Technical Report

Department of Computer Science
and Engineering
University of Minnesota
4-192 Keller Hall
200 Union Street SE
Minneapolis, MN 55455-0159 USA

TR 12-019

Understanding the complexity of 3G UMTS network performance - Modeling, Prediction, and Diagnosis

Yingying Chen, Nick Duffield, Patrick Haffner, Wen-ling Hsu, Guy Jacobson, Yu Jin, Subhabrata Sen, Shobha Venkataraman, and Zhi-li Zhang

July 30, 2012

Understanding the Complexity of 3G UMTS Network Performance - Modeling, Prediction, and Diagnosis

Yingying Chen[2], Nicholas Duffield[1], Patrick Haffner[1], Wen-Ling Hsu[1], Guy Jacobson[1], Yu Jin[1], Subhabrata Sen[1], Shobha Venkataraman[1], and Zhi-Li Zhang[2]

[1]AT&T Labs Research

[2]University of Minnesota-Twin Cities

ABSTRACT

With rapid growth in smart phones and mobile data, effectively managing cellular data networks is important in meeting user performance expectations. However, the scale, complexity and dynamics of a large 3G cellular network make it a challenging task to understand the diverse factors that affect its performance. In this paper we study the RNC (Radio Network Controller)-level performance in one of the largest cellular network carriers in US. Using large amount of datasets collected from various sources across the network and over time, we investigate the key factors that influence the network performance in terms of the round-trip times and loss rates (averaged over an hourly time scale). We start by performing the "first-order" property analysis to analyze the correlation and impact of each factor on the network performance. We then apply RuleFit – a powerful supervised machine learning tool that combines linear regression and decision trees – to develop models and analyze the relative importance of various factors in estimating and predicting the network performance. Our analysis culminates with the detection and diagnosis of both "transient" and "persistent" performance anomalies, with discussion on the complex interactions and differing effects of the various factors that may influence the 3G UMTS network performance.

1. INTRODUCTION

The wide adoption of smart phones and other mobile devices such as smart tablets and e-readers has spurred rapid growth in mobile data. In order to meet user performance expectations and enhance user experiences, effectively managing cellular data networks is imperative: for example, quickly trouble-shooting performance issues as they arise, or adding additional capacity where the network elements are overloaded. Due to the sheer scale, complexity and dynamics of a typical large scale cellular data network, there is a myriad of diverse factors that may affect the performance of a cellular data network, from the types, geographical locations and coverage of various network elements (e.g., cell towers/base stations, radio network controllers, IP gateways or routers) within the network infrastructure, to the number of users served by each network element, the amount of traffic generated by the users, user usage patterns and behaviors (e.g., mixtures of dominant applications) across different times of days and weeks, to user handset side as well as the (application) server side issues.

In this paper, utilizing the massive amount of performance and other data collected in one of the largest cellular network carriers in US, we set out to analyze and understand the various key factors that may influence the network performance in a large UMTS cellular data network. For instance, we are interested in understanding how the geographical coverages of nodeBs (*i.e.*,, base stations; please refer to Section 2 for a quick overview of the UMTS architecture and its basic terminologies) and their distances to corresponding RNCs and/or SGSNs in the radio access network subsystems in general affect the (average) round trip delays experienced by users, how the overall system utilization (e.g., in terms of overall packet, byte or flow counts), the number of subscribers and their usage patterns (e.g., the application mix) may have an impact on the (average) loss rates across the network, how the time-of-day or day-of-week effect correlate with the overall network performance over time, and so forth. Furthermore, we are interested in dissecting how the confluence of various factors may interact or counteract with each other in different parts of the network, thus having a differing effect on the overall network performance across the network. For these purposes, we aim to build *macroscopic* models that can help identify and access the various major factors that may (or may not) influence the network performance, and predict their effects *across the network* and *over time*. (This is as opposed to *microscopic* models that target specific network elements or users, e.g., to trouble-shoot (transient or persistent) performance issues experienced by certain network elements (e.g., cell towers) or users. Building effective microscopic models requires far more detailed, fine-grained (and perhaps also lower-level) measurement data.) Our goal is to provide network operators with a "big picture" understanding of the major factors that influence the overall network performance across different parts of the network and over longer time periods.

To understand the diverse factors that may influence the network performance, we gather and collect information and data from a variety of sources in the large 3G UMTS cellular network. These include "static" information regarding the UMTS infrastructure, such as the GPS location of each NodeB, the number of sectors per NodeB, the corresponding RNCs that NodeBs are associated with, the SGSNs that RNCs are connected to, the equipment type and vendor of each RNC, and so forth. We also collect IP flow-level data at each SGSN/GGSN, where each TCP/UDP flow can be attributed to one of the many RNCs connecting to an SGSN. From the flow-level data, we can obtain important usage-related factors such as the packet counts, byte counts, and flow counts as well as the average number of bytes per flow, e.g., for each RNC in every hour. In addition, we classify flows based on the known applications such as web, email, VoIP, streaming video, MMS (multimedia messaging service), Jabber (another messaging service), Appstore downloads, etc.; the application mix therefore reflects the user behavior at various locales. Clearly, the factors obtained using the flow-level data are time-varying, and thus *dynamic* in nature. Both to reduce the amount of data and to provide a "big picture" view of the large cellular network, we aggregate the flow-level statistics at the *RNC-level* and compute the averages on an *hourly* basis. Last but not the least, as two important indicators of the overall network performance, we consider *round-trip times* (RTTs) and *loss rates*, averaged on an hourly basis and aggregated at the RNC-level. The RNC-level, hourly averaged RTTs and loss rates are calcuated based on the average RTT and loss rate of each TCP flow belonging to an RNC [24].

Clearly, one may expect that certain factors (e.g., distance of nodeBs to an RNC or SGSN) are more important in terms of their contribution to the overall user delay performance, while others (e.g., heavy usage of stream videos) can be more critical in terms of their impact on the loss rate performance. Unfortunately, due to its vast diversity and complexity (for instance, the numbers of NodeBs and sectors as well as the geographical coverage can vary significantly from one RNC to another, and the usage patterns and application mixes differ also markedly across the network and over time), teasing out each factor individually is not an easy task. Not only are there a vast array of diverse factors, but many of these factors are also intertwined, effecting disparate influences in different parts of the network. Furthermore, there may be *latent* factors that are not explicitly accounted for, or captured at all, in the information and data we have collected.

To circumvent these challenges, our approach is to first take several major classes of factors and perform network-wide correlation and other "first-order" analysis to understand how significantly each of these major factors *individually* influence the overall RNC-level network performance (in terms of hourly RTTs and loss rates). This provides us with a baseline understanding of the individual effect (or lack thereof) that each of the major classes of factors has on the overall network performance at the RNC-level. Next, to examine the collective effect of various factors on the network performance and sort out their relative contributions, we apply RuleFit [12] – a powerful predictive learning method via rule-emsembles, combining both linear regression and decision trees. RuleFit provides both predictive and interpretative capabilities that are needed for our analysis. By aggregating as well as dividing the datasets along different dimensions (e.g., based on geographical locations such as states or along time such as days or weeks), we intelligently apply RuleFit to build macroscopic models to assess and dissect the various major factors that affect part or the whole network, and predict their impact on the network performance over time. For example, by analyzing the network performance predictions over time and comparing the model obtaining using the datasets from all RNCs with those obtained from the RNCs located within certain geographical regions, we can identify the major factors that have persistent network-wide effect as well as uncover those factors that have marked impact at certain locales or at certain times. Furthermore, by examining the level of the overall contributions of all factors as well as the relative importance of each factor and how they differ across locales or over times, we can also infer whether there are potential latent factors in play, affecting the network performance in a way that cannot be quantified using the factors explicitly included in the models. Significant deviations from the model predictions may also signal anomalous events, and can be used for diagnosis purposes (see Sections 6, 7 for details).

In summary, the models and insights obtained from our study provide the network operators with a "big picture" understanding of the overall RNC-level RTT and loss rate performance in a large 3G cellular network and the various factors that may influence such performance, thereby the user experiences. Such a "big picture" understanding can help network operators better cope with the management challenges posed by the scale and complexity of a large cellular network. For example, our relative importance analysis reveals that the placement of the network facilities has a major effect on the RTT performance; whereas loss rate performance may hinge on the collective effect of the type of equipment being used, the usage patterns such as the application mix, and flow size, etc. The factors identified by models suggest that the network operators can reduce latency by improving the network coverage in some parts of the network, or add additional network capacity in other parts of the network to meet the user demands due to the increasing usage of bandwidth-intensive applications. While our study is focused primarily on *macroscopic* models, as the results in Section 7 illustrate, they may also help diagnose and trouble-shoot the network performance issues associated with certain part of the network or specific network elements by detecting anomalies and pointing to potential latent factors that are in play. Finally, we believe that the methodology developed in this paper can be applied to other 3G network architectures, and possibly also the emerging 4G (LTE) cellular networks.

2. BACKGROUND AND DATASETS

In this section, we give a brief overview of the architecture of the typical 3G UMTS network, and the datasets we collected for our study.

2.1 UMTS Network Overview

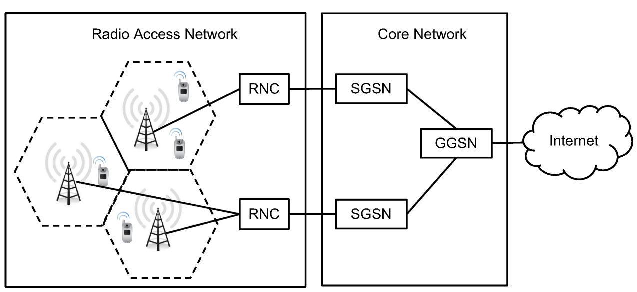

Figure 1: UMTS network architecture

As illustrated in Figure 1, a typical UMTS network consists of two major entities, the UMTS Terrestrial Radio Access Network (UTRAN) and the core network. UTRAN consists of base stations (NodeB), and Radio Network Controllers (RNC) which are connected to and control a number of NodeBs. Each NodeB is typically configured with multiple sectors, e.g., 3 sectors (the common configuration) in three different directions, each covering 120 degree range. If there are more than three sectors associated with one NodeB, multiple sectors are overlapped on each direction, and are distinguished using different frequencies. Moreover, NodeB usually supports multi-carrier technology. Via software configuration, each sector can support multiple carriers in order to further increase its capacity. The core network is comprised of the Serving GPRS Support Nodes (SGSN) and the Gateway GPRS Support Nodes (GGSN). To connect to, say, a web server located in the outside Internet, a User Equipment (UE) will first contact the nearest NodeB. After receiving the user data access requests on one of its sectors, NodeB will handover the user requests to its upstream RNC, which further forwards the data service requests to an SGSN. In the core network, the SGSN establishes a tunnel with a GGSN, using the GPRS Tunneling Protocol (GTP). The data is carried as IP packets in the tunnel and finally reaches the web server in the external Internet.

2.2 Datasets

For this study, we combine datasets collected from a variety of sources in order to gain a more comprehensive view of the various factors that may influence the network performance. As mentioned in the introduction, these data sources include "static" information regarding the UMTS infrastructure, such as the GPS location (and zipcode) of each NodeB, the number of sectors (and carriers) per NodeB, the corresponding RNCs that NodeBs are associated with, the SGSNs that RNCs are connected to, the equipment type and vendor of each RNC, and so forth. To gain a sense of the population density and demographics in each base station/RNC coverage area, we also utilized the 2000 census data[1], which contains the land area coverage of each zip code. From these static data sources, we can estimate the geographic areas that are covered by each NodeB, RNC and SGSN, and get a sense of the population density and other demographic information (e.g., rural vs. small city vs. large metro, etc.). Therefore, these static factors can help us dissect and understand the impact from the geographical coverage, regional variations, the placement of the networking facilities, and other infrastructure-related issues.

There are two major sources of *dynamic* data that are used in our study. One is the IP traffic data collected periodically at each GGSN. Because of the large volume of the traffic, all the measurements were computed at the RNC level, i.e. aggregated for all the users served by the same RNC, and at the granularity of an hour. (The time mentioned in this paper always refers to the local time of the RNC location.) From the IP traffic data, we obtain the RTT and loss rate performance measurement data and the usage related statistics such as the number of bytes, flows, packets, average flow sizes, and so forth, as well as the application classifications and mixes (e.g., email, VoIP, streaming video, MMS, Jabber, Appstore downloads) – traffic that cannot be classified is labelled *unknown*. The other dynamic data source contains more lower level information, such as the total number of Radio Resource Control (RRC) [1] attempts served by each RNC at every hour. An RRC attempt indicates a connection establishment attempt between the UEs and the

[1]While the census data is almost 12 years old, the land area coverage per zip code does not change drastically over the years. Moreover, we do not make any direct conclusion based on the census data, but rather combine it with other data and techniques for our study.

UTRAN, and therefore the #RRC attempts can be a good approximation of the #requests served at each RNC. The datasets collected span more than 6 months. However, as representative examples and for illustrative purposes, we will focus on the datasets collected during a two-week period in September 2011. To adhere to the confidentiality under which we had access to the data, at places, we present normalized views of our results while retaining the scientifically relevant bits.

3. PROBLEM SETTING & ILLUSTRATION

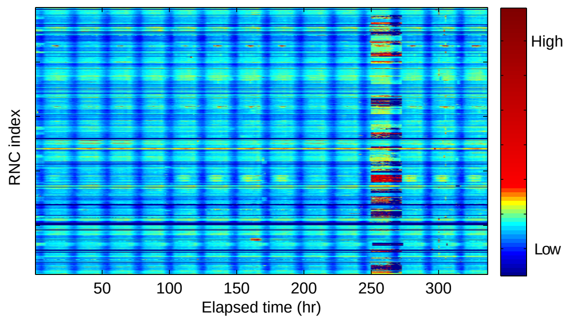

Figure 2: RTT time series across all RNCs.

Due to the sheer scale, complexity and dynamics of the large 3G UMTS cellular network, there are a myriad of complex factors that may potentially affect or influence the overall network performance. Some of these factors may depend on other factors, and interact with each other differently in different parts of the networks. To identify the major factors and tease out the effect of each factor – and understand how they affect the overall network performance in different parts of the network and over time – are extremely challenging tasks. In this paper, we consider RTT and loss rate (aggregated at the RNC-level and averaged hourly) as two key performance indicators. In Figures 2, we plot the RTTs of all RNCs in the network over a two-week time period. The x-axis is the elapsed time in hours since 12AM on Sep.2nd, 2011. The y-axis is the index of all the RNCs, which are grouped based on the state they (primarily) serve. The colorbar on the right is the indicator of the RTT performance level. [2] As we observe from the figure, not only that RNCs have large variability across times (the *temporal variability*), e.g., a strong diurnal pattern, but also across different RNCs (the *spatial variability*). For loss rate, we observe similar trends. Moreover, the RTT and loss rate time series are positively correlated in a considerable manner. The correlation coefficient of the average RTT and loss rate (averaged across all RNCs) is 0.7 (see Table 1, which also includes the pairwise correlation coefficients of RTT, loss rate and other usage statistics: flow size – average flow size, #bytes – the bytecount, # flows – the flow count at each RNC per hour, averaged across all RNCs). This suggests that time dependent factors are possibly one of the main drivers for the performance variation over time.

[2]Given the sensitivity of the data, and the privacy of the users, the abosolute values of the metrics throughout our study are anonymized.

On the other hand, the spatial variability of RTTs and loss rates cannot be attributed to time dependent factors. Clearly some geograpphic conditions and location-dependent factors (e..g., varying usage patterns or different mixtures of applications served by different RNCs) may come into play here. In terms of their spatial variability, the RTT and loss rate have very weak correlation (in most hours), with values generally smaller than 0.4 and sometimes negatively correlated (see Figure 3). In other words, suppose RNC A has larger RTT than RNC B at a particular hour, it does not necessarily imply that A is also likely to have a larger loss rate than B in the same hour. This suggests that there probably exists two separate sets of factors which contribute to the spatial diversity in the RTT vs. loss rate performance.

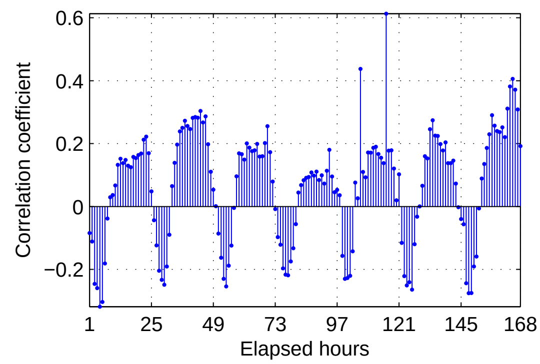

Figure 3: Diurnal pattern in the correlations between RTT and loss rate spatial variability.

It is aslo interesting to observe that although their spatial variability is weakly correlated, there exhibits a distinct diurnal pattern, as clearly shown in Figure 3. This also holds true in terms of the spatial variability of other usage metrics, such as the #bytes, #flows, flowsize. These observations indicate that the overall network performance, whether RTT or loss rate, can be affected by a combination of many factors. Their relative importance or contribution to the network performance may also differ across the network, and vary over time.

4. FIRST-ORDER PROPERTY ANALYSIS

In this section we start by looking into several major classes of factors that are expected to have likely influence on the RNC-level network performance. These include various user usage-related factors, infrastructure-related factors (e.g., locations, placement and coverage of network facilities) as well as device factors (type

Table 1: Correlation coefficient of all pairwise temporal variability.

	RTT	loss	flowsize	#bytes	#flows
RTT	1	0.7017	0.2528	0.9763	0.9746
loss	0.7	1	-0.3794	0.6582	0.7606
flowsize	0.2528	-0.3794	1	0.3269	0.1619
#bytes	0.9763	0.6582	0.3269	1	0.9839
#flows	0.9746	0.7606	0.1619	0.9839	1

and vendor of network equipment). For each of them, we present metrics for characterization and estimation (where needed), and perform network-wide correlation and other "first-order" analysis to get a better sense of how they positively (or negatively) influence the performance. This provides us with a baseline understanding of the individudal effect (or lack thereof) that each of the major classes of factors has on the overall network performance at the RNC-level. Clearly, the complexity and diversity of the network making it almost impossible to exhaust all possible factors and perform similar analysis.

4.1 Usage Factors

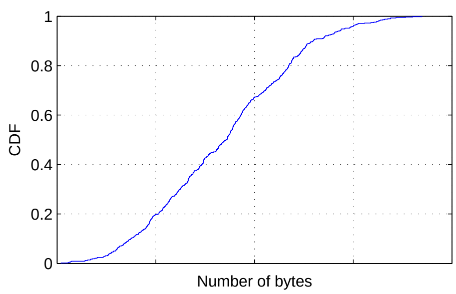

Figure 4: Distribution of traffic load on all RNCs at 11AM.

User behaviors can affect the performance in a number of ways due to the varying traffic load and application mix over different time periods (e.g., the time of the day or the day of the week) and across different parts of the network (as represented by the RNCs). Foremost, it directly shapes the diurnal patterns seen in the performance. As shown in Table 1, the number of bytes or flows served by each RNC has strong correlation with the performance, in particular, RTT. Besides, the traffic load also has a large variation across different RNCs, The cumulative distribution of the total number of bytes are depicted in Figure 4. The largest byte counts can be 10 times the smallest byte counts per hour. A similar distribution is also observed for the number of RRC establishments seen on these RNCs per hour, which can be a good approximation of the number of active subscribers and their data accesses at that hour. This large variation of traffic load across RNCs is an illustration of the huge diversity (in terms of users and their data access activity) that exists in the large UMTS network.

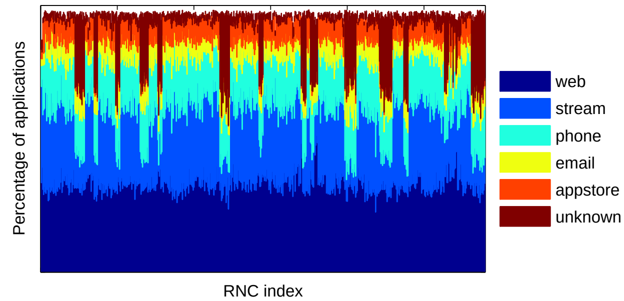

Figure 5: Fractions of major applications on all RNCs at 11AM.

In addition to the traffic load, application mix also exhibits large variation and diversity over time and across the RNCs. There are altogether 16 known categories of applications, and one unknown category. Most RNCs see a significant increase in streaming and appstore traffic during the nighttime, whereas during the daytime, web, email and other traffic tend to dominate. The breakdowns of the major applications such as web, streaming, smart phone, email, appstore, as well as the unknown traffic across different RNCs are shown in Figure 5. The x-axis is the index for all the RNCs, where the RNCs within the same state are grouped closer to each other. The composition of the traffic exihibits a clear geographical distinction. While most RNCs contain a large portion of streaming traffic, several clusters of RNCs contain much larger portions of unknown traffic. A closer look at the dataset reveals that they represent RNCs coming from 14 different states. As will be discussed later, these states are also among those that tend to persistently suffer the worst network performance in terms of loss rate.

4.2 Infrastructural Factors

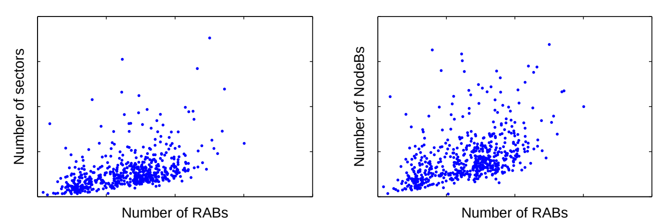

Figure 6: Correlation of user requests and #sectors deployed.

To meet the diversified user demands at different RNCs, the number of NodeBs, sectors, and carriers associated with each RNC also vary accordingly, as manifested in Figure 6. For those regions with a large or an increasing number of UMTS subsribers, multiple NodeBs are

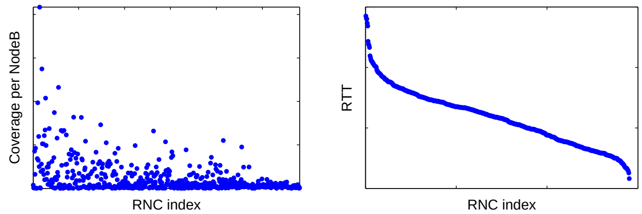

Figure 7: The effect of NodeB coverage on RTT.

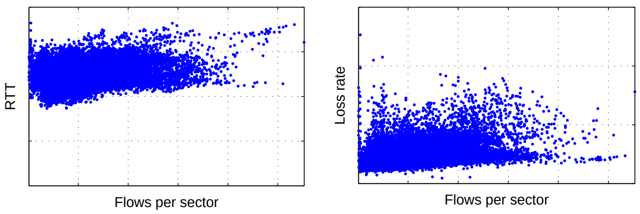

Figure 8: The impact of the number of flows per sector on performance.

likely deployed in the same geographical location. Both the number of devices being placed and the geographical placement strategies used can play a crucial role in influencing the network performance experienced by the users in each region.

First, the more sectors or carriers are deployed, the less load will incur on each carrier or sector. The impact of the load per sector or per carrier, as characterized by the number of flows per sector, on the network performance is shown in Figure 8 using a one-day dataset. The basic trend indicates that larger flows per sector tends to incur larger RTT and loss rate. Note that it is not a perfect linear curve, as the number of flows per sector is not the sole factor affecting the network performance. Other factors such as the varying behaviors of users at different hours also come into play. Other metrics such as the number of bytes per sector, users per sector, as well as the load on the carrier level exhibit similar impact on the network performance.

Second, an increasing number of NodeBs within a geographical area of a fixed size can potentially improve the RTT performance in that it leads to a decrease in the coverage area per NodeB, and therefore smaller "last mile" network latency between UEs and NodeBs. To confirm and verify this intuition, we estimate the coverage area of each nodeB (which is not directly accessible from our data) using a combination of information from the census data and inference using the Voronoi Diagram (based on the GPS locations of NodeBs). The census data provides us with information such as the land area covered by each zip code. Along with the zip code information for each NodeB, we can approximately compute the coverage area per NodeB. However, since multiple NodeBs can be deployed even within the same zip code area, as such this approximation may overestimate the real coverage area of a NodeB. To overcome this problem, we also apply the Voronoi Diagram to demarcate the serving area per NodeB, based on the GPS locations of the NodeBs. The limitation with this method is that not all US land areas (e.g., deserts and other uninhabitable areas) may be covered by the NodeBs. (Another technicality due to the use of the Voronoi Diagram is that the NodeBs on the US boundary are considered to have an "infinite" area – this effect can be mitigated by explicitly including the contour of the US boundary into the estimation process.) Therefore, both methodologies may overestimate the coverage area in some way. For this reason, we infer the coverage area per NodeB by taking the minimum of the results obtained by both methods. In Figure 7, the impact of our estimated NodeB coverage area on the RTT performance is plotted. As expected, RNCs with NodeBs that have smaller coverage areas tend to have smaller average RTT performance.

Similar to the location and placement of NodeBs, the number of RNCs deployed in a geographical area and the coverage areas of RNCs are likely driven by the user/subscriber population and the load they generate. The same also holds true for SGSNs, but to a lesser degree. In general, due to their smaller numbers, RNCs and SGSNs are not as geographically spreaded as NodeBs. For instance, RNCs and SGSNs may be placed in certain selected locations for ease of management. Overall, the distance between RNCs and SGSNs are usually stable, and there is far less variability in the distance between RNCs and SGSNs than there is between NodeBs and RNCs. The latter depends highly on the geograghical coverage of each RNC, and varies from region to region. Nonetheless, both the distance between NodeBs and RNCs and the distance between RNCs and SGSNs can potentially have a significant impact on the network performance, especially the RTT performance.

4.3 Device Factors

In addition to the aforementioned factors, the choice of vendor for the networking devices such as NodeBs and RNCs may also have an effect on the network performance, as illustrated below. This may be due to the fact that different vendors can have fairly different hardware specifications [6], resulting in different data processing speeds and capacities.

In the large cellular network studied in this paper, two major types of devices are used, vendor-x and vendor-y. The type of NodeB is usually chosen to be the same as the RNC it connects to, most likely due to the hardware level compatibility issue. The comparison of the RTT and loss rate performance distributions of these two types of RNCs at two different hours is shown in

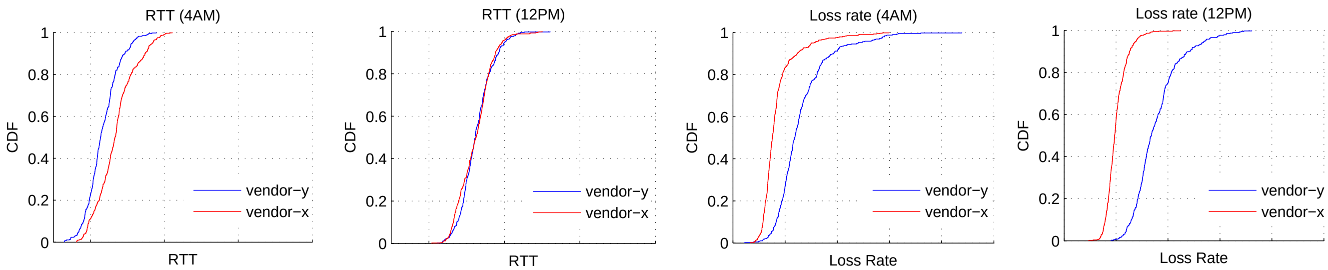

Figure 9: Comparing the performance for two major types of devices.

Figure 9, one for 12PM and the other for 4AM. While the RTT performance distribution is quite similar for both types for most of the 24 hours, all RNCs of vendor-x outperform those from vendor-y in terms of their loss rate. Moreover, such contrasting performance is persistently observed across all 24 hours. More detailed investigation reaveals that the maximum capacity for NodeBs of vendor-x is 6 sectors and 12 carriers, whereas NodeBs of vendor-y only have 3 sectors and 6 carriers. To increase the capacity, additional carriers are configured on vendor-y's NodeBs through software configuration. However, as the overall data processing capacity of the device is still constrained by the hardware capacity, NodeBs of vendor-y tend to incur more packet losses, especially during high loads.

5. PERFORMANCE MODELING

To understand the complex interaction of various factors and assess their relative importance and contribution to the network performance, we apply RuleFit [4, 12] – a powerful predictive learning method via rule-emsembles – to build macroscopic models to assess and dissect the various major factors that affect part or the whole network, and predict their impact on the network performance over time. RuleFit provides both predictive and interpretative capabilities that are useful for our study in a number of ways. Rulefit is useful for our measurement analysis in a number of ways. First, it provides better modeling accuracy, especially for such a large pool of factors by combining linear regression and decision tree. Second, Rulefit ranks the factors based on their importance to the output performance, thereby enabling the relative important analysis. Third, by analyzing the fitting accuracy of the model across states and hours, or comparing the model obtained using the datasets from all RNCs with those obtained from the RNCs located within certain geographical regions, we can identify the major factors that have persistent network-wide effect as well as uncover those factors that have marked impact at certain locales or at certain times. Furthermore, by examining the level of the overall contributions of all factors as well as the relative importance of each factor and how they differ across locales or over times, we can also infer whether there are potential latent factors in play, affecting the network performance in a way that cannot be quantified using the factors explicitly included in the models. Significant deviations from the usual model predictions may also signal anomalous events, and can be used for diagnosis purposes.

In this section, we start with an overview of Rulefit. Next, we describe the training and prediction datasets and the list of features (factors) we use as input to Rulefit as well as their mutual interactions. By focusing on a generalized model generated using one-day data, important factors associated with the performance are discussed. Interpretation of the rules generated using the model is also presented. The fitting accuracy across states and hours as well as using it as a mechanism to detect persistent performance anomalies is further discussed.

5.1 Rulefit Overview

Rulefit is a supervised rule-based ensemble learning technique. Given a set of input variables $\mathbf{x}$, the ensemble prediction is a linear combination of the predictions of each ensemble members, which takes the form,

$$F(\mathbf{x}) = a_0 + \sum_{i=0}^{m} a_m f_m(\mathbf{x}) \qquad (1)$$

where a_i is the linear combination parameters, and $f_m(\mathbf{x})$ is a set of base learners. They are different functions of the input variables possibly derived from different parametric families. We consider base learners being a set of simple rules, along with a set of linear functions of the input variables. The rules are generated using decision tree. Each rule base learner is in the form of a conjunctive statement,

$$r(\mathbf{x}) = \prod_j I(x_j \in s_j) \qquad (2)$$

where s_j is a subset of all the possible values of the input variables. $I(.)$ is an indication function, and the rule takes value 1 only if all the arguments hold true. The parameters a_i of the linear regression model are estimated with a lasso penalty. The importance of rule r_k is measured using

$$I_k = |a_k|\sqrt{s_k(1 - s_k)} \qquad (3)$$

where s_k is the support of the rule on the training data. An input variable is considered important if it defines important rule (or linear) predictors. The importance of an input variable x_l is defined as

$$J_l = I_l + \sum_{x_l \in r_k} I_k / m_k \quad (4)$$

where I_l is the importance of the linear predictor, and m_k is the total number of input variables that define the rule r_k.

5.2 Predictor Variables

To characterize the impact from all possible factors, and apply them to Rulefit as the input predictor variables to predict RNC performance, we summarize them into following 35 such metrics,

Usage factors: 1) $nRRC$: number of RRC establishment attempts, 2) $nByte$: number of bytes, 3) $nFlows$: number of flow, 4) $Bpflow$: number of bytes per flow, 5) - 21) fractions of 17 application categories, including misc, web, ftp, instant messaging (IM), P2P, navigation (Nav), VPN, email, Voip, game, Appstore, video optimization (opt), multimedia messaging service (mms), push notification (push), smartphone applications, streaming, and unknown category which is not able to be identified as any of the 16 categories.

Infrastructural factors: 22) $B2RDist$: distance between NodeB and RNC, 23) $R2SDist$: distance between RNC and SGSN, 24) $nSector$: number of sectors, 25) $nCarrier$: number of carriers, 26) $RRCpsec$: number of RRCs per sector, 27) $RRCpcar$: number of RRCs per carrier, 28) $nNodeB$: number of NodeBs, 29) $CovpNB$: land coverage per NodeB, 30) $Bpsec$: number of bytes per sector, 31) $Bpcar$: number of bytes per carrier, 32) $Fpsec$: number of flows per sector, 33) $Fpcar$: number of flows per carrier.

Device factor: 34) $device$: device type.

Beside, we also include geographical *state* as one of the input variables to capture any state specific features. All features are considered as numerical variables, except *device* and *state*, which is in the form of an unorderable categorical variable.

Mutual information among features: To better characterize the features that may affect the network performance and give as much information to the learning algorithm as possible, we present a rich list of features as input predictor variables. Some may share similar information, while others are independent from each other. Rulefit itself provides powerful learning algorithm using decision tree to perform the feature selction. Therefore, the existence of multicollinearity [3] among these features does not affect the accuracy of the model. However, when it comes to compute the relative importance of each factor, Rulefit may consider a feature as unimportant if another feature that is highly correlated with it is selected to build the model. Instead of taking the risk of losing the accuracy of the predictive model [15] or losing the intuitive interpretation power of the model by applying a separate feature selection technique on the dataset before applying Rulefit, we provide more insights of the interactions among these features using the concept *mutual information* [8]. The mutual information of two variables x and y takes the form,

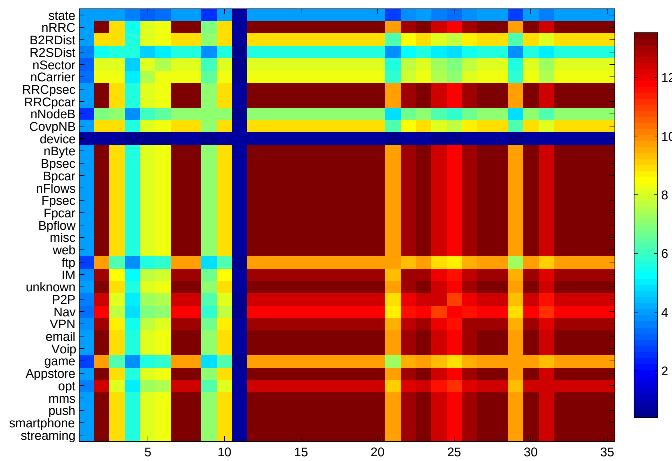

Figure 10: Pairwise mutual information among all features.

$$I_{xy} = \sum_x \sum_y p(x,y) log \frac{p(x,y)}{p(x)p(y)} \quad (5)$$

The idea of mutual information is similar to Pearson correlation except that the former one also takes categorical variables. I_{xy} measures the amount of information that one variable contains about another, which is the reduction in the uncertainty of one variable due to the knowledge of the other. $I_{xy} = 0$ if and only if x and y are independent from each other, meaning knowing one of them give no information about the other. If all the information contained in x is also contained in y, then I_{xy} is the same as the entropy of x. In the case when x and y are identical, I_{xy} is the same as the entropy of x or y.

The pairwise mutual information of all the features are shown in Figure 10. The colorbar on the right indicates the mutual information level. The y-axis lists all the features from top to bottom, and x-axis indicates the same set of features in the same order from left to right, i.e., the value in the ith row from the top and the ith column from the left is the entropy value of the ith variable. As expected, the entropy values on the diagonal of the matrix suggests that the categorical variables such as state and device have less uncertainty, as they only have limited number of possible values associated with them. Second to the categorical variables, the infrastural related variables have less uncertainty than all the application mix related variables. In particular, R2SDist, the distance between RNCs and their SGSNs are more stable due to the fact that only few selected locations are used to place RNCs and SGSNs compared to the span of NodeBs. Among all the applications, ftp and game are two of the applications that have smaller entropy values than the rest of applications, most likely due to the fact that there are fewer users use ftp and game via their mobile UEs.

The values other than the diagonal entries also shown interesting observations. All the values on ith row or ith column indicates the mutual information between the ith variable and all the other variables. The 2nd row or column implies that the number of requests as characterized by nRRC share pretty large amount of information with the traffic load per sector or carrier as well as the usage related variables such as nBytes, Bpsec, etc. Meanwhile, nRRC also contains much of the information as contained in such application mix related variables as misc, web, email, Voip, Appstore, mms, push notification, smartphone apps, and streaming, most likely due to the fact that they are widely used among people, thereby leading to a high correlation between the nRRC and the fractions of these applications.

5.3 Training & Prediction Sets

For the varying purpose of modeling, prediction, and diagnosing the network performance, we organize our whole datasets into subsets of data in three different ways.

The most intuitive way to organize the data is based on the date they were collected. Each subset of the datasets is a one-day data, containing 24 hourly aggregated measurements for all RNCs. This will facilitate our modeling and prediction of the datasets on a daily basis. Starting from Sep.2nd, 2011, each such dataset is labeled as *TS-D-*{*mm/dd*}. However, this organization of the datasets mixes the measurements performed at different hours as well as from RNCs within different states. To better understand the modeling accuracy at specific hours or at particular RNCs, as well as help diagnose the problem specific to certains hours or RNCs, we further reorganize the datasets based on their hour *TS-H-*{*hour*}, and state information. For simplicity, we perform this finer granularity analysis only on the datasets collected at the first 10 days which observe less anomalies compared to the rest of the days. By comparing the model and important factors from different angles, we are potentially able to diagnose the performance and narrow down the issues to certain days, hours, or states.

5.4 Generalized Model

To get better a understanding of the important factors associated with the performance, a generalized model is created and discussed using the first day data TS-D-09/02. The model is generalized in the sense that it is able to *model* the major performance behaviors for RNCs within different states, across different hours, as well as *predict* the performance in the near future.

Important factors: As one of the output of our model, input predictor variables are ranked based on their relative importance to the performance. The top 10 most important factors contributing to RTT and loss rate for the generalized model are listed in Table 2, 3, respectively. The number below each factor is the *relative* importance level of that factor, where 100 is most important and 0 is least important. Both RTT and loss rate performance prove to be highly state dependent, which might due to the varying geographical conditions, localized user interest or application mix, as well as other state specific latent factors.

Beyond the state factor, other important factors suggest that RTT is mainly dependent on such architectural factors as the coverage of NodeB, the distance between NodeB and RNC, etc. To better improve the RTT performance, a better placement strategy of the networking devices is critical. For instance, replacing NodeBs covering huge areas with larger number of NodeBs covering smaller area and transmitting at a lower power might be more cost-effective. Moreover, instead of using a few selected locations for the placement of RNCs, spreading them out may help reduce the distance between NodeB and RNC.

On the other hand, loss rate is more dependent on such factors as the device type, the application mix, and flow size (Bpflow). As we discussed earlier, devices of vendor-x exhibit persistently better loss rate performance than those of vendor-y. Application mix and flow size are highly dependent though. The streaming application is ranked as the most important factor is due to the fact that its flow size is much larger than all other applications.

Another observation is that the metrics associated with NodeBs, such as, flows per sector or flows per carrier play a more important role than those associated with RNCs, such as the load on RNCs, as characterized by the total number of bytes or RRC establishments served on them. That implies the network performance is more constrained by the NodeB performance than RNC. Some problems can possibly get fixed by simply deploying more NodeBs or more sectors on NodeBs, or even configuring more carriers on each sector, which is far more cost-effective than adding more RNCs.

Rules interpretation: The model generated by Rulefit takes the form of a linear combination of a set of easy-to-interpret rules. The coefficients and the support of a rule defines its importance as to how much improvement or degradation it brings to the overall performance. One of the important rules of the generalized model suggests that reducing CovpNB to lower than

Table 2: Top 10 factors and their relative importance to RTT performance.

TS \ Rank	1	2	3	4	5	6	7	8	9	10
TS-D-09/02	state 100	CovpNB 39.72	mms 37.04	B2RDist 15.98	jabber 13.20	Fpsec 5.96	nFlow 4.65	email 4.41	Bpflow 4.24	Bpcar 4.07
TS-D-09/12	state 100	nByte 77.24	CovpNB 53.19	nFlow 33.28	Fpcar 22.64	B2RDist 19.06	Bpcar 17.85	mms 13.81	Fpsec 12.58	RRCpsec 9.34
TS-H-10	State 100	CovpNB 43.30	B2RDist 17.02	R2SDist 10.73	Bpflow 8.29	jabber 6.33	RRCpsec 5.53	nNodeB 4.00	nCarrier 3.76	mms 3.74

Table 3: Top 10 factors and their relative importance to loss rate performance.

TS \ Rank	1	2	3	4	5	6	7	8	9	10
TS-D-09/02	state 100	stream 61.74	device 46.30	Bpflow 36.43	Appstore 28.57	mms 19.13	Voip 15.75	CovpNB 9.84	R2SDist 9.09	web 7.91
TS-D-09/12	Bpflow 100	state 71.45	unknown 45.72	smartphone 22.23	web 21.08	nByte 18.09	email 15.25	nFlow 13.82	RRCpcar 11.97	Fpcar 11.94
TS-H-10	Bpflow 100	state 93.09	device 43.57	stream 20.68	email 15.79	misc 14.57	jabber 12.20	Appstore 11.77	mms 8.18	unknown 7.01

$12.35 miles^2$ in 11 of the states , RTT can be improved by 7.3ms. The support of this rule is 0.33, which is the fraction of data points falling into this category. Another such important rule states that if more than 78 sectors are deployed for RNCs in 16 of the states , their RTT can be improved by 6.69ms.

Similarly, for loss rate, the generated rules also suggest a list of ways that can potentially improve the performance. For instance, if the number of bytes per flow gets larger than 6115 by combining smaller flows, the loss rate can potentially get improved by around 0.16 percent within 13 of the states. On the other hand, for three of the states , if the distance between RNC and SGSN is larger than 270.9 miles, the loss rate will be penalized for 0.28 percent.

Fitting accuracy and detecting *persistent* performance anomalies: As discussed in Section 5 and 4, the performance, as well as the traffic load and application mix have large variations across different RNCs and hours. Our generalized model may not fit equally well for all these scenarios. As depicted in Figure 11, the performance at early morning hours, from 2AM to 4AM, behave fairly different from the rest of the hours. In fact, it improves during those hours, possibly due to the decrease of the overall traffic load and the change of application mix. While most traffic originates from web, email, etc. during daytime, entertainment traffic such as streaming, Appstore becomes more prevalant during nighttime.

In addition, as mentioned in previous section, both RTT and loss rate are highly dependent on the state factor itself. Not surprisingly, the fitting error across different states using the same generalized model also observes a large variation, as shown in Figure 12. The x-axis is the state index. In particular, the performance observed in states labeled as 29, 30, 31 is much worse than other states. These performance anomalies (improving or regressing) exist *persistently*, not just for one or two days. By comparing the fitting accuracy across states and hours, or even individual RNC, Rulefit is able to help us detect them and narrow down to specific hours or states.

Moreover, loss rate is observed to have persistent worse fitting accuracy than RTT, which suggests the possible missing latent factors in our model. Loss rate performance is therefore more complicated than we expected and could depend on lots of other factors.

6. PREDICTION

One important application of the performance modeling is to help predict future network performance from the past. A model that can provide powerful prediction accuracy will help the operators gain more insights of the future needs, and therefore proactively take measures to prevent any predictable performance issues. Any transient performance issue will also be easily detected as a result of that.

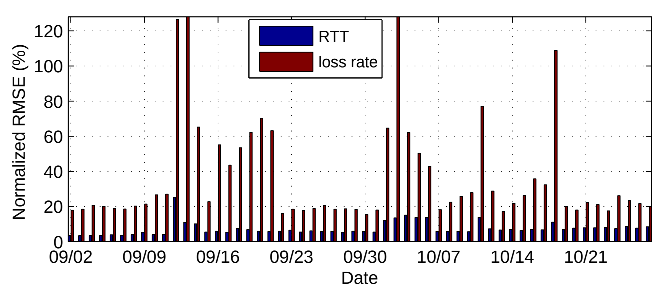

Figure 13: Prediction of a two-month duration.

For this purpose, we conduct a prediction for a two-month duration using the observations in the first week (i.e., Sep.2nd to Sep.8th). Considering the day of week

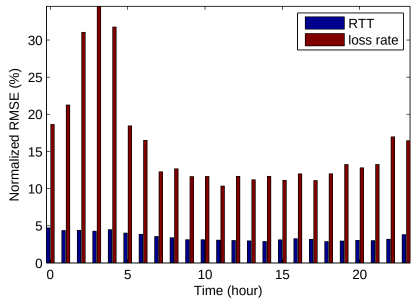

Figure 11: Hourly fitting accuracy on Sep.2nd.

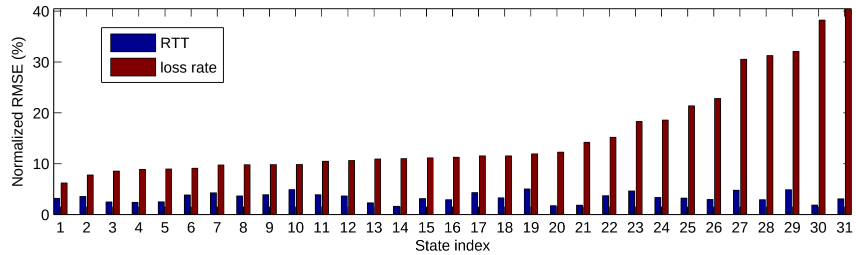

Figure 12: Fitting accuracy across states on Sep.2nd.

pattern of the overall performance, we use the same day within the first week to predict the same day performance of furture weeks. The daily prediction errors as characterized by normalized root mean square eros(NRMSE) are shown in Figure 13. Note that the prediction error for the first week is the indication of the fitting accuracy of the model generated by Rulefit for the real data.

Our observations are two-fold. First, similar to the fitting accuracy across different states and hours of the generalized model, the prediction error is much larger in loss rate than in RTT, partially implying the "complexity" of loss rate performance. On the other hand, most of the future RTT performance is quite predictable mainly due to the fact that it is more dependent on infrastructure related factors. Second, there are more variations in terms of the loss rate performance as it is more susceptible to the change of user related factors. Though the long term trend (even beyond 2-month period) suggests that the loss rate performance is mostly predictable, there are quite a few days showing quite abnormal behaviors. One of the anomaly happens on Sep.12th and 13th, as was displayed in Figure 2. Another anomaly happens the week after, where we observe an increase of loss rate (but not RTT) on quite a number of RNCs. Several other anomalies also happened in October.

In real system, in order to provide better prediction accuracy and alleviate the impact from anomalies in the historical data, as well as be more responsive to more recent trend variations, we can train the model using more (than just one week) historical data and give more weights towards recent data.

7. PERFORMANCE DIAGNOSIS

By adopting Rulefit, two types of preformance regressions were unveiled, *transient* and *persistent.* However, Rulefit is more than just a tool to detect performance issue. Rather, it is also helpful in diagnosing these issues. In this section, we present our approach to tackling with these two types of performance regressions and revealing the possible existence of latent factors using Rulefit.

7.1 Transient Performance Regression

For diagnosing transient performance regression, we take a step-by-step approach to solve this problem. First, we extract the key factors associated with the anamaly by applying the relative importance analysis. Next, we further narrow down the issue to the specific group of states, hours, or even RNCs by drilling down into the fitting accuracy across different states or RNCs as well as over time. In addition, with the macroscopic information suggested by Rulefit, we need to conduct a more detailed or even lower-level diagnosis on the original data to provide a better understanding of the underlying behavioral change associated with the factors as suggested by Rulefit.

Using the performance regression on Sep.12th as one case study, we gain insights of the possible underlying factors attributing to the performance issue by comparing the important factors generated from Rulefit using TS-D-09/02 and TS-D-09/12, respectively. As summarized in Table 2, 3, the number of bytes starts to play a more critical role in RTT performance, while the number of bytes per flow also replaces the state factor and becomes the most important, both suggesting the possible issue related to the change of traffic load on that day.

To further investigate whether the performance degradation on that day was only observed on particular states or all states, we compare the fitting accuracy across states of these two models. As indicated in Figure 14, it was a problem associcated with a subset of states, as much larger fitting errors are observed in those states. The state index is the same as used in Figure 12. Similarly, the fitting accuracy across hours suggests that the performance issue occurs in all hours, not just for particular hours.

With the heuristics obtained from Rulefit, we speculate that the problem is most likely due to the drastic change of the traffic load in those states. To verify our hypothesis, we compare the number of bytes served on

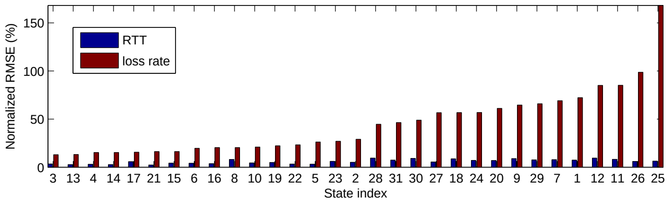

Figure 14: Prediction error across states on Sep.12th.

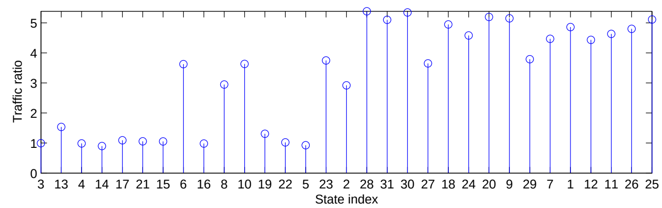

Figure 15: Traffic ratio of Sep.2nd to Sep.12th.

Sep.2nd and Sep.12th for each individual state. Their respective ratio of the number of bytes on Sep.2nd to Sep.12th is shown in Figure 15. As we can see, those states with larger fitting errors also see a drastic drop of traffic on that particular day. Therefore, our assumption is indeed borne out by our investigation of the measurement data itself. A more detailed analysis reveals that the drop of traffic load not only happened to a subset of RNCs within those states, but rather, all RNCs within those states. Moreover, these states are by chance all located on the east coast. Based on these induction, we believe that the degradation of the performance was not attributed to holidays or particular events, which normally incur more traffic than usual. Rather, it could be an issue associated with the network carrier itself, such as the failure of SGSN or GGSN that can have a large impact on all the RNCs on east coast. Similar strategies can also be applied to figure out the underlying reasons leading to the other transient anomalies observed in the weeks after that as shown in Figure 13.

7.2 Persistent Performance Anomalies

To diagnose the performance regression persists within a subset of problematic states as discussed earlier, we intend to adopt the same methodology in diagnosing the transient performance issue. Similarly, we apply Rulefit to the training sets collected in different states. We might have expected that the predictive model will achieve better accuracy if we apply Rulefit to each states individually. Surprisingly, the predictive loss rate performance model generated for those states with poor performance is rather inaccurate, whose explained variance is only around 50% or even lower. On the other hand, the explained variance of the loss rate model improves for those states with good performance. This suggests that the factors presented in our study are not complete enough to fully capture the loss rate behavior. Rather, there may exist possible latent factors that play a critical role but was missing in our feature list. This phenomenon is also partly revealed by our previous observation as shown in Figure 5, where proportions of the "unknown" category of traffic in those problematic states are much larger than the rest of the states. The missing latent factors might be quite correlated with the unknown type of traffic, and therefore suggesting a finer-granularity investigation of this type of traffic by the operators.

Besides the latent factors coming from the traffic itself, there may also exist other latent factors associated with the specific geographical and demographical conditions. Using one of the biggest and busiest states as one example, the numerous tall buildings can lead to very poor performance that are not considered in our model. Besides, the large user density in that state leads to the placement of many microcells, i.e. base stations of smaller sizes and transmitting at a lower power [13]. As a result, users traverse more quickly through different cells. The poor handoff process therefore may also lead to the interuption of service.

The anomaly behavior in the early morning hours on the other hand is not due to its performance regression, but a large performance improvement compared to other daytime hours after 7AM. To get a better understanding of why early morning hours have much better performance, or in other words, why other hours have much worse performance, we similarly apply Rulefit to the training sets collected at different hours. The important factors for one of the daytime hours TS-H-10 are shown in Table 2, and 3. While the important factors for RTT is similar to our generalized model, Bpflow becomes the most important factor during the daytime hours. This is due to the fact that most applications prevalant during those hours have smaller flow sizes, resulting in a large number of smaller flows, and therefore degraded loss rate performance. On the other hand, the flow size become much larger during the early morning hours due to the increase of streaming traffic. Similar to the problem with the states, the model generated for early morning hours also has very poor predictive power, having only around 40% explained variance, whereas the model improves for the rest of the daytime hours. This also suggests the missing latent factors in those hours. The existence of the latent factors both during early morning hours and within the problematic states illustrate the challenges coming from the "complicated" nature of the UMTS network performance. With the help of measurement datasets at a finer granularity, i.e. NodeB or even UE level, we believe the operators can incorporate some of the missing latent factors, and perform more accurate diagnosis using the similar methodologies discussed here.

8. RELATED WORK

There have been quite a few efforts regarding the evaluation or improvement of celluar network performance, such as [18] which studied the important factors affecting the network performance as well as the web browsering performance, and MobiPerf [17] which designed a measurement tool widely installed on mobile end users in order to collect the client side measurement data and provide them with a set of real time network related information, Online tools, e.g., [2, 5], are also available for measuring the performance. However, their measurement studies are all based on client side factors, whereas the datasets in our study were collected inside the carrier, containing more insights of the factors pertaining to the carrier network itself. Besides, unlike previous studies, which either consider the performance indicator for a specific set of applications such as Ping, HTTP, FTP, P2P, VoIP and streaming performance [16, 20, 22, 27], or consider the performance impact from only a few factors [10, 14, 25, 26], we evaluate the RTT and loss rate performance that are critical to most of the applications and try to catch up as many factors as we can in modeling the performance. Other performance evaluation works include a study of the impact of variable 3G wireless link latency and bandwidth on the TCP/IP performance [7], the design of a better handoff decision algorithm using the mobile terminal location and area information [21], performance evaluation of cellular networks using more realistic assumptions [11], and a remodeling of handoff interarrival time [9]. However, they are all simulation based work, and the studies were performed under certain restricted assumptions. Similarly, in [23], authors observed a significant temporal and spatial variations in terms of the network resource usage and subscriber behavior. No systematic approach was proposed to model the performance and understand the underlying important factors. In terms of methodology, a comparative study of various analysis methods of network performance was performed in [19]. However, none of them was applied to real data and verified to be an effective approach in detecting and diagnosing the performance issue. To our best knowledge, our work is the first attempt to identify a large set of factors from both user's and carrier's perspective via a large-scale collection of the datasets inside the carrier. A systematic approach is presented to build a macroscopic model of the cellular network performance, as well as perform prediction, and diagnosis.

9. CONCLUSIONS

In this paper we have studied the RNC-level performance in a large UMTS cellular network, with the goal to provide a "big picture" understanding of the various major factors that may influence the overall network performance across the network and over time. The major contributions and findings of our study are the following. *i) Large scale data collection:* Our study utilizes massive data periodically collected from diverse sources over more than six months within one of the largest UMTS cellular network carriers, whose coverage spans the whole United States. The combination of different datasets provides us a quite comprehensive view of the network performance in the network. The conclusions drawn in our study will therefore not be restricted to any particular group of users or devices. *ii) Identification of a rich set of factors:* As the network performance is not a simple function of a few factors, we identify a rich set of features or factors along different dimensions, including usage patterns, user behavior or application mixes, network coverage, geographical regions, hardware and other factors associated with the network infrastucture. While the features gathered in our study are clearly not exhaustive, new factors can be easily accommodated in our models. *iii) Macroscopic models and relative importance analysis:* The macroscopic models developed in our study have both predictive and interpretative capabilities. They provide network operators with a "big picture" view of the overall network performance and a high-level understanding how various factors may have a differing effect on the network performance across the network or over time. *iv) Unveiling potential latent factors and detecting anomalies:* Our models can also help unveil potential latent factors that are not directly captured in our datasets, or are hard to be captured using either numerical values or non-orderable categorical values. The effect of potential latent factors can be inferred, for example, by comparing the prediction results produced by the models feeding on different (sub)sets of data and by performing the relative importance analysis. Anomalies may also be detected via significant deviations from usual performance predictions.

In a nutshell, the models and insights obtained from our study provide the network operators with a "big picture" understanding of the overall RNC-level RTT and loss rate performance in a large 3G cellular network and the various factors that may influence such performance, thereby the user experiences. Such a "big picture" understanding can help network operators better cope with the management challenges posed by the scale and complexity of a large cellular network. While the paper focuses only on a 3G UMTS network, we believe that our methodology can also be applied to other 3G (and perhaps also 4G) network architectures.

10. REFERENCES

[1] 3gpp specification: 25.331. http://www.3gpp.org/ftp/Specs/html-info/25331.htm.
[2] Consumer broadband test. http://www.broadband.gov/qualitytest/.
[3] Multicollinearity. http://en.wikipedia.org/wiki/Multicollinearity.
[4] Rulefit. http://www-stat.stanford.edu/~jhf/R-RuleFit.html.
[5] Speedtest.net. http://speedtest.net/.
[6] Umts/3g nodeb/bts comparison. http://www.umtsworld.com/technology/ran/ran.htm.
[7] M. Chan and R. Ramjee. Tcp/ip performance over 3g wireless links with rate and delay variation. *Wireless Networks*, 11(1-2):81–97, 2005.
[8] T. Cover, J. Thomas, J. Wiley, et al. *Elements of information theory*, volume 6. Wiley Online Library, 1991.
[9] S. Dharmaraja, K. Trivedi, and D. Logothetis. Performance analysis of cellular networks with generally distributed handoff interarrival times. In *Symposium on Performance Evaluation of Computer and Telecommunication Systems (SPECTS 2002)*, 2002.
[10] J. Fajardo, F. Liberal, and N. Bilbao. Impact of the video slice size on the visual quality for h. 264 over 3g umts services. In *Broadband Communications, Networks, and Systems, 2009. BROADNETS 2009. Sixth International Conference on*, pages 1–8. IEEE, 2009.
[11] Y. Fang. Performance evaluation of wireless cellular networks under more realistic assumptions. *Wireless Communications and Mobile Computing*, 5(8):867–885, 2005.
[12] J. Friedman. Rulefit with r, 2005.
[13] A. Goldsmith. *Wireless communications*. Cambridge Univ Pr, 2005.
[14] A. Haque, M. Kyum, M. Al Sadi, M. Kar, and M. Hossain. Performance analysis of umts cellular network using sectorization based on capacity and coverage. *International Journal*, 2011.
[15] A. Haury, P. Gestraud, and J. Vert. The influence of feature selection methods on accuracy. *Stability and Interpretability of Molecular*, 2011.
[16] T. Hoßfeld, A. Binzenhöfer, M. Fiedler, and K. Tutschku. Measurement and analysis of skype voip traffic in 3g umts systems. *Proc. of IPS-MoMe*, pages 52–61, 2006.
[17] J. Huang, C. Chen, Y. Pei, Z. Wang, Z. Qian, F. Qian, B. Tiwana, Q. Xu, Z. Mao, M. Zhang, et al. Mobiperf: Mobile network measurement system. Technical report, Technical report). University of Michigan and Microsoft Research, 2011.
[18] J. Huang, Q. Xu, B. Tiwana, Z. Mao, M. Zhang, and P. Bahl. Anatomizing application performance differences on smartphones. In *Proceedings of the 8th international conference on Mobile systems, applications, and services*, pages 165–178. ACM, 2010.
[19] P. Lehtimäki. *Data Analysis Methods for Cellular Network Performance Optimization*. PhD thesis, Helsinki University of Technology, 2008.
[20] S. Liu, W. Jiang, and J. Li. Architecture and performance evaluation for p2p application in 3g mobile cellular systems. In *Wireless Communications, Networking and Mobile Computing, 2007. WiCom 2007. International Conference on*, pages 914–917. IEEE, 2007.
[21] A. Markopoulos, P. Pissaris, S. Kyriazakos, and E. Sykas. Cellular network performance analysis: handoff algorithms based on mobile location and area information. *Wireless Personal Communications*, 30(2):97–117, 2004.
[22] B. Orstad and E. Reizer. *End-to-end key performance indicators in cellular networks*. PhD thesis, MasterâĂŹs thesis, Information and Communication Technology, Agder University College, Faculty of Engineering and Science, Grimstad, Norway, 2006.
[23] U. Paul, A. Subramanian, M. Buddhikot, and S. Das. Understanding traffic dynamics in cellular data networks. In *INFOCOM, 2011 Proceedings IEEE*, pages 882–890. IEEE, 2011.
[24] F. Qian, Z. Wang, A. Gerber, Z. Mao, S. Sen, and O. Spatscheck. Characterizing radio resource allocation for 3g networks. In *Proceedings of the 10th annual conference on Internet measurement*, pages 137–150. ACM, 2010.
[25] M. Rossi, L. Scaranari, and M. Zorzi. On the umts rlc parameters setting and their impact on higher layers performance. In *Vehicular Technology Conference, 2003. VTC 2003-Fall. 2003 IEEE 58th*, volume 3, pages 1827–1832. IEEE, 2003.
[26] P. Svoboda, F. Ricciato, W. Keim, and M. Rupp. Measured web performance in gprs, edge, umts and hsdpa with and without caching. In *World of Wireless, Mobile and Multimedia Networks, 2007. WoWMoM 2007. IEEE International Symposium on a*, pages 1–6. IEEE, 2007.
[27] Q. Zhang, W. Zhu, and Y. Zhang. Network-adaptive scalable video streaming over 3g wireless network. In *Image Processing, 2001. Proceedings. 2001 International Conference on*, volume 3, pages 579–582. IEEE, 2001.