Predictive Impact Analysis for Designing a Resilient Cellular Backhaul Network

SEN YANG, Georgia Institute of Technology, USA
YAN HE, AT&T Labs - Research, USA
ZIHUI GE, AT&T Labs - Research, USA
DONGMEI WANG, AT&T Labs - Research, USA
JUN XU, Georgia Institute of Technology, USA

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*;

Additional Key Words and Phrases: Tower outage impact prediction; cellular network

ACM Reference Format:
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 (December 2017), 33 pages. https://doi.org/10.1145/3154488

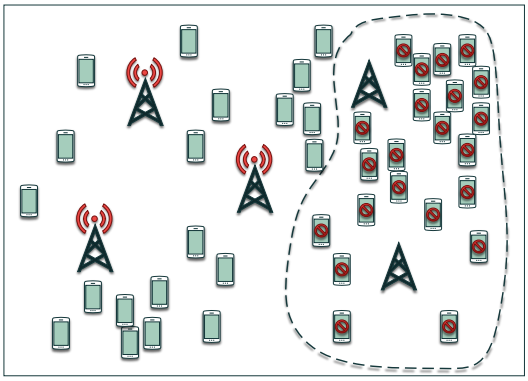
1 INTRODUCTION

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

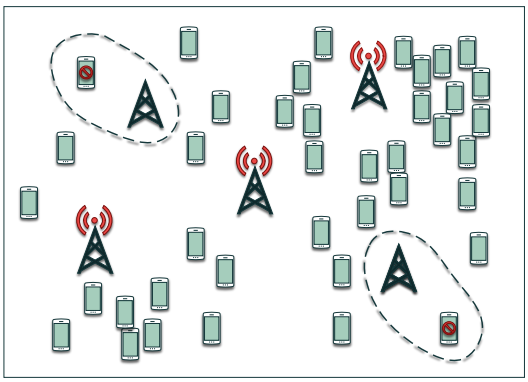
Authors' addresses: Sen Yang, Georgia Institute of Technology, School of Electrical and Computer Engineering, Atlanta, GA, USA, sen.yang@gatech.edu; Yan He, AT&T Labs - Research, Bedminster, NJ, USA, yanhe@research.att.com; Zihui Ge, AT&T Labs - Research, Bedminster, NJ, USA, gezihui@research.att.com; Dongmei Wang, AT&T Labs - Research, Bedminster, NJ, USA, mei@research.att.com; Jun Xu, Georgia Institute of Technology, School of Computer Science, Atlanta, GA, USA, jx@cc.gatech.edu.

2476-1249/2017/12-ART30 $15.00
https://doi.org/10.1145/3154488

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

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

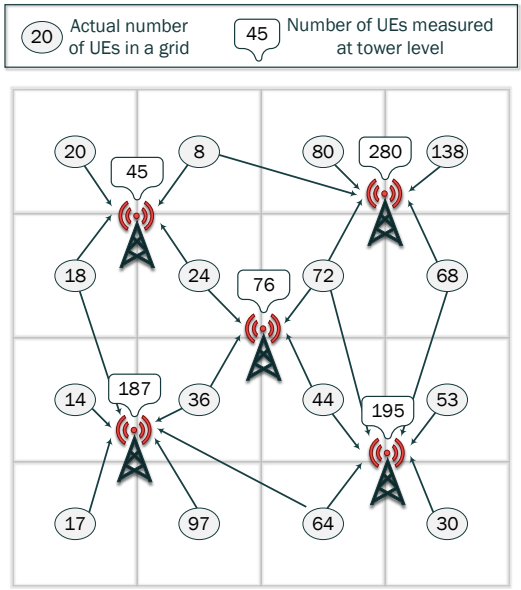
This challenge stems from the built-in redundancy of radio links. Typically an end-user's equipment (UE) is within the radio reach of multiple cellular towers. Hence, not every cellular tower outage scenario has a significant impact on the service it provides, as nearby towers are often able to take 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.

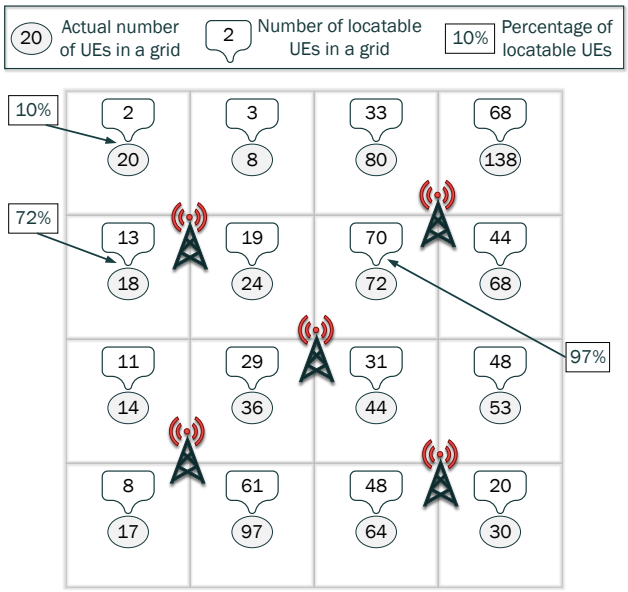
In this paper, we focus on the problem of predicting the service impact on UEs given any hypothetical cellular-tower outage scenario, using two existing data sources collected from a major cellular provider network in North America as a part of routine cellular network operations. These two data sources, 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 hypothetic 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%.

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 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:



(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.

Fig. 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.

(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.

The rest of the paper is organized as follows. In Section 2, we provide important background on the cellular backhaul transport design and radio link redundancy. We formulate the problem and present our solution strategies in Section 3. We evaluate the performance of TOIP experimentally in Section 4. In Section 5, we present and evaluate a real application of TOIP, which solves the tower NTE rehoming problem in the backhaul networks. In Section 6, we discuss the background and related work most pertinent to this paper. Finally, we conclude the paper in Section 7.

2 BACKGROUND

In this section, we present a high level overview of the backhaul transport network that connects to and from the cellular towers, and the built-in redundancy of cellular radio links that can obscure the service impact of individual tower outages. We also describe the two aforementioned datasets that make it possible to predict the impact of any hypothetical tower outage scenario. We emphasize that, although these two datasets are gathered from a 3G UMTS network, the mathematical formulation of our aforementioned three-step prediction methodology is general enough for the methodology to be applicable to networks with other underlying technologies, such as LTE and 5G, as long as the corresponding measurement data (i.e., tower-level workloads, UE distributions, and tower-UE associations) for such networks are available.

2.1 Mobility Backhaul

The backhaul transport network of a cellular service provider provides connectivity between its cellular towers, where mobile base stations (2G), NodeBs (3G), or eNodeBs (4G) are located, and its Mobile Telephone Switching Offices (MTSOs), where various voice and data applications are hosted (or aggregated and routed further upstream for processing). These mobility backhauls are usually provided via Ethernet Virtual Circuits (EVCs) over a metro Ethernet transport.

Since a cellular carrier has to provide (cellular) access to a massive geographic area (e.g., the entire continental US), most carriers lay out their metro Ethernet transports in a hierarchical (or "hub-and-spoke") manner. A typical such layout is as follows. A carrier-owned termination switch, called Network Terminating Equipment (NTE), is placed on a cellular tower and connects to Ethernet switches (each called an Emux) in Wire Centers (WCs) via a direct *feeder* fiber link or loop. Several Emux's connect to a small router that is usually located at another larger WC. Several such small routers then *home* (i.e., connect) to a larger router. For reliability purposes, such larger routers form a mesh topology with single failure restoration capability while smaller routers are dual-homed to a pair of larger routers via different uplinks. There can be variations to this typical

layout in a cellular carrier network however. For example, a cellular site NTE may bypass an Emux and connect directly to a smaller router when it is less costly to do so. For another example, a chain of point-to-point microwave links can be used instead of a (wireline) direct feeder or loop link over hard-to-access terrain or over water.

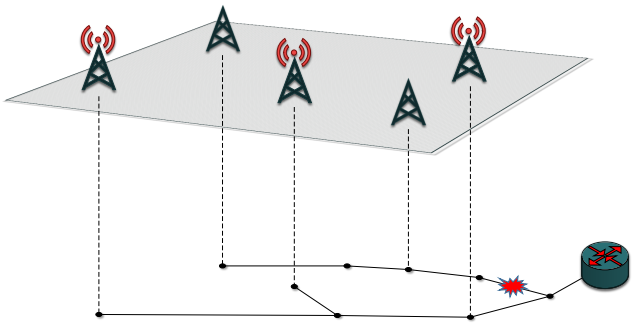


Fig. 3. A single fiber cut brings down backhaul links to 2 cellular sites

While a feeder fiber sub-network usually forms a tree structure rooted at a WC, the inter-WC paths typically route over segments of direct fiber or Dense Wavelength Division Multiplex (DWDM) transmission systems. Multiple fibers may each have a segment share a common conduit and risk being cut together when this common conduit is cut. Hence for service reliability, diversity routing should be used whenever possible to minimize the service impact due to such a fiber/conduit cut. Figure 3 illustrates an example backhaul network that connects the five towers shown in Figure 1 and a cut to a fiber link that results in the tower outage scenario shown in Figure 1b.

To evaluate different network topology layouts and routing designs for cellular service reliability, network planners need to understand the associated risks and costs, such as the probabilities of various types of failure scenarios, the service impact (e.g., loss-of-service) when the corresponding set of cellular towers lose network connectivity, and the costs for possible reliability enhancement strategies (e.g., using a diversely-routed fiber segment to replace the aforementioned common conduit). However, due to the inherent redundancy of cellular radio links, the service impact of cellular towers losing network connectivity is hard to assess, which we will discuss next.

2.2 Redundancy in Cellular Radio Links

In this subsection, we discuss the radio link redundancy in a 3G UMTS networks. While the exact mechanisms for realizing the service redundancy through radio links are quite different in 2G or 4G, the intrinsic requirement for radio redundancy remains the same: to smoothly hand over an ongoing voice/data call from one cellular tower to another.

Figure 4 shows the overall architecture for a 3G UMTS network. It consists of a UMTS Radio Access Network (UTRAN) and a UMTS core network. As we are primarily concerned with cellular tower outages caused by failures in the backhaul transport, we will focus only on the UTRAN part in the sequel. The UTRAN consists primarily of the User Equipments (UE), the NodeBs (at the cellular towers), and the Radio Network Controllers (RNCs, at the MTSOs). A cellular tower performs wireless transmissions/receptions to/from the UEs via its Uu radio interfaces, and communicates with the RNC via an Iu-B link routed on top of the underlying backhaul transport network. The geographical area covered by a UMTS cellular tower is typically divided into between 3 and 9 sectors, each of which is covered by a directional antenna using a different frequency. At a cellular

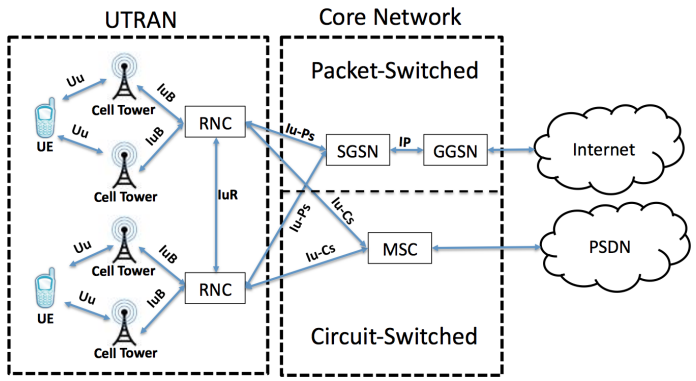


Fig. 4. 3G UMTS network architecture

tower, each directional antenna is uniquely characterized by its azimuth angle, tilt, and transmission power. These parameters, along with the terrain and landscape features around the cellular tower, determine the coverage range of the sector served by the antenna. This coverage range is, however, extremely difficult to compute accurately in reality; for example, it is hard, if not impossible, to accurately incorporate the terrain or building information into the mathematical model of the radio coverage. Each RNC typically manages tens to hundreds of cellular towers and serves as their gateway to the UMTS core network.

Once a UE is on the 3G UMTS network, the associations between the UE and the sectors/towers are managed via maintaining a so-called active Radio Link Set (RLS) at both the UE and the RNC ends. The RLS contains the identities of the sectors that a UE is simultaneously listening to and transmitting to. These sectors may belong to different towers, which could be controlled by different RNCs. When receiving data from the UE, all sectors in the RLS participate in receiving and decoding the data frames, and when transmitting data to the UE, a primary sector in the RLS is responsible for sending the data frames. Each UE continuously monitors the received signal strengths from all "visible" sectors and communicates this information to its RNC periodically so that an RNC is able to optimize the channel allocations for all UEs in its "jurisdiction". Any update to an RLS (addition of a new sector or removal of an existing one) is signaled back to the corresponding UE and the affected cellular tower(s). This signaling and RLS updating mechanism in UTRAN not only supports user mobility, but also facilitates load-rebalancing and fault-tolerance in the event of a cellular outage.

2.3 Measurement Data

As mentioned in Section 1, for this service impact prediction analysis, we rely on two types of measurement data collected from a major cellular provider network in North America: tower-level measurements and UE-level measurements.

2.3.1 Tower-level measurements. The first type of data are the tower-level measurements, which are either the UTRAN counters or metrics derived from them. In a UMTS network, the Element Management System (EMS) for NodeBs and RNCs maintains a wide range of service load and performance measurement counters. Similar to the SNMP counters in routers, these UMTS counters track the frequency of network/service events (e.g., the number of successful/unsuccessful Radio Resource Control (RRC) requests), the packets and bytes flowing through each network interface, and other service and device self-monitoring metrics (e.g., queue length and CPU utilization). In

the cellular provider network that we study, there are over 300 different types of UMTS counters operating at three different levels: sector-level, tower-level, and RNC-level. Each counter is tracked and collected at 15-minute granularity. These raw counter values are further aggregated (e.g., to the hourly granularity) and processed according to a vendor- or carrier-defined formula into service metrics reflecting (1) the service demand/load such as the number of connection requests, the Erlang, or the data volume, and (2) the service performance quality such as the accessibility (e.g., voice and data call setup success rate and delay), the retainability (e.g., call drops rate), the mobility (e.g., hand-over frequency and delay), and the application performance (e.g., voice call quality and data throughput). Roughly tens of gigabytes of raw measurements are generated per month for each type of the UMTS counters throughout US, and the aggregated data, which we use to make the prediction, are about one order of magnitude smaller in size.

To quantify the service impact of cellular tower outages, some loss-of-service metrics are often needed. One key metric used in network operations to quantify service load is the number of UEs in service, which is derived from a counter statistic associated with each sector called "number-of-UEs". An instance of this statistic tracks the time-averaged value of the total number of UEs associated with a certain sector within the data collection time window. Each such counter is maintained (i.e., updated when needed) at the corresponding RNC periodically (e.g., every second). To ensure that each UE is counted exactly once by all the counters associated with the sectors (say there are k of them) belonging to the UE's active RLS, the counter associated with each such sector is incremented by $1/k$.

2.3.2 UE-level measurements. The second type of data are UE-level measurements concerning the location and the signal strength information of each UE. Roughly tens of terabytes of raw data are generated per month throughout US for such measurements. Its grid-level aggregation, which we use for the service impact predictions, is about one order of magnitude smaller. The location estimates of UEs are based on two type of information: GPS coordinates (if available) and locations inferred from relative timing offset information reported by individual UEs [5]. GPS coordinates are always the preferred information (source) if available. However, UEs know their GPS coordinates only if they are exposed to satellites, which essentially limits the availability of GPS coordinates information to the UEs that are outdoors, in-vehicles or by windows. When GPS coordinates are unavailable, the locations of UEs can be inferred by measuring the propagation delay between the UE and one or more cellular sites[1]. The timing offsets of the pilot signals from different sectors are calibrated using the locations of UEs in the region that have access to GPS/satellites. This calibration process takes place aperiodically. The calibration-adjusted timing offset information from multiple cellular towers is then used to geo-locate the UEs in the region via identifying the intersection points of different hyperbolic curves in reference to the GPS coordinates of these cellular towers. A system located at each MTSO observes all signaling message exchanges between the UEs in the region and the corresponding RNCs, each of which contains the Received Signal Strength Indication (RSSI) and Time Difference of Arrival (TDoA) information for the reported sectors visible to a certain UE. All such RSSI and TDoA information is used collectively to geo-locate the UEs in the region. A location determined by this technique is typically accurate to within a few meters.

Since only a small fraction of all UEs know their own GPS coordinates and to obtain the non-GPS-based location information of a UE requires the UE to exchange signaling messages with RNC concerning multiple cellular towers, not all UEs are "locatable". As explained above, the "locatability" of UEs is roughly proportional to their signaling activity levels, which can vary significantly from one grid to another. On one hand, a UE in a grid covered by multiple towers is more "locatable" as it is more likely to exchange signaling messages with the RNC(s) for call handoffs. On the other hand,

[1]Each site consists of a tower and its affiliated facilities.

a UE in a grid dominated by a single tower (e.g., at the foot of a tower) is harder to locate using the geo-locating algorithm described above due to the lack of multiple aforementioned hyperbolic curves that need to be intersected. Each UE-level measurement data item contains a timestamp, the location estimate (latitude-longitude coordinates) of the UE, and a list of cellular sectors and their corresponding radio signal qualities at the UE. The UE identification information is anonymized in this dataset to protect the privacy of mobile users.

3 TOWER OUTAGE IMPACT PREDICTION

To evaluate the potential impact of a transport circuit outage, one need to understand and quantify the service impact when a group of its associated cellular towers become out-of-service. This 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 service workload. In this section, we describe our methodology in deriving these information from the measurement data that we obtain.

3.1 Notation and Problem Formulation

We denote as S the set of radio sectors in the provider network. For each sector $s \in S$, the service load measured on the sector (over time) is a time series $\{y_{s,t}\}$, where $t \in T$ indicates the time bin from which the measurement data are collected, and T is the set of such time bins. As described in Section 2.3.1, this measurement is derived from the UTRAN counter values in hourly aggregates that are contained in the aforementioned tower-level measurement. We use *the number of UEs in service* as the service load metric as it is commonly used operationally when loss-of-service is reported. Note that our approach remains applicable when other metrics, such as bytes transferred, are used. We will refer to this dataset as $\mathcal{D}_{sector}$ in the rest of this paper.

We divide the geographical space within the service area of the provider network into small grids. Based on the average location accuracy of the deployed geo-locating algorithm, which ranges from meters to tens of meters, we set the grid size to 100-meter by 100-meter. We denote as G the set of geographic grids. We aggregate the aforementioned UE geo-location measurement data (described in Section 2.3.2) both in time and in space, and report the observed number of UEs as a time series $\{x_{g,t}\}$ for each grid $g \in G$, where $t \in T$. We will refer to this dataset as $\mathcal{D}_{grid}$ in the rest of this paper.

Lastly, based on the reported radio signal quality measures in the UE geo-locating measurement data, we can also derive the association mapping of the geographic grids G to the subset of sectors S that can deliver a functional level of radio signal strength to the grid. We assume such radio footprint is relatively stable in the course of days to weeks. The association relationship is defined by a binary matrix C of size $|S|$ by $|G|$; a matrix element $c_{s,g} = 1$ if and only if sector $s \in S$ has radio coverage for grid $g \in G$. We refer to this data as $\mathcal{D}_{assoc}$ in the rest of this paper. The data processing described above corresponds roughly to the "radio signal profiling" step of our three-step methodology described in in Section 1.

We define the hypothetic outage scenario as a tuple $O = (S_O, t_O)$, where $S_O \subseteq S$ is the set of out-of-service sectors and $t_O \in T$ is the time bin of this outage. Outages spanning over multiple time bins can be straightforwardly decomposed into multiple sub-outages that each lasts a single time bin. Our objective is to estimate the loss-of-service due to the service outage at S_O, which we denote as H_O.

If there were no radio coverage redundancies among sectors, H_O can simply be computed as

$$H_O = \sum_{s \in S_O} y_{s,t_O}$$

However, with redundant radio coverages, the actual service impact is smaller than the above, and has to be determined at the grid level. We define as $G_O \subseteq G$ the geographical space that loses service coverage during outage O:

$$G_O \triangleq \{g \in G \mid c_{s,g} = 0 \; for \; \forall \, s \in S - S_O\}$$

The analysis of G_O corresponds roughly to the "survival analysis" step of our three-step methodology.

We define the service workload, measured by number of UEs, in each grid g at time t as random variables $z_{g,t}$. It follows that

$$H_O = \sum_{g \in G_O} z_{g,t_O} \tag{1}$$

Our objective is to identify an unbiased estimator $\hat{H}_O$ for H_O based on our measurements, which ideally also has a low variance. This is a challenging goal since the grid-level measurements in $\mathcal{D}_{grid}$ is "skewed" by the non-uniform locatability of UEs, which is hard to "invert" by itself. Fortunately, the dataset $\mathcal{D}_{sector}$ is quite accurate, which allows a ridge regression model to be used for calibrating $\mathcal{D}_{grid}$ with $\mathcal{D}_{sector}$ to arrive at a more accurate estimator. This process, to be described next, is the "grid-level UE count calibration" step of our three-step methodology.

The skewness in $\mathcal{D}_{grid}$ can be modeled as a sampling process: For example, if there are 100 UEs in a certain grid but only 70 of them are located by our system, we "imagine" that these located UEs are obtained by sampling the 100 UEs with a sampling ratio of 70%. Let $p_{g,t}$ denote the sampling ratio of grid g at time t, i.e., $p_{g,t} \triangleq x_{g,t}/z_{g,t}$. Let $r_{g,t}$ be the inverse of $p_{g,t}$ so that $z_{g,t} = r_{g,t} \cdot x_{g,t}$. With a slight abuse of notation, we define $x_{g,t_O} \triangleq p_{g,t_O} \cdot z_{g,t_O}$. Our estimator $\hat{H}_O$ is then defined as

$$\hat{H}_O \triangleq \sum_{g \in G_O} \hat{r}_g \cdot \hat{x}_{g,t_O} \tag{2}$$

where $\hat{r}_g$ is a time-independent estimator for $\mathbf{E}\left[r_{g,t}\right]$ and $\hat{x}_{g,t_O}$ is an unbiased estimator for x_{g,t_O}. The detailed definition of $\hat{r}_g$ and $\hat{x}_{g,t_O}$, through a ridge regression formulation, will be given in Section 3.2.

As will be shown shortly, the service impact estimator $\hat{H}_O$ defined above is unbiased, and arguably also has a relatively small mean square error. Detailed proof and discussion will be provided in Section 3.3 and Appendix A.

3.2 Definition of the Estimators

We now introduce the detailed definition of $\hat{r}_g$ and $\hat{x}_{g,t_O}$. A more rigorous justification for the following formulation will be provided in Appendix A.1.

3.2.1 Definition of $\hat{r}_g$. The estimated scaling factors $\{\hat{r}_g\}$ are defined based on the following arguments. For any given time bin t, let $a_{s,g} \in [0,1]$ be the average proportion of UEs that are located in grid g and connected to radio sector s, such that $\mathbf{E}[y_{s,t}] = \sum_{g \in G} a_{s,g} \cdot \mathbf{E}\left[z_{g,t}\right]$ for each sector $s \in S$. Let $\{r_g\}_{g \in G}$ be a group of scaling factors such that $\mathbf{E}\left[z_{g,t}\right] = r_g \cdot \mathbf{E}\left[x_{g,t}\right]$ for all $g \in G$. Then for any sector $s \in S$, we have

$$\begin{aligned} \mathbf{E}[y_{s,t}] &= \sum_{g \in G} a_{s,g} \cdot r_g \cdot \mathbf{E}\left[x_{g,t}\right] \\ &= \sum_{g \in G} u_{s,g} \cdot \mathbf{E}\left[x_{g,t}\right] \end{aligned} \tag{3}$$

where $u_{s,g} \triangleq r_g \cdot a_{s,g}$. Note that for any given grid $g \in G$, we should always have $\sum_{s \in S} a_{s,g} = 1$ and hence

$$\sum_{s \in S} u_{s,g} = r_g \cdot \sum_{s \in S} a_{s,g} = r_g$$

For any given $g \in G$, we assume that the scaling factor r_g is independent of time t. This assumption is reasonable because the radio footprint in a certain area is relatively stable in the course of days to weeks, so does the locatability of UEs and the corresponding sampling ratio in our model. With a similar justification, we also assume that $\{a_{s,g}\}$ are independent of time t, and so are the factors $\{u_{s,g}\}$ (since $u_{s,g}$ is determined by r_g and $a_{s,g}$). We can then estimate the above scaling factors $\{r_g\}$ through a ridge regression formulation as follows. Let $\hat{u}_{s,g}$ be the Least Squares estimator for $u_{s,g}$ in Equation (3), i.e., $\hat{u}_{s,g}$ is the solution for the following least square problem

$$\underset{\{\hat{u}_{s,g} \mid s \in S,\, g \in G\}}{\operatorname{argmin}} \quad \sum_{s \in S} \sum_{t \in T} \left(y_{s,t} - \sum_{g \in G} \hat{u}_{s,g} \cdot x_{g,t} \right)^2 \tag{4}$$

The scaling factors $\{r_g\}$ can then be estimated as

$$\hat{r}_g \triangleq \sum_{s \in S} \hat{u}_{s,g} \qquad g \in G \tag{5}$$

3.2.2 Definition of $\hat{x}_{g,t_O}$. The definition of estimator $\hat{x}_{g,t_O}$ is much more straightforward – it is simply defined as the "sample mean" of the the corresponding historical data in $\mathcal{D}_{grid}$. More specifically, as shown in Figure 5, both datasets, namely $\mathcal{D}_{sector}$ and $\mathcal{D}_{grid}$, exhibit strong daily and weekly seasonal patterns. Thus for $g \in G$, time series $\{x_{g,t}\}$ can be partitioned into $7 \times 24 = 168$ weakly stationary processes[2] as follows. For time bins t and t', $x_{g,t}$ and $x_{g,t'}$ belong to the same weakly stationary process if $(t - t') \mod (7 \times 24) = 0$, written as $t \sim t'$. We can then define a "congruence class" of t as:

$$\mathcal{T}(t) \triangleq \{t' \mid t' \in T \text{ and } t' \sim t\}$$

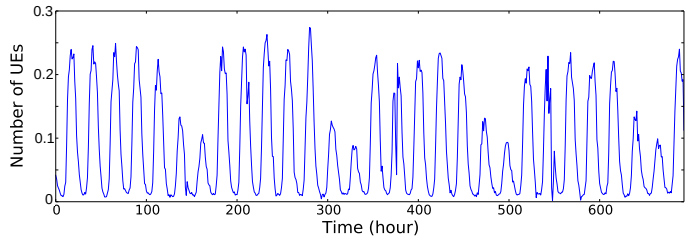
and have $\mathbf{E}[x_{g,t'}] = \mathbf{E}[x_{g,t}]$ (due to the weak stationarity assumption) if $t' \in \mathcal{T}(t)$. We can then estimate the value of x_{g,t_O} by the mean of the historical data within the same congruence class of t_O, i.e.,

$$\hat{x}_{g,t_O} \triangleq \frac{1}{|\mathcal{T}(t_O)|} \sum_{t \in \mathcal{T}(t_O)} x_{g,t} \tag{6}$$

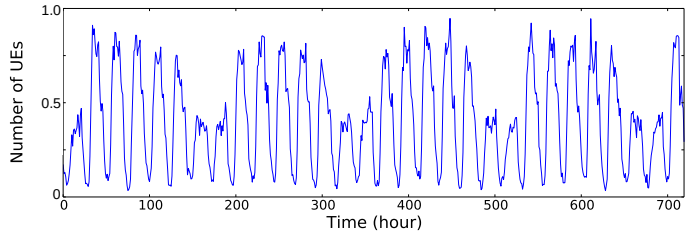
3.3 Performance Analysis

In this section, we analyze the goodness of the service impact estimator $\hat{H}_O$ defined above in terms of Mean Square Error (MSE). Since the MSE of an estimator is equal to the square of its bias plus its variance, it is desirable for the estimator to be unbiased (i.e., have 0 bias). In this section, we prove that the service impact estimator $\hat{H}_O$ defined in Section 3.2 is indeed unbiased, and arguably also has a relatively small variance, so its MSE is small. The proof of unbiasedness is given by Theorem 1. To prove this theorem, we first present the following two lemmas, which establish the unbiasedness of estimators $\hat{u}_{s,g}$ and $\hat{r}_g$. The proofs of Lemma 1, Lemma 2 and Theorem 1 are quite straightforward and can be found in Appendix A.

[2]A stochastic process $\{x_t\}$ is said to be weakly stationary if for $\forall\, t, \tau \in \mathbb{R}$ we have $\mathbf{E}[x_t] = \mathbf{E}[x_{t+\tau}]$ and $\mathbf{Cov}(x_t, x_{t+\tau}) = C_x(\tau)$ is only a function of τ (i.e., not of t).



(a) Observed number of UEs (normalized) in a grid



(b) Observed service load (normalized) on a sector

Fig. 5. Service load measured by number of UEs shows strong daily and weekly seasonal patterns.

Lemma 1. *$\hat{u}_{s,g}$ is an unbiased estimator of $u_{s,g}$ given the data set $\mathcal{D}_{grid}$, i.e.,*

$$\mathbf{E}\left[\hat{u}_{s,g} \,\middle|\, \mathcal{D}_{grid}\right] = u_{s,g}$$

Lemma 2. *$\hat{r}_g$ is an unbiased estimator of r_g given the data set $\mathcal{D}_{grid}$, i.e.,*

$$\mathbf{E}\left[\hat{r}_g \,\middle|\, \mathcal{D}_{grid}\right] = r_g$$

Theorem 1. *$\hat{H}_O$ is an unbiased estimator of H_O, i.e.,*

$$\mathbf{E}[\hat{H}_O] = \mathbf{E}[H_O]$$

However, $\hat{H}_O$ is not guaranteed to be a best linear unbiased estimator (BLUE) of H_O if we solve the ridge regression model in (4) directly using the the ordinary least squares (OLS) method [27], i.e., it's not guaranteed to give the lowest variance of the estimate, as compared to other unbiased linear estimators of H_O. This is partly because random variables $\{y_{s,t}\}_{t\in T}$ (i.e., the workloads of sector s at different time t) are not guaranteed to have the same variance at different time t, and hence the ridge regression model in (4) is not guaranteed to result in a BLUE estimator for $\{u_{s,g}\}$ when solved by the OLS method [8]. Generalized least squares (GLS) estimator [2] is not feasible here either since the covariance matrix for random variables $\{y_{s,t}\}_{t\in T}$ is generally unknown in our scenario. Fortunately we can estimate the covariance matrix of $\{y_{s,t}\}_{t\in T}$ using the residuals of the OLS estimator and then generate the GLS estimator for $\{\hat{u}_{s,g}\}$ using this estimated covariance matrix. This standard method is called feasible generalized least squares (FGLS) [10], which is empirically more efficient (i.e., has a smaller variance) than OLS in such cases. The efficiency of this FGLS estimator is verified in our simulation experiments in Section 4, which demonstrate that this FGLS estimator achieves a fairly low empirical estimate error.

3.4 Implementation Issues

In this section, we discuss two implementation issues of TOIP.

3.4.1 A low complexity solution of the ridge regression model. The least square problem in Equation (4) involves solving for $|S| \times |G|$ variables. Since $|S|$ and $|G|$ for the entire continental US is gigantic ($|S| \approx 10^5$, $|G| \approx 10^9$), solving this problem would be very time- and space-consuming. Fortunately, it can be decomposed into $|S|$ small problems as follows, which are much easier and faster to solve.

First, since we have $a_{s,g} = 0$ if $c_{s,g} = 0$, Equation (3) can be simplified as

$$\mathbf{E}[y_{s,t}] = \sum_{g:\, c_{s,g}=1} u_{s,g} \cdot \mathbf{E}\left[x_{g,t}\right]$$

As the radio coverage of a single sector is limited, number of grids involved in the above equation can be upper bounded by a constant. As a consequence, the minimization problem in Equation (4) can be split into $|S|$ disjoint subproblems as follows.

$$\underset{\{\hat{u}_{s,g} \mid g \in G\}}{\operatorname{argmin}} \quad \sum_{t \in T} \left(y_{s,t} - \sum_{g:\, c_{s,g}=1} \hat{u}_{s,g} \cdot x_{g,t} \right)^2 \qquad (s \in S)$$

Each of these subproblems involves only $O(1)$ variables and can be solved separately. Such simplification can also mitigate the over-fitting problem since less variables are involved in each of the least square problems.

3.4.2 Accommodating tower capacity constraints. Another way to improve the accuracy of the above estimator is to add tower capacities as constraints when calculating the customer impacts as it is possible that not all UEs from the failed towers can be accommodated by the adjacent towers. To take into consideration the capacity constraints, we need to know how UEs are relocated to alternate towers. This knowledge however is generally very hard to obtain in practice. A feasible, and arguably “common-sensical" alternative approach is to assume that UEs from the failed towers first try to connect to their nearest alternate towers, and continue on to try the second nearest if rejected by the first one, and so on. This assumption however may not alway be true in practice according to our observations: the closest tower is not necessarily the one that delivers the strongest signals to the UEs and thus may not be their first choice when they are trying to reconnect. Factoring in the capacity constraints using such a “common-sensical" approach may not necessarily benefit the performance of our estimator and may even do some harms in certain scenarios.

As a consequence, we did not incorporate the above “common-sensical" approach in the initial formulation of TOIP, but instead propose the following possible solution in practice. When deploying TOIP in the real system, we associate an empirical workload threshold to each of the towers in the network and estimate the workload increase in the alternate towers using the aforementioned “nearest tower” approach; although this method is not accurate enough for a “quantitative” analysis, we can use it for a “qualitative” guidance for issuing the following warning messages that network planners need to look further into: A warning message is generated whenever the (thus) estimated workload of any alternate tower exceeds the corresponding threshold, which indicates there is a potential risk in service quality degradation.

4 EVALUATION

In this section, we evaluate the accuracy of the cellular service impact prediction made by TOIP in terms of number of UEs that lost cellular service using both synthetic data and real outage data from a large operational cellular network.

4.1 Simulation using Synthetic Data

In this section, we first evaluate TOIP using synthetic data. These synthetic data, generated using a simple model shown below, may not capture all the characteristics in the real network. But it allows us to control and isolate variables in different aspects, e.g., the outage area G_O or the tower-grid association matrix C, and to investigate how these variables affect the accuracy of TOIP. Also, with synthetic data, we are able to evaluate the accuracy of the intermediate steps of our estimator, e.g., the accuracy of the estimated scaling factors $\hat{r}_g$, and have a better understanding of the detailed behavior of the estimator.

More specifically, to generate the synthetic data, we consider a $10\,km \times 10\,km$ geographical area with 10 randomly located cellular towers. This area is divided into 100×100 grids, each representing a $100\,m \times 100\,m$ geographical bin. Per the best practice in cellular networking, the cellular service providers usually deploy more towers in the areas with larger UE densities (e.g., an urban area), and in this case to limit the interferences between different towers, the antennas are tuned so that each grid is only covered by a small number of towers. In accordance with this, we enforce the following two constraints in our simulation model: the UE density in each grid is positively correlated with the tower density in its nearby area; and the coverage of a tower is negatively correlated with the number of towers in its nearby area. More specifically, in the simulation, we assume that the antennas are tuned so that each grid is covered only by 4 towers nearest to it. The actual number of UEs within a grid g at time t, denoted as $z_{g,t}$, is modeled as a Gaussian random variable with its mean and standard deviation set proportional to the tower density in the nearby areas of this grid. We assume that the radio transmission in this area follows the log-distance path loss model and the probability that a UE connecting to a certain tower (with this UE in its coverage) is proportional to the power strength it receives from the tower. The number of UEs connecting to each tower is then calculated according to the above tower-grid associations. Furthermore, as described in Section 2.3 and 3.1, due to the non-uniform locatability of UEs, the number of UEs in each grid observed by our measurement system can be modeled as a random sampling process. As UEs in grids under coverage of multiple towers are more "locatable" as they are more likely to exchange signaling messages with RNC to achieve hand-over among multiple towers, we assume that the sampling rate (proportion of located UEs) within a certain grid is proportional to the density of towers in its nearby area. Figure 6 shows an example of the network topology generated in this way. Standard kernel density estimator (KDE) with Gaussian kernels is invoked here to estimate the tower densities at different locations. Average sampling rate of grid g, denoted as p_g, is then calculated by multiplying the tower densities with a scaling factor such that the maximum sampling rate over the whole area is normalized to 1. Observed number of UEs in grid g at time t, denoted as $x_{g,t}$, is then generated according to the corresponding sampling rate, i.e, $x_{g,t} = Binomial(z_{g,t}, p_g)$.

Our first set of experiments evaluates the accuracy of the per-grid scaling factor estimation made by TOIP. For the randomly generated network topology and load described above, we calculate the estimated scaling factor $\hat{r}_g$ for each grid using our TOIP algorithm[3] (trained with number-of-UE time series of length 100,000) and compare it with the ground truth $r_g = 1/p_g$. The empirical cumulative distribution function (CDF) of the ratio $\hat{r}_g/r_g$ is shown in Figure 7. As shown in the figure, this ratio is concentrated around 1 throughout the grids, demonstrating that the estimated scaling factors $\hat{r}_g$ are indeed unbiased and empirically also have a small variance.

To further evaluate the service impact estimation accuracy for potential tower outages, we randomly select a block of grids as the outage area G_O and estimate the total number of UEs in this area using TOIP. We randomly generate 10 different network topologies, repeat this process 10,000

[3]The the ridge regression model in (4) is solved using the FGLS method mentioned in Section 3.3.

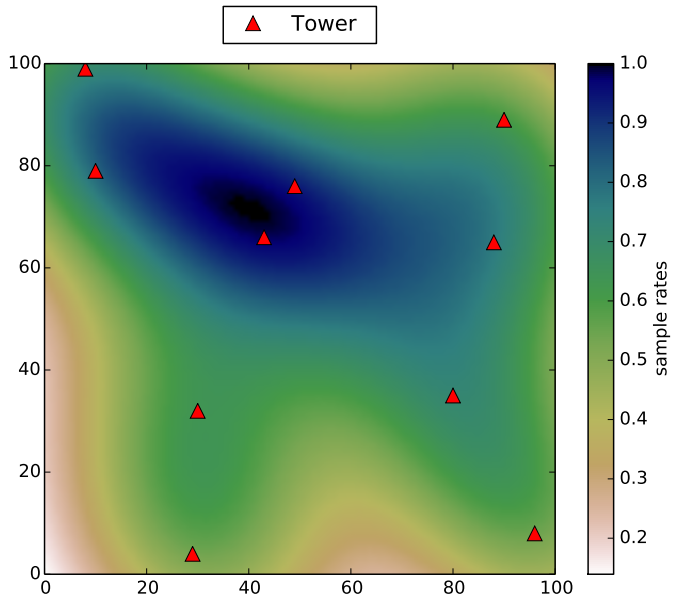


Fig. 6. An example for the randomly generated network topology.

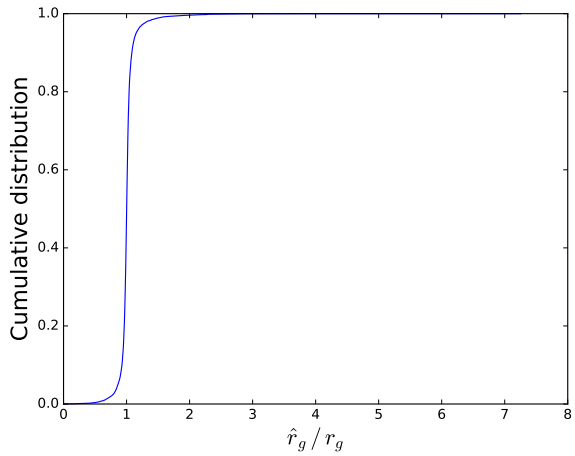


Fig. 7. Empirical cumulative distribution function (CDF) of the ratio $\hat{r}_g/r_g$.

times for each of such topologies and report the overall average estimate error ratio[4] as the result. We use a strawman algorithm here to compare with our TOIP, which simply uses the inverse of

[4]The average error ratio (ER) metric used in our evaluation is defined as the square root of the normalized mean square error, i.e.,

$$ER = \sqrt{\frac{\sum_{i=1}^{N}\left(H_O^{(i)} - \hat{H}_O^{(i)}\right)^2}{\sum_{i=1}^{N}\left(H_O^{(i)}\right)^2}}$$

the average sampling rate over the whole area to recover the number of UEs instead of the per-grid scaling factors $\hat{r}_g$. We vary the length of the number-of-UEs time series used to train the estimator and the sizes of the outage areas $|G_O|$ in the experiments. The results are shown in Figures 8, 9 and 10. We can observe that with sufficient amount of training data, the estimation error ratio of TOIP is only around 3% $\sim$ 5% while the strawman algorithm is around 30%. The estimation accuracy of TOIP is significantly better than the strawman algorithm.

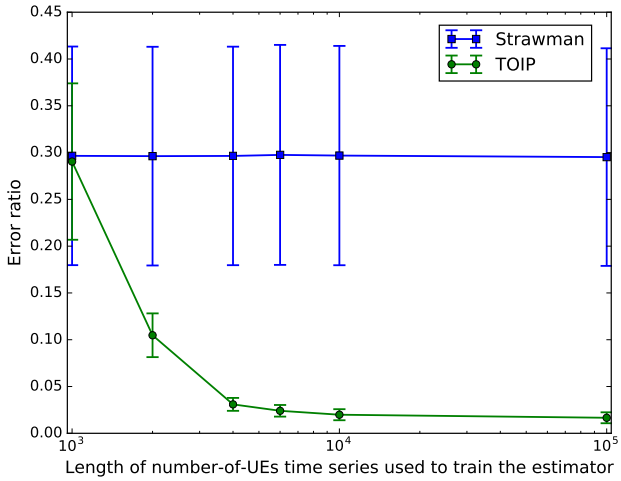


Fig. 8. Error ratios of the service impact estimations made by the strawman algorithm and TOIP when $|G_O| = 1,000$. Their empirical standard deviations are shown as "error bars" in the figure.

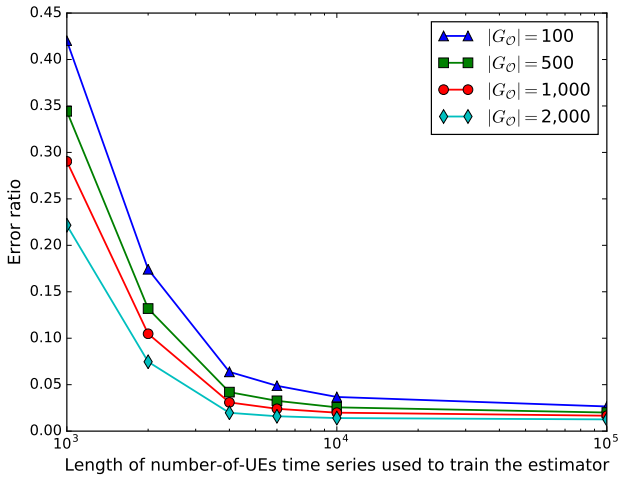


Fig. 9. Error ratio of service impact estimation made by TOIP when varying the sizes of the outage areas $|G_O|$.

where $H_O^{(i)}$, $i = 1, \ldots, N$, is the actual outage customer impacts (i.e., the ground truth) in the i-th experiment and $\hat{H}_O^{(i)}$, $i = 1, \ldots, N$, is its estimate.

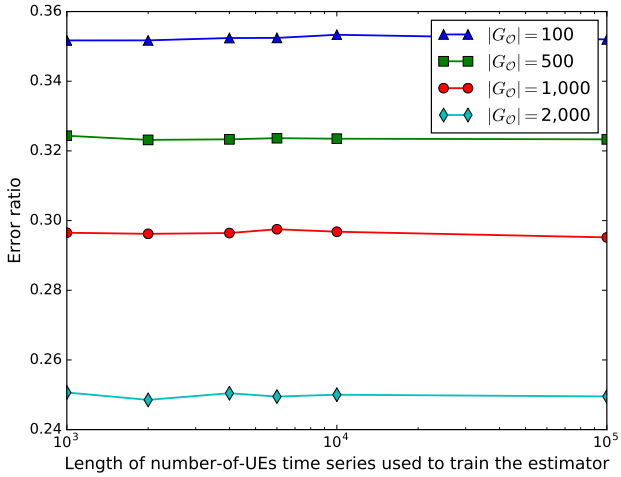


Fig. 10. Error ratio of service impact estimation made by the strawman algorithm when varying the sizes of the outage areas $|G_{\mathcal{O}}|$.

The third set of experiments investigate the robustness of TOIP on incomplete tower-gird association data. As mentioned in Section 3.1, dataset $\mathcal{D}_{assoc}$ is inferred from the UE geo-locating measurement data, but due to the aforementioned non-uniform locatability of UEs, we may fail to infer a small percentage of tower-gird associations when there are insufficient number of locatable UEs in certain areas. To evaluate the robustness of TOIP to such data incompleteness, we randomly drop 50% of the tower-grid association information in the synthetic data and use the remaining part to train the TOIP estimator. The result is shown in Figure 11. When part of the tower-grid association information is missing, TOIP will automatically cast more weights on the remaining tower-grid associations. This may lead to inaccurate estimation of the scaling factors $\hat{r}_g$. More specifically, it may tend to underestimate scaling factors of the grids with more missing data while overestimate the others. However, such distortion could cancel out with each other when we consider a block of grids. As shown in Figure 11, the estimation error ratio is increase by only $3\% \sim 5\%$ even if 50% of the tower-grid association information is missing[5]. In other words, TOIP is quite robust against the data incompleteness.

4.2 Systematic Evaluation using Real Operational Data

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

[5]Note that the left most point in Figure 11 is even smaller than that in Figure 9. It's because less "features" are involved when the tower-grid association information is incomplete, which mitigates the over-fitting problem when there's *insufficient* amount of training data.

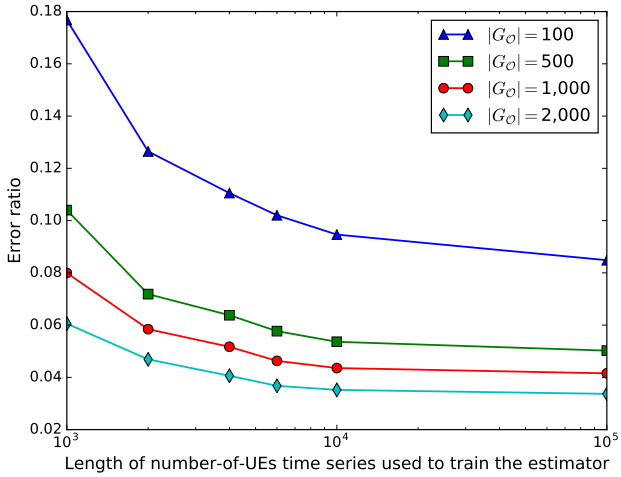


Fig. 11. Error ratio of service impact estimation made by TOIP using incomplete tower-gird association data.

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	$0 \sim 0.002$
Level 2	$0.002 \sim 0.02$
Level 3	$0.02 \sim 0.1$
Level 4	$0.1 \sim 0.2$
Level 5	$0.2 \sim 0.6$
Level 6	> 0.6

As shown in Figure 12, 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.

5 USE CASE: TOWER NTE REHOMING

As TOIP accurately estimates the cellular service availability given a potential backhaul network failure scenario, it becomes a powerful tool for cellular backhaul network planners to re-optimize their backhaul design, from time to time, so that the impact of any actual backhaul network failure, such as fiber cut or Ethernet card failure, on cellular service is minimized. In this section, we demonstrate the efficacy of TOIP in backhaul network re-optimization via simulation studies on a large real-world cellular network topology. More specifically, we show that the use of TOIP can lead

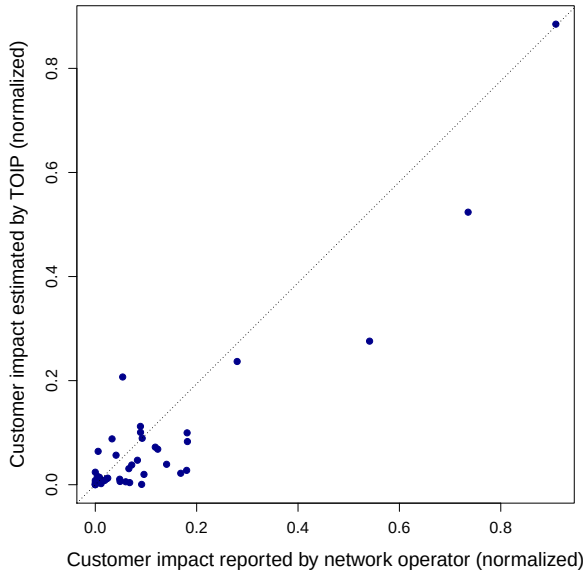


Fig. 12. Customer impact estimation made by TOIP

to design changes in the backhaul network that would result in significantly improved resiliency to failures.

5.1 Rehoming Optimization Problem

Since the cellular backhaul network was designed in phases and cellular towers were built gradually, cellular tower NTEs may not always be optimally homed to backhaul network access devices at all times. Hence cellular backhaul network planners are required to analyze service availability and periodically re-home some of the tower NTEs to backhaul network access devices that provide better cellular accesses to them.

As mentioned in section 2.1, each tower backhaul connection from its NTE to MTSO consists of two segments: one segment is from NTE to backhaul network access device and the other is from access device to MTSO. The backhaul network is usually a mesh topology with full failure restoration capability whereas the first segment from an NTE to its backhaul network access device is usually single homed over access fibers. At the early stage of backhaul network deployment, when the number of access devices is relatively small, a tower NTE may have to be initially homed to an access site that is relatively far away, resulting in a relatively long unprotected backhaul connection segment. Later on, however, with the expansion of the backhaul network, this tower NTE could become much closer to another access site. In this case, rehoming this tower NTE from the original (faraway) access site to the new (nearby) access site can shorten the unprotected backhaul connection segment and improve tower reliability. There is however another factor to consider in this rehoming: As deployments of access fibers and cellular towers are always limited, multiple tower NTE backhaul connections often have to share a common fiber segment. Should this shared common fiber segment be cut, all towers with backhaul connections routed over it would fail together, which could impact a large number of customers. Hence a rehoming operation needs to take into consideration the service impact of potential backhaul segment failures.

5.2 The Prior Rehoming Solution

Due to budget constraints, only a limited percentage of tower NTEs can be re-homed during a certain time period and network planners need to prioritize the rehoming operations for different towers. Before TOIP, the state-of-the-art solution for rehoming tower NTEs uses the number of affected cellular towers as the measure to assess the service impact of a backhaul network failure. Under this measure, the service impact of a backhaul failure was simply defined by those towers routed through the failed network segment. Hence, given the high cost associated with any reliability enhancement to backhaul network topology, the planners typically focus on identifying backhaul network segments that are shared by a large number of cellular towers; to mitigate the potential impact of a backhaul failure, they either design additional protection (e.g, adding a secondary circuit) to these network segments or re-home some of the connected cellular towers for them to gain network access via alternative paths.

The assumption made by this solution that every disconnected tower (due to failure) has the same impact on users is a crude and questionable one in that it ignores a large amount of useful information, such as the service redundancy of the radio network. For one thing, the number of affected cellular towers often has a low correlation with the cellular service availability perceived by users. In other words, depending on the spatial distribution of cellular towers and UEs and the radio environments it induces, it is very possible that losing a network segment (e.g., a fiber cut) serving more cellular towers may be less disruptive than the ones serving fewer cellular towers. For example, some cellular towers are located in dense urban business areas that have high loads while some other towers are located in suburban areas with low loads; the antennas on some towers are tilted up to cover a large geographical area whereas those on some others are tilted down purposely to avoid interference with neighboring towers.

TOIP takes a fundamentally different approach to analyzing service availability and offers a more optimized solution to rehoming cellular tower NTEs. In the following, we provide a vis-a-vis comparison between our TOIP-based solution and the prior solution, which we refer to as *network impact solution* since it considers only the network-tower relationship.

Network Impact Solution (Previous solution):

(1) Find the network segment failure with the largest network impact, i.e., the network segment failure that will bring down the largest number of towers in the network.
(2) Among all the towers affected by this network segment failure, find the one with the largest risk score. Here the risk score of a tower is the total network impact of all the network segment failures that would affect with this tower.
(3) Rehome this tower.
(4) Repeat Steps (1) - (3).

TOIP-based Customer Impact Solution:

(1) Find the network segment failure with the largest customer impact (estimated by TOIP).
(2) Among all the towers affected by this network segment failure, find the one that, if it's rehomed, will lead to the largest reduction in the customer impact (estimated using TOIP).
(3) Rehome this tower.
(4) Repeat Steps (1) - (3).

5.3 Experiment Results

To demonstrate the efficacy of using TOIP in improving the resilience of the network, we compare the performance of these two tower rehoming approaches by simulation experiments using a real network topology. The network area we investigated is illustrated in Figure 13, which is a medium size local access and transport area (LATA) in a large operational cellular network. There are a

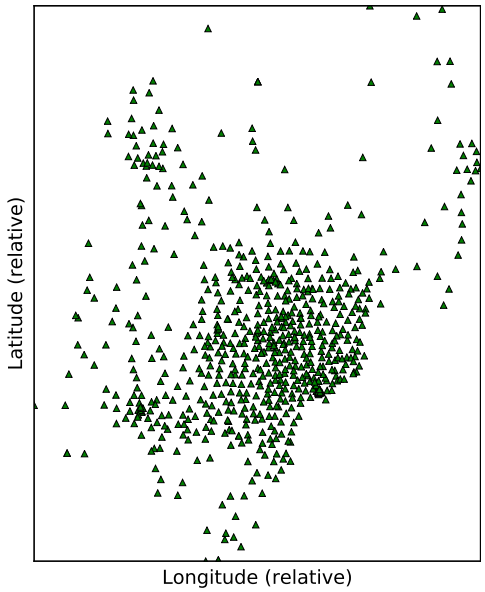
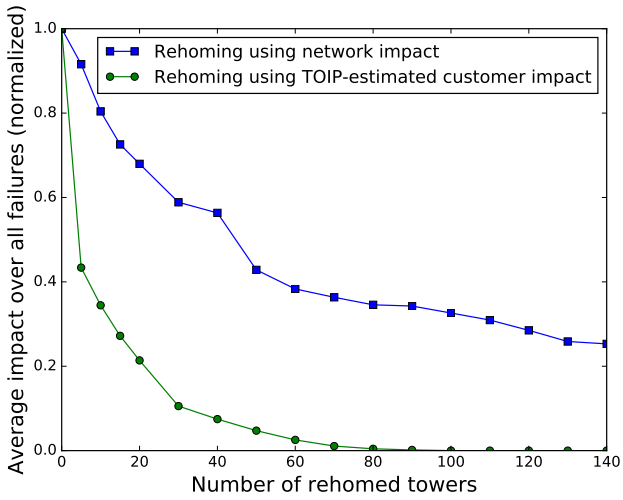


Fig. 13. A real network topology with 693 towers in a medium size local access and transport area (LATA).

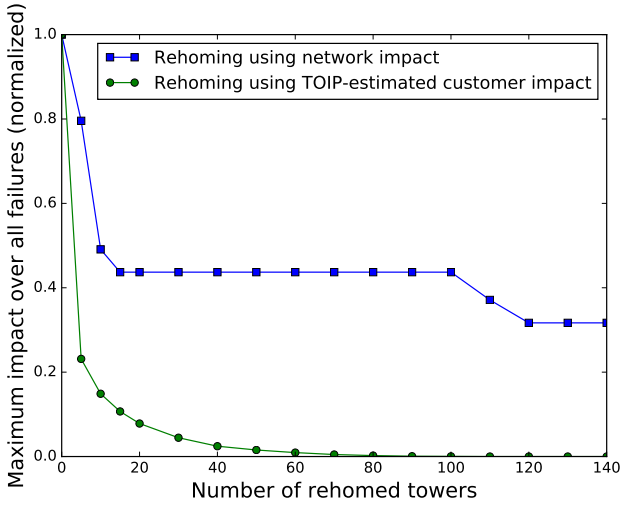
total of 693 towers and 844 potential network segment failures. This is a coastal metropolitan area. Towers are dense in the city and is sparser in the surrounding suburban and rural areas. This area is divided into about 2×10^4 grids, each representing a $1\,km \times 1\,km$ geographical bin (water area is excluded in our simulation). The number of UEs in each grid and the proportion of locatable UEs within the grid are generated using the synthetic model described in Section 4.1.

We use synthetic UE distribution data here because the actual number of UEs in each grid is generally unavailable in practice due to the aforementioned non-uniform locatability of UEs. Such data is however indispensable in calculating the *real* customer impact (not the one estimated by TOIP) of each hypothetical network segment failure, since we need to know its real customer impact after the tower-rehoming in order to compare the efficacies of these two approaches in improving the resiliency of the network. It is also inappropriate to evaluate the efficacy of the TOIP-based approach using the customer impact estimated by TOIP itself. Hence, as a compromise solution, we use the synthetic UE distribution data here to fill this gap. We emphasize this compromise solution is innocuous for the following reason. As described in Section 4.1, our statistical model of generating such synthetic UE distribution data is inferred directly from the given network topology, and clearly is independent of the statistical models used by TOIP. Hence, intuitively the use of such synthetic data should neither help nor hurt the outcome statistically of the comparison between our TOIP based approach and the prior approach.

In our simulation experiments, we rehome a given number of the towers using the above two tower-rehoming approaches respectively, and calculate the "real" customer impacts (using the aforementioned synthetic data) of a potential network segment failure after the rehoming. We do not consider the combination of multiple network segment failures in the evaluation since simultaneous multiple failures almost never happen in practice. We vary the number of towers to be rehomed, and report the average and the maximum customer impacts over all potential single network segment failures after rehoming the corresponding number of towers. The results, plotted in Figure 14, show conclusively that our TOIP approach significantly outperforms the



(a) *Average* customer impact (normalized) over all the 844 network segment failures after tower NTE rehoming.



(b) *Maximum* customer impact (normalized) over all the 844 network segment failures after tower NTE rehoming.

Fig. 14. Tower rehoming using TOIP results in a more reliable network design.

prior approach. More specifically, both the average (Figure 14a) and the maximum (Figure 14b) customer impacts of the potential single network segment failures for the TOIP approach drop much faster than those for the prior approach. Furthermore, perhaps a bit surprisingly, the TOIP tower-rehoming approach is able to reduce the customer impact of each single network segment failure to a negligible value after rehoming only about 10% of the 693 towers. This implies that the affected towers of each single failure are sufficiently "scattered" in the network after rehoming, and will not have a significant user impact when they are brought down.

We have also performed additional experiments using two other real cellular topology datasets and the experimental results, shown in Appendix B, lead to similar conclusions. To summarize, with the help of TOIP, we are able to identify the tower NTEs that need rehoming most urgently and to put capital investments in the segments that need the protection the most. This capability clearly will result in an improved resiliency to failures and a more reliable cellular network design. Consequently, TOIP has been adopted by the cellular backhaul network planner team of a large cellular service provider in North America as a primary tool to re-optimize its backhaul network since early 2016.

6 RELATED WORK

To the best of our knowledge, no prior work has been devoted to predicting the customer impact of hypothetic cellular tower outages. We have mainly been inspired by prior works in three areas: what-if analysis, network resilience and cellular network modeling.

What-if analysis. What-if analysis have recently received much attention in many fields. Various methods are explored to design such kind of systems. In [21], a workload-based what-if analysis system was designed and implemented for cloud computing applications. Network of queues was utilized in this paper to analytically model the behavior of large distributed cloud applications. Thereska et. al. presented a self-predicting storage system in [24], which continuously monitors the system status and uses the activity records to answer the what-if questions about the impact of a decision on the performance of the system. WISE [23] is another what-if analysis system designed for content distribution network (CDN). It uses statistical learning techniques to predict the service response-time distribution when infrastructure deployment and configuration are changed, such as deploying a new data center or changing the mapping of clients to servers. Though these designs all enable the administrators to ask a variety of "what-if" questions about the corresponding systems, none of them considered the "what-if" scenarios in a wireless cellular network. Our TOIP scheme, as described above, is designed to fill this gap.

Network resilience. Researchers have made great efforts towards improving the resilience of cellular networks. For example, MERCURY [16], Litmus [15] and PRISM [14] are proposed to detect and assess the impact of planned or unplanned network changes (e.g., network upgrade or maintenance) in cellular networks. Several other works [1, 3, 4, 6, 9, 13, 20, 26, 28, 30], such as Cell Outage Compensation (COC) in Self-Organizing Networks (SON) and the recently proposed Magus scheme [29], focus on improving the service performance during network upgrades or outages. Network resilience under extreme conditions has also attracted considerable attention in recent years and many efforts are devoted to measuring and minimizing the service disruptions due to natural or manmade disasters (e.g., Hurricane Sandy and 9/11 Attack) [12, 17–19, 22, 25, 31]. Our problem scope is different from that of these prior works, which are focused on assessing or improving the service performance *during or after* network changes. TOIP on the other hand is designed to predict the service impact on end-user's equipments for *hypothetic* cellular-tower outage scenarios. It allows the service provider to invest in the enhancements of backhaul transport network efficiently so that the improvement to cellular service reliability is maximized.

Modeling Cellular Networks. Assessing the performance of cellular networks and the end-users' behaviors is challenging partially because of overlapping external factors such as terrain, buildings, and foliage. These external factors could be very different in different geographical regions. As it is extremely hard to estimate such details by a simple model, people make various statistical assumptions in their model when studying the performance of cellular networks [7, 11, 13, 26, 28, 30]. Inspired by Magus [29], which divides the coverage area into $100m \times 100m$ grids and calculates each grid's SINR and throughput rate independently, TOIP uses a measurement-based model (rather than the idealized analytical model) to avoid making simple assumptions for these external factors.

TOIP v.s. Magus. The problem scopes of these two schemes are mostly orthogonal to each other. At a high level, Magus is more network-centric while TOIP is more customer-centric. More specifically, Magus is concerned primarily with the change in the coverage of towers and the degradation of service quality (e.g., downlink rate) during network upgrades, but it does not take the number of UEs that are affected by this change into account. TOIP, on the other hand, provides an estimation of the number of UEs within the affected area, and thus can be used to complement Magus to further improve its efficacy. Furthermore, Magus is designed mainly for the real-time mitigation of a short-term service disruption caused by a planned network upgrade; it does so by proactively tuning the transmission power and the tilts of nearby towers to minimize the service impact of the tower outages caused by the upgrade. TOIP, on the other hand, is designed mainly to assist long-term network planning and does not account for such proactive adjustments when estimating the outage impact.

7 CONCLUSION

In this paper, we study the service impact analysis/prediction problem for cellular service providers in the application of transport network design. We rely on two classes of measurement data collected from a major cellular provider network in north America. We invent and prototype a novel system called TOIP to answer what-if questions regarding end-users' experience given hypothetical cell tower outage scenarios. TOIP first models the radio coverage for each spatial grid (100m $\times$ 100m area) based on the signal quality readings from individual UEs and determines the set of spatial grids without radio coverage given a particular outage scenario. It then estimates the number of UEs in the affected spatial grids by aggregating the UE level records and calibrating it with tower level measurements. We evaluate our system with both synthetic data and historical real tower outages in a large operational cellular network. The result demonstrates close matches between the predicted end-users' experience and the observed service impact in retrospect.

A PROOF OF THE UNBIASEDNESS

In this section, we give the detailed proof for the unbiasedness of the customer impact estimator $\hat{H}_O$ defined in Section 3. We first describe a more rigorous mathematical modeling of the relationship between the grid-level measurement $\mathcal{D}_{grid}$ and the sector-level measurement $\mathcal{D}_{sector}$ in Section A.1, and then give the proof of Lemma 1, Lemma 2 and Theorem 1 in Sections A.2, A.3 and A.4, respectively. Readers may refer to Table 2 in Section A.5 for a summary of the major notations used in the proof.

A.1 Modeling

For any t, random variables $z_{g,t}$, $x_{g,t}$ and $y_{s,t}$ are correlated because they are measurements of the same quantities from different vantage points. Specifically, for any $s \in S$ and $t \in T$, we have

$$\begin{aligned} y_{s,t} &= \sum_{g \in G} a_{s,g,t} \cdot z_{g,t} \\ &= \sum_{g \in G} a_{s,g,t} \cdot r_{g,t} \cdot x_{g,t} \end{aligned} \tag{7}$$

where $0 \leq a_{s,g,t} \leq 1$ is the proportion of UEs that are located in grid g and connected to radio sector s at time t. For any $t \in T$ and $g \in G$, we have

$$\sum_{s \in S} a_{s,g,t} = 1 \tag{8}$$

We assume that, given any fixed t, random variable $r_{g,t}$ is independent of each of the random variables $x_{g,t}$ and $z_{g,t}$ (but not independent of both because $r_{g,t} \triangleq 1/p_{g,t}$ and $p_{g,t} \triangleq x_{g,t}/z_{g,t}$). This assumption is reasonable since when the provider network is not overloaded, the locatability of UEs is independent of the number of UEs in the area. As a consequence, the sampling process that produces measurements $x_{g,t}$ in our model – and hence its actual average sampling ratio $p_{g,t}$ during time bin t – is independent of the set of UEs being sampled (with a population of $z_{g,t}$) and also independent of the set of UEs actually sampled (with a population of $x_{g,t}$). We also assume that given any fixed t, random variable $a_{s,g,t}$ is independent of each of the random variables $x_{g,t}$ and $z_{g,t}$, and $r_{g,t}$. This assumption can be similarly justified.

For any given $s \in S$ and $g \in G$, we assume that time series $\{r_{g,t}\}$ is a weakly stationary process. This assumption is reasonable because the radio footprint in a certain area is relatively stable in the course of days to weeks, so does the locatability of UEs and the corresponding sampling ratio in our model. With a similar justification, we also assume that $\{a_{s,g,t}\}$ is a weakly stationary process. Hence, there exist constants r_g and $a_{s,g}$ such that for any time t we always have $\mathbf{E}[r_{g,t}] = r_g$ and $\mathbf{E}[a_{s,g,t}] = a_{s,g}$ (Note that there is no guarantee that $\mathbf{E}[r_{g,t}] = 1/\mathbf{E}[p_{g,t}]$). Define $\xi^{(r)}_{g,t}$ and $\xi^{(a)}_{s,g,t}$ as

$$\xi^{(r)}_{g,t} \triangleq r_{g,t} - r_g \tag{9}$$

$$\xi^{(a)}_{s,g,t} \triangleq a_{s,g,t} - a_{s,g} \tag{10}$$

i.e., $\xi^{(r)}_{g,t}$ and $\xi^{(a)}_{s,g,t}$ are the noise components of $r_{g,t}$ and $a_{s,g,t}$ respectively, and we have $\mathbf{E}\left[\xi^{(r)}_{g,t}\right] = \mathbf{E}\left[\xi^{(a)}_{s,g,t}\right] = 0$.

Combining Equations (7), (9) and (10), we have

$$\begin{aligned} y_{s,t} &= \sum_{g\in G} a_{s,g,t} \cdot r_{g,t} \cdot x_{g,t} \\ &= \sum_{g\in G} \left(a_{s,g} + \xi^{(a)}_{s,g,t}\right) \cdot \left(r_g + \xi^{(r)}_{g,t}\right) \cdot x_{g,t} \\ &= \sum_{g\in G} a_{s,g} \cdot r_g \cdot x_{g,t} + \sum_{g\in G} a_{s,g} \cdot \xi^{(r)}_{g,t} \cdot x_{g,t} \\ &\quad + \sum_{g\in G} \xi^{(a)}_{s,g,t} \cdot r_g \cdot x_{g,t} + \sum_{g\in G} \xi^{(a)}_{s,g,t} \cdot \xi^{(r)}_{g,t} \cdot x_{g,t} \\ &= \sum_{g\in G} u_{s,g} \cdot x_{g,t} + \xi^{(y)}_{s,t} \end{aligned} \tag{11}$$

where

$$\begin{aligned} u_{s,g} &= a_{s,g} \cdot r_g \\ \xi^{(y)}_{s,t} &= \sum_{g\in G} a_{s,g} \cdot \xi^{(r)}_{g,t} \cdot x_{g,t} + \sum_{g\in G} \xi^{(a)}_{s,g,t} \cdot r_g \cdot x_{g,t} \\ &\quad + \sum_{g\in G} \xi^{(a)}_{s,g,t} \cdot \xi^{(r)}_{g,t} \cdot x_{g,t} \end{aligned} \tag{12}$$

A.2 Proof of Lemma 1

Proof. It's sufficient to prove that $\mathbf{E}\left[\xi^{(y)}_{s,t} \,\middle|\, \mathcal{D}_{grid}\right] = 0$. For any $s \in S$, $g \in G$ and $t \in T$, we have

$$\mathbf{E}\left[a_{s,g} \cdot \xi^{(r)}_{g,t} \cdot x_{g,t} \,\middle|\, \mathcal{D}_{grid}\right]$$

$$
\begin{aligned}
&= \mathbf{E}\left[a_{s,g} \cdot \xi_{g,t}^{(r)} \cdot x_{g,t} \,\middle|\, x_{g,t}\right] \\
&= \mathbf{E}\left[a_{s,g} \cdot x_{g,t} \,\middle|\, x_{g,t}\right] \cdot \mathbf{E}\left[\xi_{g,t}^{(r)} \,\middle|\, x_{g,t}\right] \\
&= \mathbf{E}\left[a_{s,g} \cdot x_{g,t} \,\middle|\, x_{g,t}\right] \cdot \mathbf{E}\left[\xi_{g,t}^{(r)}\right] \\
&= \mathbf{E}\left[a_{s,g} \cdot x_{g,t} \,\middle|\, x_{g,t}\right] \cdot 0 \\
&= 0
\end{aligned}
$$

The third equality holds since $r_{g,t}$ is independent of $x_{g,t}$, so does $\xi_{g,t}^{(r)}$. Similarly, we can prove that

$$
\begin{aligned}
\mathbf{E}\left[\xi_{s,g,t}^{(a)} \cdot r_g \cdot x_{g,t} \,\middle|\, \mathcal{D}_{grid}\right] &= 0 \\
\mathbf{E}\left[\xi_{s,g,t}^{(a)} \cdot \xi_{g,t}^{(r)} \cdot x_{g,t} \,\middle|\, \mathcal{D}_{grid}\right] &= 0
\end{aligned}
$$

Thus

$$
\begin{aligned}
\mathbf{E}\left[\xi_{s,t}^{(y)} \,\middle|\, \mathcal{D}_{grid}\right] = &\sum_{g \in G} \mathbf{E}\left[a_{s,g} \cdot \xi_{g,t}^{(r)} \cdot x_{g,t} \,\middle|\, \mathcal{D}_{grid}\right] \\
&+ \sum_{g \in G} \mathbf{E}\left[\xi_{s,g,t}^{(a)} \cdot r_g \cdot x_{g,t} \,\middle|\, \mathcal{D}_{grid}\right] \\
&+ \sum_{g \in G} \mathbf{E}\left[\xi_{s,g,t}^{(a)} \cdot \xi_{g,t}^{(r)} \cdot x_{g,t} \,\middle|\, \mathcal{D}_{grid}\right] \\
=&0
\end{aligned}
$$

□

A.3 Proof of Lemma 2

Proof. From Equation (8), we have

$$
\sum_{s \in S} a_{s,g} = \sum_{s \in S} \mathbf{E}[a_{s,g,t}] = \mathbf{E}\left[\sum_{s \in S} a_{s,g,t}\right] = 1
$$

Combining this with Equation (12), we can obtain

$$
\sum_{s \in S} u_{s,g} = r_g \sum_{s \in S} a_{s,g} = r_g
$$

Thus, by Equation (5) and Lemma 1, we have

$$
\begin{aligned}
\mathbf{E}\left[\hat{r}_g \,\middle|\, \mathcal{D}_{grid}\right] &= \mathbf{E}\left[\sum_{s \in S} \hat{u}_{s,g} \,\middle|\, \mathcal{D}_{grid}\right] \\
&= \sum_{s \in S} \mathbf{E}\left[\hat{u}_{s,g} \,\middle|\, \mathcal{D}_{grid}\right] \\
&= \sum_{s \in S} u_{s,g} \\
&= r_g
\end{aligned}
$$

□

A.4 Proof of Theorem 1

Proof. By the definition of H_O in Equation (1), we have

$$\begin{aligned}
\mathbf{E}[H_O] &= \sum_{g \in G_O} \mathbf{E}\left[z_{g,t_O}\right] \\
&= \sum_{g \in G_O} \mathbf{E}\left[r_{g,t_O} \cdot x_{g,t_O}\right] \\
&= \sum_{g \in G_O} \mathbf{E}\left[r_{g,t_O}\right] \cdot \mathbf{E}\left[x_{g,t_O}\right] \\
&= \sum_{g \in G_O} r_g \cdot \mathbf{E}\left[x_{g,t_O}\right]
\end{aligned}$$

The third equality holds since r_{g,t_O} is independent with x_{g,t_O}.

By Equation (2), Equation (6) and Lemma 2, we have

$$\begin{aligned}
\mathbf{E}[\hat{H}_O] &= \sum_{g \in G_O} \mathbf{E}\left[\hat{r}_g \cdot \hat{x}_{g,t_O}\right] \\
&= \sum_{g \in G_O} \mathbf{E}\left[\hat{r}_g \cdot \frac{1}{|\mathcal{T}(t_O)|} \sum_{t \in \mathcal{T}(t_O)} x_{g,t}\right] \\
&= \sum_{g \in G_O} \frac{1}{|\mathcal{T}(t_O)|} \sum_{t \in \mathcal{T}(t_O)} \mathbf{E}\left[\hat{r}_g \cdot x_{g,t}\right] \\
&= \sum_{g \in G_O} \frac{1}{|\mathcal{T}(t_O)|} \sum_{t \in \mathcal{T}(t_O)} \mathbf{E}\left[\mathbf{E}\left[\hat{r}_g \cdot x_{g,t} \,\middle|\, \mathcal{D}_{grid}\right]\right] \\
&= \sum_{g \in G_O} \frac{1}{|\mathcal{T}(t_O)|} \sum_{t \in \mathcal{T}(t_O)} \mathbf{E}\left[\mathbf{E}\left[\hat{r}_g \,\middle|\, \mathcal{D}_{grid}\right] \cdot x_{g,t}\right] \\
&= \sum_{g \in G_O} \frac{1}{|\mathcal{T}(t_O)|} \sum_{t \in \mathcal{T}(t_O)} \mathbf{E}\left[r_g \cdot x_{g,t}\right] \\
&= \sum_{g \in G_O} \frac{1}{|\mathcal{T}(t_O)|} \sum_{t \in \mathcal{T}(t_O)} r_g \cdot \mathbf{E}\left[x_{g,t}\right] \\
&= \sum_{g \in G_O} \frac{1}{|\mathcal{T}(t_O)|} \sum_{t \in \mathcal{T}(t_O)} r_g \cdot \mathbf{E}\left[x_{g,t_O}\right] \\
&= \sum_{g \in G_O} r_g \cdot \mathbf{E}\left[x_{g,t_O}\right] \\
&= \mathbf{E}[H_O]
\end{aligned}$$

In the fifth equality, $\mathbf{E}\left[\hat{r}_g \cdot x_{g,t} \,\middle|\, \mathcal{D}_{grid}\right] = \mathbf{E}\left[\hat{r}_g \,\middle|\, \mathcal{D}_{grid}\right] \cdot x_{g,t}$ since $x_{g,t}$ is in the dataset $\mathcal{D}_{grid}$. The eighth equality holds because $\mathbf{E}\left[x_{g,t}\right] = \mathbf{E}\left[x_{g,t_O}\right]$ for $\forall\, t \in \mathcal{T}(t_O)$. □

A.5 Table of the Notations

In this section, we provide a table (Table 2) for the major notations used in Sections A.1 - A.4.

Table 2. Important notations in Sections A.1 - A.4

$\mathcal{D}_{sector}$	Dataset of the sector-level number-of-UE measurements.
$\mathcal{D}_{grid}$	Dataset of the grid-level number-of-UE measurements.
$\mathcal{D}_{assoc}$	Dataset of the sector-grid associations.
S	The set of radio sectors in the provider network.
G	The set of geographic grids in the network area.
T	The set of time bins.
$\mathcal{T}(t)$	The "congruence class" of time bins, i.e., $\mathcal{T}(t) \triangleq \{t' \mid t' \in T \text{ and } t' \sim t\}$ where $t \sim t'$ iff $(t - t') \mod (7 \times 24) = 0$
$c_{s,g}$	A binary variable indicating the association relationship between sector $s \in S$ and grid $g \in G$. $c_{s,g} = 1$ if and only if sector $s \in S$ has radio coverage for grid $g \in G$.
$y_{s,t}$	The service load (in terms of number-of-UEs) measured on sector $s \in S$ at time $t \in T$.
$x_{g,t}$	The observed number of UEs in grid $g \in G$ at time $t \in T$.
$z_{g,t}$	The actual number of UEs in grid $g \in G$ at time $t \in T$.
O	$O = (S_O, t_O)$ represents a given outage scenario.
S_O	$S_O \subseteq S$ is the set of out-of-service sectors during outage O.
t_O	$t_O \in T$ is the time at which outage O happens.
G_O	$G_O \subseteq G$ is the set of geographical grids that loses service coverage during outage O
H_O	Service impact of outage O.
$\hat{H}_O$	Estimated service impact of outage O.
$p_{g,t}$	$p_{g,t} \triangleq x_{g,t}/z_{g,t}$ is the sampling ratio of grid g at time t
$r_{g,t}$	$r_{g,t} \triangleq 1/p_{g,t}$ is the scaling factor of grid g at time t
r_g	$r_g \triangleq \mathbf{E}[r_{g,t}]$ is the time-independent expectation of random variable $r_{g,t}$ (This definition is valid only when assuming $\{r_{g,t}\}$ is a stationary time series)
$\hat{r}_g$	An estimator for r_g
$\hat{x}_{g,t_O}$	An estimator for x_{g,t_O}
$a_{s,g,t}$	$a_{s,g,t} \in [0, 1]$ is the proportion of UEs that are located in grid g and connected to radio sector s at time t
$a_{s,g}$	$a_{s,g} \triangleq \mathbf{E}[a_{s,g,t}]$ is the average proportion of UEs that are located in grid g and connected to radio sector s
$u_{s,g}$	$u_{s,g} \triangleq r_g \cdot a_{s,g}$
$\hat{u}_{s,g}$	The Least Squares estimator for $u_{s,g}$
$\xi^{(r)}_{g,t}$	$\xi^{(r)}_{g,t} \triangleq r_{g,t} - r_g$ is the noise component of $r_{g,t}$
$\xi^{(a)}_{s,g,t}$	$\xi^{(a)}_{s,g,t} \triangleq a_{s,g,t} - a_{s,g}$ is the noise component of $a_{s,g,t}$

B ADDITIONAL EXPERIMENTS FOR TOWER NTE REHOMING

The network topology we used for tower NTE rehoming experiment in Section 5.3 is that of a medium size LATA. In this section, we present the results of two additional experiments performed on a smaller topology and a larger topology respectively. We first introduce these two topologies in Sections B.1 and B.2 respectively, and then describe the experiment results in Section B.3.

B.1 The Smaller Topology

The first network area we investigated in this section is illustrated in Figure 15, which is from a small size LATA. This is a inland metropolitan area. There are a total of 444 towers and 427

potential network segment failures. Its corresponding simulation results will be shown in Figure 17 in Section B.3.

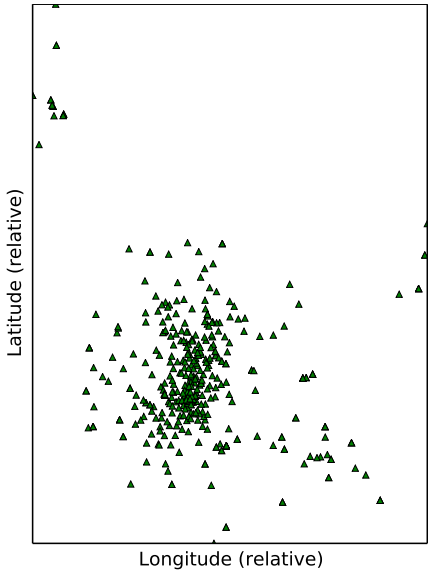


Fig. 15. A real network topology with 444 towers in a *small* size LATA.

B.2 The Larger Topology

The second network area we investigated is illustrated in Figure 16, which is from a large size LATA. This is a lakeside metropolitan area. There are a total of 1,364 towers and 1,445 potential network segment failures. Its corresponding simulation results will be shown in Figure 18 in Section B.3.

B.3 Experimental Results

The two sets of experimental results for the topologies described in Sections B.1 and B.2 are shown in Figures 17 and 18, respectively. These experimental results lead to the same conclusion as we can draw from the results of the first experiment that are described in Section 5.3: These results demonstrate that our TOIP approach significantly outperforms the prior approach. More specifically, both the average (Figures 17a and 18a) and the maximum (Figures 17b and 18b) customer impacts of the potential single network segment failures for the TOIP approach drop much faster than those for the prior approach and the TOIP tower-rehoming approach is able to reduce the customer impact of each single network segment failure to a negligible value after rehoming only about 10% of the towers.

ACKNOWLEDGMENTS

We thank our shepherd Gil Zussman and the anonymous reviewers for their insightful feedback on the paper. We are grateful to Rick Weeks from AT&T's Network Planning and Engineering team for his invaluable suggestions to improve TOIP's usability in the field and case-study analysis.

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

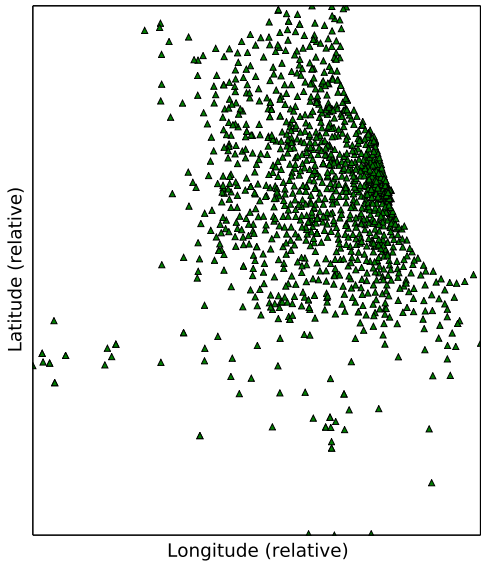


Fig. 16. A real network topology with 1,364 towers in a *large* size LATA.

REFERENCES

[1] 3GPP. 2017. *Telecommunication management; Self-Organizing Networks (SON); Self-healing concepts and requirements.* TS 32.541. 3rd Generation Partnership Project (3GPP). http://www.3gpp.org/ftp/Specs/html-info/32541.htm

[2] AC Aaitken. 1935. On least Squares and Linear Combination of Obsevations.". *Proc. Of the Royal Society of Edinburgh* 55 (1935), 42–48.

[3] Mehdi Amirijoo, L Jorguseski, T Kurner, R Litjens, M Neuland, LC Schmelz, and U Turke. 2009. Cell outage management in LTE networks. In *Wireless Communication Systems, 2009. ISWCS 2009. 6th International Symposium on.* IEEE, 600–604.

[4] Mehdi Amirijoo, L Jorguseski, R Litjens, and R Nascimento. 2011. Effectiveness of cell outage compensation in LTE networks. In *Consumer Communications and Networking Conference (CCNC), 2011 IEEE.* IEEE, 642–647.

[5] M. Austin, J. Fix, S. Meredith, S. Puthenpura, and G. Meempat. 2012. Location Estimation of a Mobile Device in a UMTS Network. (March 1 2012). http://www.google.com/patents/US20120052883 US Patent App. 12/870,254.

[6] Ulrich Barth. 2009. Self-X RAN: Autonomous self organizing radio access networks. In *Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks, 2009. WiOPT 2009. 7th International Symposium on.* IEEE, 1–2.

[7] Simon C Borst, Arumugam Buvaneswari, Lawrence M Drabeck, Michael J Flanagan, John M Graybeal, Georg K Hampel, Mark Haner, William M MacDonald, Paul A Polakos, George Rittenhouse, et al. 2005. Dynamic optimization in future cellular networks. *Bell Labs Technical Journal* 10, 2 (2005), 99–119.

[8] John S. Chipman. 2011. *Gauss-Markov Theorem.* Springer Berlin Heidelberg, Berlin, Heidelberg, 577–582. https://doi.org/10.1007/978-3-642-04898-2_270

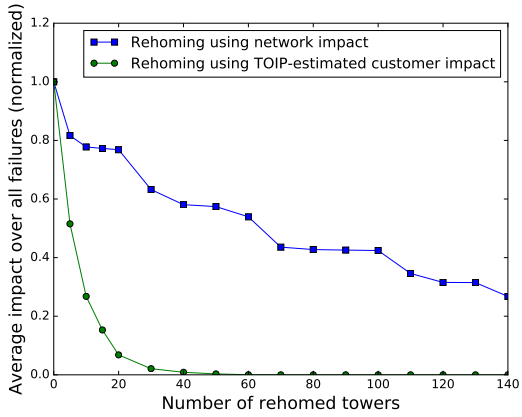
[9] Christopher Cox. 2014. Self-Optimizing Networks. *Introduction to LTE, An: LTE, LTE-Advanced, SAE, VoLTE and 4G Mobile Communications* (2014), 289–299.

[10] Thomas B Fomby, R Carter Hill, and Stanley R Johnson. 2012. *Advanced econometric methods.* Springer Science & Business Media.

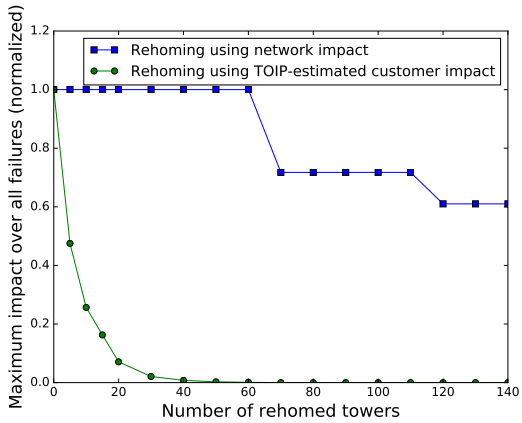
[11] Muhammad Aatiq Ismail, Xiang Xu, and Rudolf Mathar. 2013. Autonomous antenna tilt and power configuration based on CQI for LTE cellular networks. In *Wireless Communication Systems (ISWCS 2013), Proceedings of the Tenth International Symposium on.* VDE, 1–5.

[12] Alexis Kwasinski. 2013. Lessons from field damage assessments about communication networks power supply and infrastructure performance during natural disasters with a focus on Hurricane Sandy. In *FCC Workshop on Network Resiliency 2013.*

[13] Fu-qiang Li, Xue-song Qiu, Luo-ming Meng, Heng Zhang, and Wenzhe Gu. 2011. Achieving cell outage compensation in radio access network with automatic network management. In *GLOBECOM Workshops (GC Wkshps), 2011 IEEE.* IEEE, 673–677.



(a) *Average* customer impact (normalized) over all the 427 network segment failures after tower NTE rehoming.



(b) *Maximum* customer impact (normalized) over all the 427 network segment failures after tower NTE rehoming.

Fig. 17. Tower rehoming experiment results for the network topology given in Section B.1.

[14] Ajay Mahimkar, Zihui Ge, Jia Wang, Jennifer Yates, Yin Zhang, Joanne Emmons, Brian Huntley, and Mark Stockert. 2011. Rapid detection of maintenance induced changes in service performance. In *Proceedings of the Seventh COnference on emerging Networking EXperiments and Technologies*. ACM, 13.

[15] Ajay Mahimkar, Zihui Ge, Jennifer Yates, Chris Hristov, Vincent Cordaro, Shane Smith, Jing Xu, and Mark Stockert. 2013. Robust assessment of changes in cellular networks. In *Proceedings of the ninth ACM conference on Emerging networking experiments and technologies*. ACM, 175–186.

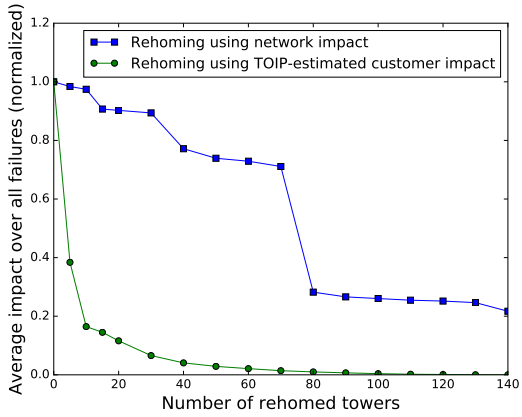
[16] Ajay Anil Mahimkar, Han Hee Song, Zihui Ge, Aman Shaikh, Jia Wang, Jennifer Yates, Yin Zhang, and Joanne Emmons. 2010. Detecting the performance impact of upgrades in large operational networks. In *ACM SIGCOMM Computer Communication Review*, Vol. 40. ACM, 303–314.

[17] Biswanath Mukherjee, M Habib, and Ferhat Dikbiyik. 2014. Network adaptability from disaster disruptions and cascading failures. *IEEE Communications Magazine* 52, 5 (2014), 230–238.

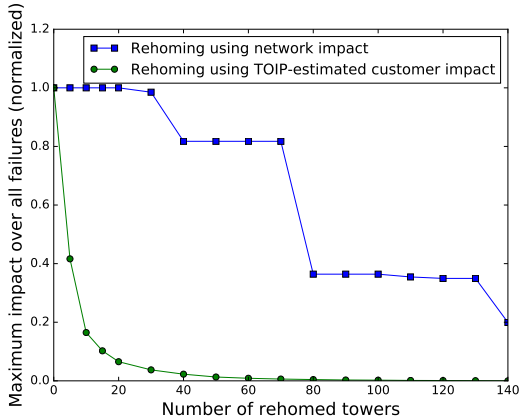
[18] Shivendra Panwar. 2013. Building Robust Cellular Networks. In *FCC Workshop on Network Resiliency 2013*.

[19] Jennifer Rexford. 2013. Lessons Learned from the 9/11 Attacks. In *FCC Workshop on Network Resiliency 2013*.

[20] LC Schmelz, JL Van Den Berg, R Litjens, K Zetterberg, M Amirijoo, K Spaey, I Balan, N Scully, and S Stefanski. 2009. Self-organisation in wireless networks use cases and their interrelation. In *Wireless World Res. Forum Meeting*, Vol. 22. 1–5.



(a) *Average* customer impact (normalized) over all the 1,445 network segment failures after tower NTE rehoming.



(b) *Maximum* customer impact (normalized) over all the 1,445 network segment failures after tower NTE rehoming.

Fig. 18. Tower rehoming experiment results for the network topology given in Section B.2.

[21] Rahul Singh, Prashant Shenoy, Maitreya Natu, Vaishali Sadaphal, and Harrick Vin. 2013. Analytical modeling for what-if analysis in complex cloud computing applications. *ACM SIGMETRICS Performance Evaluation Review* 40, 4 (2013), 53–62.

[22] James P. G. Sterbenz. 2013. Diverse Network Infrastructure for Resilience and Rapid Recovery from Large-Scale Disasters. In *FCC Workshop on Network Resiliency 2013.*

[23] M.B. Tariq, K. Bhandankar, V. Valancius, A. Zeitoun, N. Feamster, and M. Ammar. 2013. Answering Deployment and Configuration Questions With WISE: Techniques and Deployment Experience. *Networking, IEEE/ACM Transactions on* 21, 1 (Feb 2013), 1–13. https://doi.org/10.1109/TNET.2012.2230448

[24] Eno Thereska, Michael Abd-El-Malek, Jay J Wylie, Dushyanth Narayanan, and Gregory R Ganger. 2006. Informed data distribution selection in a self-predicting storage system. In *Autonomic Computing, 2006. ICAC'06. IEEE International Conference on.* IEEE, 187–198.

[25] John Thomas. 2013. Minimizing the Risk of Communication Failure. In *FCC Workshop on Network Resiliency 2013.*

[26] Li Wenjing, Yu Peng, Jiang Zhengxin, and Li Zifan. 2012. Centralized management mechanism for cell outage compensation in LTE networks. *International Journal of Distributed Sensor Networks* (2012).

[27] Halbert White. 1980. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica: Journal of the Econometric Society* (1980), 817–838.

[28] Lingfeng Xia, Wenjing Li, Heng Zhang, and Zhili Wang. 2011. A cell outage compensation mechanism in self-organizing RAN. In *Wireless Communications, Networking and Mobile Computing (WiCOM), 2011 7th International Conference on*. IEEE, 1–4.

[29] Xing Xu, Ioannis Broustis, Zihui Ge, Ramesh Govindan, Ajay Mahimkar, NK Shankaranarayanan, and Jia Wang. 2015. Magus: minimizing cellular service disruption during network upgrades. In *Proceedings of the 11th ACM Conference on Emerging Networking Experiments and Technologies*. ACM, 21.

[30] Jin Yang and Jinsong Lin. 2000. Optimization of power management in a CDMA radio network. In *Vehicular Technology Conference, 2000. IEEE-VTS Fall VTC 2000. 52nd*, Vol. 6. IEEE, 2642–2647.

[31] Gil Zussman. 2013. The Vulnerability of Fiber Networks and Power Grids to Geographically Correlated Failures. In *FCC Workshop on Network Resiliency 2013*.

Received August 2017; revised October 2017; accepted December 2017.