

Predictive Impact Analysis for Designing a Resilient Cellular Backhaul Network

Sen Yang
Georgia Institute of Technology
sen.yang@gatech.edu

Yan He
AT&T Labs - Research
yanhe@research.att.com

Zihui Ge
AT&T Labs - Research
gezihui@research.att.com

Dongmei Wang
AT&T Labs - Research
mei@research.att.com

Jun Xu
Georgia Institute of Technology
jx@cc.gatech.edu

ABSTRACT

Backhaul transport network design and optimization for cellular service providers involve a unique challenge stemming from the fact that an end-user's equipment (UE) is within the radio reach of multiple cellular towers: It is hard to evaluate the impact of the failure of the UE's primary serving tower on the UE, because the UE may simply switch to get service from other nearby cellular towers. To overcome this challenge, one needs to quantify the cellular service redundancy among the cellular towers riding on that transport circuit and their nearby cellular towers, which in turn requires a comprehensive understanding of the radio signal profile in the area of the impacted towers, the spatial distribution of UEs therein, and their expected workload (e.g., calls, data throughput). In this work, we develop a novel methodology for assessing the service impact of any hypothetical cellular tower outage scenario, and implement it in an operational system named Tower Outage Impact Predictor (TOIP). Our evaluations, using both synthetic data and historical real tower outages in a large operational cellular network, show conclusively that TOIP gives an accurate assessment of various tower outage scenarios, and can provide critical input data towards designing a reliable cellular backhaul transport network.

CCS CONCEPTS

• **Networks** → **Network reliability**; *Wireless access points, base stations and infrastructure; Mobile networks*;

KEYWORDS

Tower outage impact prediction; cellular network

ACM Reference Format:

Sen Yang, Yan He, Zihui Ge, Dongmei Wang, and Jun Xu. 2018. Predictive Impact Analysis for Designing a Resilient Cellular Backhaul Network. In *SIGMETRICS'18 Abstracts: ACM SIGMETRICS International Conference on Measurement & Modeling of Computer Systems Abstracts, June 18–22, 2018, Irvine, CA, USA*. ACM, New York, NY, USA, 3 pages. <https://doi.org/10.1145/XXXXXX.XXXXXX>

Permission to make digital or hard copies of part or all 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 third-party components of this work must be honored. For all other uses, contact the owner/author(s).

SIGMETRICS'18 Abstracts, June 18–22, 2018, Irvine, CA, USA

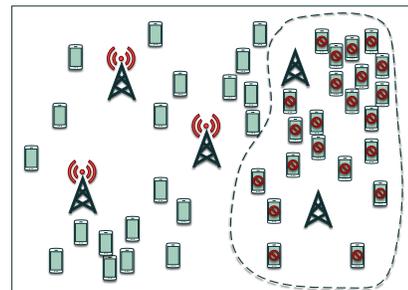
© 2018 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-5846-0/18/06.

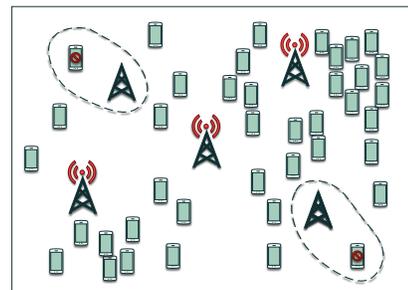
<https://doi.org/10.1145/XXXXXX.XXXXXX>

1 INTRODUCTION

The design and optimization of a layer-1/2 transport network requires critical inputs that can be broadly divided into three different aspects: risk, service impact, and cost. Risk models the likelihood of the occurrence or co-occurrence of various types of outages, such as a fiber cut or component failure, given a certain network topology and routing design. Service impact quantifies the consequences (e.g., loss-of-service) of different failure scenarios. Cost reflects the expenditures associated with the implementation and maintenance of a certain (failure-resistant) network design. For cellular service providers, the service impact analysis can be extremely challenging when designing the backhaul transport network connecting to and from the cellular towers.



(a) Significant service impact when tower outages are clustered



(b) Negligible service impact when tower outages are scattered

Figure 1: Service impact varies dramatically under different outage scenarios with similar number of cellular towers involved

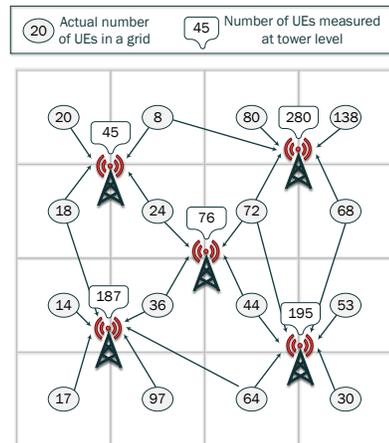
This challenge stems from the built-in redundancy of radio links. Typically an end-user's equipment (UE) is within the radio reach

of multiple cellular towers. Hence, not every cellular tower outage scenario has a significant impact on the service it provides, as nearby towers are often able to take over the servicing of a certain percentage of UEs abandoned by the failed tower(s). For example, multiple simultaneous tower outages that are close-by usually lead to much larger “out-of-service” areas than scattered simultaneous outages. Furthermore, the actual service impact of an outage also depends on the spatial distribution of UEs surrounding, and the workloads they impose on, the impacted towers (including the failed towers and other towers close to them). The service impact could vary significantly when a similar outage happens in different areas (e.g., metropolis versus suburb) and different time periods (e.g., morning versus midnight). Figure 1 illustrates two tower outage scenarios in which the service impact (measured by the number of UEs that lose cellular services) varies significantly while the number of out-of-service cellular towers is the same. Understanding and quantifying the service impact of a tower outage scenario calls for a comprehensive understanding of the radio signal profiles in the areas surrounding the impacted towers, the spatial distribution of UEs therein, and the workloads these UEs impose on the impacted towers. None of this information is easy to obtain.

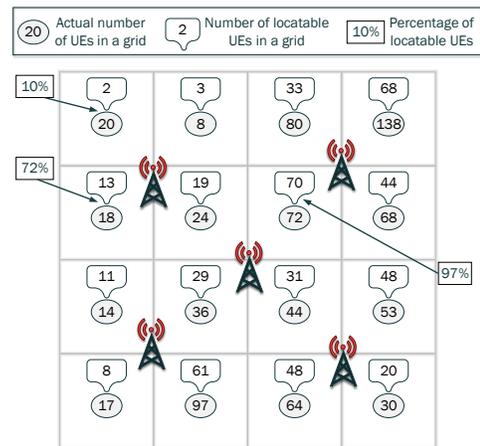
In this paper, we focus on the problem of predicting the service impact on UEs given any hypothetical cellular-tower outage scenario, using two existing data sources collected from a major cellular provider network in North America as a part of routine cellular network operations. These two data sources, and the challenges in using them for solving this prediction problem, are illustrated in Figure 2, in which the cellular service area is divided into small square-shaped grids of the same size. The first data source, shown in Figure 2a, is the tower-level data that includes the accurate counts of the total number of UEs served by each tower. However, as illustrated in Figure 2a, the data shed no light on the spatial distribution of UEs in different grids. For example, in Figure 2a, virtually no information can be learned from the total UE count of 45 recorded on the top-left tower concerning the actual UE counts (20, 8, 18, and 24) in the four grids surrounding the tower. The spatial distribution of UEs is however critical for predicting the service impact on UEs given any hypothetical cellular-tower outage scenario, as just explained in the above paragraph. The second data source, shown in Figure 2b, is the grid-level aggregations of UE-level measurements: the counts and the radio profiles of UEs within each grid. However, these UE-level measurements are not accurate either, since they account for only those UEs that are locatable (i.e., whose geo-locations can be determined by the system); the likelihood for a UE to be locatable (called “locatability” in the sequel), is roughly proportional to its signaling activity level, which can be non-uniform over space. For example, in Figure 2b, this likelihood ranges from 10% to 97%.

2 OUR APPROACH

In this work, we develop the following three-step methodology to overcome the challenge in using these two inaccurate and incomplete data sources for predicting the service impact of a cellular-tower outage. First, we construct the radio signal profiles for each grid (e.g., a list of cellular towers covering the grid) based on the (locatable) UE-level measurements associated with each grid. We



(a) Accurate number of UEs measured on each cellular tower. Note when a UE is connected to k towers, each tower only count it as $1/k$ to avoid double counting.



(b) Inaccurate number of UEs in each grid due to the non-uniform “locatability” of UEs over space. For example, UEs under coverage of multiple towers are typically more locatable while UEs at the foot of a tower are typically less locatable.

Figure 2: Examples of tower-level and grid-level measurements for the same area as shown in Figure 1. We divide this area into 4×4 grids to better illustrate the idea.

refer to this first step as *radio signal profiling*. Second, we propose a ridge regression model that can fairly accurately estimate the time series of the actual UE counts in each grid from that of the locatable UE counts contained in the second data source and that of the tower-level UE counts contained in the first data source. We refer to this second step as *grid-level UE count calibration*. The resulting estimator is unbiased and efficient, and has an appealing property of being decomposable in the following sense. The original computational problem (of the ridge regression) has a gigantic size of $|S| \times |G|$, where $|S| \approx 10^5$ is the total number of towers or antennas and $|G| \approx 10^9$ the total number of grids across US, and hence is prohibitively expensive. Under this estimation framework,

however, it can be decomposed into $|S|$ subproblems, each of which is only of size $O(1)$ and can be solved separately. This decomposition not only significantly reduces the total size (and hence the total computational complexity or, in parallel computing terms, the total amount of *work*) of this estimation problem, but also allows the best possible running time (i.e., the *depth* in parallel computing terms) to be further reduced, possibly by several orders of magnitude, via massively parallel computing. Third, for any tower outage scenario, we determine the cellular service survivorship of each grid based on its radio signal profile, and estimate, for each unsurvivable grid, the number of UEs in it (that will lose service as a result) according to the parameters computed from the second step. We refer to this third step as *survival analysis*. While this three-step methodology is used mainly for the service impact analyses of backhaul outages in this work, it can be generalized to a broader category of tower outages such as those caused by power failures or planned maintenance events.

We summarize the contributions of this work as follows. The full version of this paper is available in [1].

- (1) We suggest that cellular service impact of tower outages be taken into account in designing a cellular backhaul transport network and identify the major challenges in predicting this impact.
- (2) We propose a novel data-driven three-step methodology for predicting this impact using two types of data collected from a major cellular provider network in North America, and implement the methodology in a system called *Tower Outage Impact Predictor (TOIP)*.
- (3) We evaluate TOIP using both synthetic data in a simulation setup and historical real tower outages in a large operational cellular network. For a diverse set of real outages in the network, we compare the prediction (in retrospect) from TOIP and the ground truth provided by operators and show that the overall prediction accuracy is above 90%.
- (4) We further demonstrate the efficacy of TOIP in backhaul network re-optimization via simulation studies on a large real-world cellular network topology. The simulation results show that the use of TOIP can lead to design changes in the backhaul network that would result in significantly improved resiliency to failures.

3 EVALUATION

In this section, we evaluate the performance of TOIP by comparing with the ground truth from historical real tower outages in a large operational cellular network¹. We trained TOIP using 45 days of measurement data collected from a large operational cellular network. In order to evaluate the prediction accuracy of TOIP, we collect a set of 43 historical real tower outages that occurred in a 3-month time period. The number of out-of-service towers varies from 1 to about 50 in these outage cases. The ground truth of service impact (in terms of number of UEs that lost cellular service) for these outage cases are generated by network operators using reliable network monitor and analysis tools. The operator-reported service impacts are normalized (as required by the service provider

¹Please refer to our full paper [1] for more evaluation experiments.

we are working with), which served as the ground truth in our evaluation.

As shown in Table 1, the operation team classifies the service impact (in terms of number of UEs that lost cellular service) of different outages into 6 categories according to their severities. In other words, minor errors in the prediction can be ignored as long as they are not crossing categories. For example, on one hand, it is not critical if TOIP predicts the service impact as 0.01 while the ground truth is 0.011 as they both are still in the same category. But, on the other hand, we call it an error if the prediction and ground truth are in two different categories.

Table 1: Outage severity categories

Outage severity category	Service impact (normalized)
Level 1	0 ~ 0.002
Level 2	0.002 ~ 0.02
Level 3	0.02 ~ 0.1
Level 4	0.1 ~ 0.2
Level 5	0.2 ~ 0.6
Level 6	> 0.6

As shown in Figure 3, for the above 43 real outages in the network, we compare our “predicted” (in retrospect) number of UEs that lost service with the operator-reported service impact. We find remarkably good matches between our prediction and the ground truth - showing a prediction accuracy of 92.68%, which is high enough for operational use.

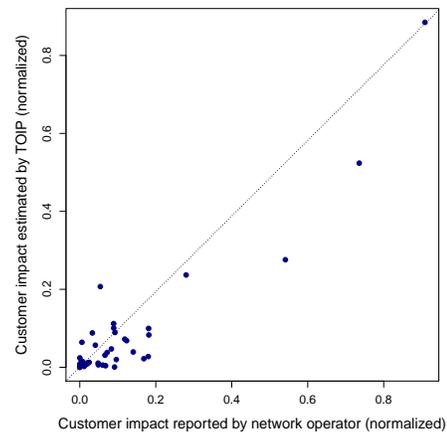


Figure 3: Customer impact estimation made by TOIP

Acknowledgment: This project is supported in part by NSF grants CNS-1423182 and CNS-1218092.

REFERENCES

- [1] Sen Yang, Yan He, Zihui Ge, Dongmei Wang, and Jun Xu. 2017. Predictive Impact Analysis for Designing a Resilient Cellular Backhaul Network. *Proc. ACM Meas. Anal. Comput. Syst.* 1, 2, Article 30 (Dec. 2017), 33 pages. <https://doi.org/10.1145/3154488>