operators for coverage assessment and performance monitoring. The drawbacks of this technique (e.g., high cost, spatio-temporal sampling) act as incentive for the design of new methodologies that address the before-mentioned issues [11], [12]. In this sense, leveraging crowdsourced dataset may help verify coverage maps [13], but this also suffer from a series of limitations (e.g., lack of control and repeatability). In this paper, we argue that piggy-backing mobile broadband measurements onto public transport infrastructure is an efficient, cost-effective and automated alternative to drive tests.

Data analytics approaches are receiving much attention from the community, due to their capabilities to draw useful information from large databases collected from the network [14], [15]. Coverage prediction methodologies based on geostatistics [16], [17] in wireless networks constitute another approach in the direction of data analytics. To the best of our knowledge, this paper is the first attempt in mobile coverage profiling using hierarchical clustering of multivariate time series. Similar solutions have been proposed in the area of spatio-temporal data mining with different applications in real life e.g. [18], [19]. This technique enables us to generate adaptive coverage profiles, which are based on real measurements and reflect the deployment reality of MBB connectivity solutions and their evolution in time.

Unlike previous efforts that ran performance measurements aboard public transport [20], [21], we focus on coverage with a time-domain component and its implications in terms of performance as experienced by end users. The resulting geo-located mosaic of coverage profiles further enables our analysis in terms of network performance characterization and reliability and whether coverage profiles can be used as an indicator for performance. Along these lines, in the past years we have seen increased interest in the networking community from different parties (e.g., researchers, operators, regulators, policy makers) in measuring the performance of fixed broadband networks e.g. [22], [23] and mobile broadband networks e.g. [24]–[26].

VII. Conclusions and Future Work

MBB networks are the key infrastructure for people to stay connected, especially in high mobility scenarios (e.g., when using public transport). MBB coverage profiling from the end-user experience while on critical public transport routes are of great importance to many stakeholders. At the same time, this is a challenging problem, since even a straightforward classification of coverage into ”good” or ”bad” is very difficult to grasp in quantitative thresholds. In this paper, we evaluate the use of hierarchical clustering to build a coverage mosaic of MBB technologies in an area and analyze its implications in terms of performance. By piggy-backing network measurements onto public transportation vehicles via the NNE platform, we first obtained a unique dataset that (i) captures the coverage and performance from user's perspective and (ii) provides repetitive measurement runs on the same route, in similar conditions. We then leveraged hierarchical clustering in order to identify and characterize prevalent coverage profiles. Though in this study we look at the case of railways in Norway, the methodology can easily be generalized for running a similar study in other regions or applying it to a different datasets, (e.g. crowd-sourced data).

Our results reveal that the clustering approach can accurately group together regions with high similarity in terms of coverage. Based on the mixture of RATs and the time-domain evolution, two main coverage profiles emerge, corresponding to the two main clusters of route segments: *Profile A* -where 3G dominates, and *Profile B* - where *No Service* dominates. This maps onto the general intuition of "good" and, respectively, "bad" coverage. We then analyze the identified coverage profiles, both in terms of stability and performance. The stability analysis investigates the similarity between different runs over the same route, with the express purpose of informing the amount of measurement repetitions we require to accurately observe stable coverage profiles. We find that we need at least between 5 to 10 measurement runs in order to achieve a stable coverage profile in an area. We then focus on how coverage profiles correlate with the performance of the networks from the end-user point of view. For this, we assess packet loss performance per coverage profile. We observe that the packet loss performance highly varies for areas with *Profile A* coverage. This results is counter-intuitive because *Profile A* present a high percentage of superios RATs. This indicates that, although we can derive this profile with few measurement runs, further characterization of the performance requires more analysis, e.g., correlation with the network congestion and measurement time of the day.

Using the coverage mosaic as underlying structure for performance prediction will foster the design of context-aware adaptive algorithms for improving the application performance and, subsequently, the experience of the end-user. For future work, we plan to use the coverage mosaic in the design of tailored solutions that avoids disturbances due to patchy or lack of coverage, hence, improve the application performance and end-user experience. Additionally, we aim to understand how our dataset compares with data generated through crowd-sourced efforts. Our goal is to investigate whether merging different datasets is beneficial and can lead to building more accurate coverage maps.

VIII. Acknowledgements

This work was partially supported by the European Union's Horizon 2020 research and innovation program under grant agreement No. 644399 (MONROE). The views expressed are solely those of the author(s). We would like to thank NSB for hosting our measurement nodes aboard their operational passenger trains.

References

[1] O. B. Karimi, J. Liu, and C. Wang, "Seamless wireless connectivity for multimedia services in high speed trains," *Selected Areas in Communications, IEEE Journal on*, vol. 30, no. 4, pp. 729–739, 2012.

[2] M.-h. Han, K.-S. Han, and D.-J. Lee, "Fast ip handover performance improvements using performance enhancing proxys between satellite networks and wireless lan networks for high-speed trains," in *Vehicular Technology Conference, 2008. VTC Spring 2008. IEEE.* IEEE, 2008, pp. 2341–2344.

[3] A. Kvalbein, D. Baltrūnas, 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.

[4] A. J. Scott and M. J. Symons, "Clustering methods based on likelihood ratio criteria," *Biometrics*, vol. 27, no. 2, pp. pp. 387–397, 1971. [Online]. Available: http://www.jstor.org/stable/2529003

[5] L. Kaufman and P. J. Rousseeuw, *Finding groups in data: an introduction to cluster analysis*. John Wiley & Sons, 2009, vol. 344.

[6] H. Alt and M. Godau, "Computing the fréchet distance between two polygonal curves," *International Journal of Computational Geometry & Applications*, vol. 5, no. 01n02, pp. 75–91, 1995.

[7] P. Jaccard, "The distribution of the flora in the alpine zone," *New phytologist*, vol. 11, no. 2, pp. 37–50, 1912.

[8] J. H. Ward Jr, "Hierarchical grouping to optimize an objective function," *Journal of the American statistical association*, vol. 58, no. 301, pp. 236–244, 1963.

[9] Y. Liu, Z. Li, H. Xiong, X. Gao, and J. Wu, "Understanding of internal clustering validation measures," in *IEEE International Conference on Data Mining (ICDM)*. IEEE, 2010, pp. 911–916.

[10] P. J. Rousseeuw, "Silhouettes: a graphical aid to the interpretation and validation of cluster analysis," *Journal of computational and applied mathematics*, vol. 20, pp. 53–65, 1987.

[11] Tektronix, "Reduce Drive Test Costs and Increase Effectiveness of 3G Network Optimization," Tektronix Communications, Tech. Rep., 2009.

[12] W. A. Hapsari, A. Umesh, M. Iwamura, M. Tomala, B. Gyula, and B. Sebire, "Minimization of drive tests solution in 3gpp," *Communications Magazine, IEEE*, vol. 50, no. 6, pp. 28–36, 2012.

[13] M. Molinari, M.-R. Fida, M. K. Marina, and A. Pescape, "Spatial interpolation based cellular coverage prediction with crowdsourced measurements," in *Proceedings of the 2015 ACM SIGCOMM Workshop on Crowdsourcing and Crowdsharing of Big (Internet) Data*. ACM, 2015, pp. 33–38.

[14] I. Leontiadis, A. Lima, H. Kwak, R. Stanojevic, D. Wetherall, and K. Papagiannaki, "From cells to streets: Estimating mobile paths with cellular-side data," in *Proceedings of the 10th ACM International on Conference on emerging Networking Experiments and Technologies*. ACM, 2014, pp. 121–132.

[15] F. Murtagh and A. E. Raftery, "Fitting straight lines to point patterns," *Pattern recognition*, vol. 17, no. 5, pp. 479–483, 1984.

[16] B. Sayrac, A. Galindo-Serrano, S. B. Jemaa, J. Riihijärvi, and P. Mähönen, "Bayesian spatial interpolation as an emerging cognitive radio application for coverage analysis in cellular networks," *Transactions on Emerging Telecommunications Technologies*, vol. 24, no. 7-8, pp. 636–648, 2013.

[17] H. Braham, S. Ben Jemaa, B. Sayrac, G. Fort, and E. Moulines, "Low complexity spatial interpolation for cellular coverage analysis," in *Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt), 2014 12th International Symposium on*. IEEE, 2014, pp. 188–195.

[18] T. Dasu, D. F. Swayne, and D. Poole, "Grouping Multivariate Time Series: A Case Study," in *Proceedings of the IEEE Workshop on Temporal Data Mining: Algorithms, Theory and Applications, in conjunction with the Conference on Data Mining, Houston*, 2005, pp. 25–32.

[19] A. Strehl, J. Ghosh, and R. Mooney, "Impact of similarity measures on web-page clustering," in *Workshop on Artificial Intelligence for Web Search (AAAI 2000)*, 2000, pp. 58–64.

[20] 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.

[21] F. P. Tso, J. Teng, W. Jia, and D. Xuan, "Mobility: A Double-Edged Sword for HSPA Networks," in *Proc. of MobiHoc*, 2010.

[22] S. Sundaresan, W. de Donato, N. Feamster, R. Teixeira, S. Crawford, and A. Pescapè, "Broadband Internet performance: A view from the gateway," *SIGCOMM CCR.*, vol. 41, pp. 134–145, 2011.

[23] FCC, "2013 Measuring Broadband America February Report," FCC's Office of Engineering and Technology and Consumer and Governmental Affairs Bureau, Tech. Rep., 2013.

[24] OFCOM, "2014 Measuring mobile broadband performance in the UK 4G and 3G network performance," OFCOM UK, Tech. Rep., 2014.

[25] J. Sommers and P. Barford, "Cell vs. WiFi: On the Performance of Metro Area Mobile Connections," in *Proc. of IMC*, 2012.

[26] D. Baltrūnas, A. Elmokashfi, and A. Kvalbein, "Measuring the Reliability of Mobile Broadband Networks," in *Proc. of IMC*, 2014.