# Modeling Web Quality-of-Experience on Cellular Networks

Athula Balachandran†, Vaneet Aggarwal‡, Emir Halepovic‡, Jeffrey Pang‡, Srinivasan Seshan†, Shobha Venkataraman‡, He Yan‡
†Computer Science Department, Carnegie Mellon University, Pittsburgh, PA, USA
‡AT&T Labs – Research, Bedminster, NJ, USA
{abalacha,srini}@cs.cmu.edu, {vaneet,emir,jeffpang,shvenk,yanhe}@research.att.com

## ABSTRACT

Recent studies have shown that web browsing is one of the most prominent cellular applications. It is therefore important for cellular network operators to understand how radio network characteristics (such as signal strength, handovers, load, etc.) influence users' web browsing Quality-of-Experience (web QoE). Understanding the relationship between web QoE and network characteristics is a pre-requisite for cellular network operators to detect when and where degraded network conditions actually impact web QoE. Unfortunately, cellular network operators do not have access to detailed server-side or client-side logs to directly measure web QoE metrics, such as abandonment rate and session length. In this paper, we first devise a machine-learning-based mechanism to infer web QoE metrics from network traces accurately. We then present a large-scale study characterizing the impact of network characteristics on web QoE using a month-long anonymized dataset collected from a major cellular network provider. Our results show that improving signal-to-noise ratio, decreasing load and reducing handovers can improve user experience. We find that web QoE is very sensitive to inter-radio-access-technology (IRAT) handovers. We further find that higher radio data link rate does not necessarily lead to better web QoE. Since many network characteristics are interrelated, we also use machine learning to accurately model the influence of radio network characteristics on user experience metrics. This model can be used by cellular network operators to prioritize the improvement of network factors that most influence web QoE.

## Categories and Subject Descriptors

C.4 [**Performance and Systems**]: measurement techniques, performance attributes; C.2.3 [**Computer System Organization**]: computer communication system—network operations

## Keywords

Cellular Network; Quality of Experience (QoE), Web Browsing, Performance


*MobiCom'14,* September 7-11, 2014, Maui, Hawaii, USA.

http://dx.doi.org/10.1145/2639108.2639137.

## 1. INTRODUCTION

Mobile web data usage is predicted to increase eleven-fold between 2013 and 2018 [1], and web browsing is already one of the most dominant applications on cellular networks [33]. Therefore, it is important for cellular operators to ensure that web browsing sessions provide a better Quality-of-Experience (QoE), i.e., a set of web user experience metrics, such as session length (the number of pages a user clicks through) and abandonment (whether a user leaves a website after visiting the landing page), which we hereafter refer to as *web QoE*. The ability to monitor web QoE is essential to determining when and where degraded network conditions actually impair user experience. Moreover, understanding the relationship between web QoE and radio factors can help troubleshoot such conditions and help operators evaluate the inherent trade-offs in potential solutions. For example, an operator may want to decide whether to increase a cell's transmit power to improve signal strength, or decrease it to reduce handovers from overlapping cells.

Prior work on monitoring web QoE relies heavily on client-side or server-side instrumentation such as browser plugins and server logs. Past work has studied the impact of web page complexity on user experience [18, 24], developing better browsers [28], detecting inefficiencies in HTTP [30] etc. These works have led to best practices that have helped improve website designs, browsers and network protocols. However, the network between the website and the user also plays a critical role in the user experience, particularly in wireless mobile networks. To complement these studies, we take a "cellular operator view" of web QoE. Understanding web QoE from an operator's perspective is more challenging because, unlike other stakeholders, network operators do not have access to detailed client-side or server-side logs, any feedback from the end hosts, or any a priori domain knowledge about website structure. Hence, it is imperative for network operators to accurately estimate QoE metrics using only network measurements.

To complicate matters, websites have evolved from serving relatively static objects, such as hypertext and images, to hosting rich mobile media applications. These sites typically deliver dynamic and personalized content that often includes third-party content such as advertising [18]. Such design typically involves fetching large number of objects from multiple domains and servers. This significant change in web page structure and content over the last decade makes accurate estimation of web QoE metrics from mobile network traces even more challenging. A key challenge is that there is not yet a scalable and accurate method to distinguish between different mobile browsing sessions or to associate each HTTP transaction with a browsing session, by only observing flow-level network traffic. Previous approaches (e.g., [24]) were designed for the "desktop" web and fare poorly when applied to mobile websites.

To the best of our knowledge, this paper presents the first large-scale measurement-driven study that characterizes and models mobile web QoE and relates it to the measurable radio network characteristics, such as radio signal strength, handovers, data rate, etc. To this end, we use a month-long anonymized data set collected from a Tier-1 US-based cellular network and analyze web browsing sessions of 3 leading mobile websites that consistently appear in the top 100 [6]. For this analysis, we design mechanisms to measure and evaluate the two key web QoE metrics that quantify user experience using only network data: session length (number of pages a user clicks through) and abandonment rate (if a user leaves the website after visting the landing page). Moreover, we show that partial download ratio (fraction of page objects that download incomplete) is a measure that likely captures user disatisfaction even for single-click sessions (the majority for some websites), as it correlates well with the other two metrics.

We make the following contributions:

- We design and evaluate a novel technique to reconstruct mobile web sessions and detect user clicks[1] from HTTP traces, and demonstrate that it significantly outperforms current state-of-the-art method. Our approach is based on bag-of-words and Naive Bayes, an approach borrowed from text classification, and extracts features from HTTP headers. It detects clicks on mobile websites with about 20% higher recall and higher precision compared to the previous state-of-the-art [24].
- We quantify the individual impact of various network characteristics on mobile web QoE in the wild, and derive actionable findings for cellular operators. For example, we find that web QoE is very sensitive to inter-radio-access-technology (IRAT) handovers: most sessions with IRAT handovers were abandoned. Somewhat surprisingly, we find that web QoE is not noticeably influenced by the mean radio download or uplink rate (in contrast to mobile video QoE [32]). Moreover, higher radio signal strength (RSSI) does not correlate with higher web QoE, suggesting that web QoE is not power limited. Further, we establish which radio-level metrics are strong indicators of QoE, and which should not be relied upon.
- We capture the complex relationships between the various network parameters and user experience using *intuitive* and *accurate* machine-learning models. Given only radio network characteristics, which are available to network operators even without traffic monitoring, our model can predict the web QoE with accuracy as high as 84%, improving accuracy by 20% compared to the obvious baseline. Network operators can use this model to continuously monitor and improve web QoE by adjusting network parameters.

The rest of the paper is organized as follows. In Section 2, we present the background and details of our data collection process. In Section 3, we discuss related work. In Section 4 we describe and evaluate our approach for estimating user experience metrics from network traces. In Section 5 we present a characterization of how different network parameters affect web browsing experience. We develop a unified web QoE model for web browsing experience and present our findings in Section 6. We conclude in Section 8.

# 2. BACKGROUND

Mobile network users care about the web experience rather than individual network metrics such as throughput and latency. Thus, cellular carriers have a significant interest in using their infrastructure to measure and improve web QoE rather than traditional network metrics, especially when there are trade-offs. To better understand the challenges in measuring web QoE, this section first provides a brief overview of the cellular network architecture, focusing on most relevant aspects for our study, the datasets we use, and the applications of web QoE.

## 2.1 Cellular Network Architecture

A Universal Mobile Telecommunication System (UMTS) is a 3rd Generation (3G) mobile data network, consisting of two major components: a Radio Access Network (RAN) and a Core Network (CN). The RAN includes user equipment (UE), base transceiver stations (i.e., NodeBs), and Radio Network Controllers (RNCs). The CN consists of Serving GPRS Support Nodes (SGSNs) and Gateway GPRS Support Nodes (GGSNs). A UE is a mobile device (smartphone, 3G card, etc.) that connects to the NodeB over the radio channel.

Each base station has multiple antennas (typically 3-6), each of which provides radio coverage for an area called a cell sector, which has a particular frequency and other channel characteristics. The primary cell sector is periodically selected based on the signal strength information, while the UE maintains connections to a set of sectors in range called the active set. The traffic to and from the UE is sent to the corresponding NodeB, via RNC, which controls multiple NodeBs, schedules transmissions, and performs all Radio Resource Control (RRC), such as signaling, handovers, and assignment of Radio Access Bearers (RABs).

Within the CN, an SGSN transfers data between RNC and GGSN on behalf of the UE. A GGSN acts as a packet gateway and router between the cellular network and external networks such as the Internet. A GGSN also maintains IP connectivity between UEs and external IP networks.

## 2.2 Data Collection Apparatus

Mobile operators often collect metrics derived from the traffic that passes through network elements in order to manage the network. For example, radio statistics such as RSSI and handovers are often collected from RNCs and end-to-end throughput and latency metrics are often derived from measurements in the GGSN. This paper is interested in whether such low-level network measurements can be used to measure and understand mobile web QoE.

Thus, for the purposes of this study, we simultaneously collect two anonymized data sets, HTTP transaction records from the interfaces between GGSNs and SGSNs, and radio data from a set of RNCs. The datasets cover a major metropolitan area in the western United States over the duration of one month in 2012.

The HTTP records contain IP flow-level information for web-browsing sessions, and it includes items like client and server IP addresses and TCP ports, flow duration, anonymized device identifier (IMEI), bytes transfered, and TCP flags. Also included are relevant HTTP headers, which include information on URL, user agent, content type, content length, etc. The query parameters in URLs are anonymized via hashing.The radio data contains event-level information for each anonymized user. For example, this data includes RRC measurement reports that periodically report the RSSI, signal to noise ratio, etc. of each UE to the RNC, handover events, RRC throughput utilization, etc. The signal strength and throughput measurements are reported every 2 seconds. Other measurements are reported based on discrete event level data (e.g., when a handover happens, when user connects, disconnects etc.). A full list of events that we use is in Section 5.

Throughout this paper, our analysis focuses on three leading mobile websites (*News*, *Social*, *Wiki*) that consistently appear in the top 100 [6]. Our one month long HTTP trace contains informa-

[1]We use the term *clicks* to refer to mobile "taps" as well as traditional mouse "clicks."

tion on 2 million web sessions to these 3 websites comprising 70 million HTTP requests and around 1 million different UEs. Our radio dataset contains complete information about 100,000 of these sessions.

We emphasize that all the device and user identifiers are anonymized before any analysis is conducted in order to protect privacy. In addition, the outputs of models in this paper are aggregated (e.g., per region and/or network element), so it does not permit the reversal of anonymization or re-identification of users.

## 2.3 Web QoE Applications

Mobile operators monitor network metrics for several purposes, and the ability to monitor web QoE would complement the same applications. First, continuous measurement of metrics permits early detection of network problems. Monitoring web QoE would permit operators to prioritize problems that have the most impact on actual user experience and the understanding of how network factors influence web QoE would help troubleshooting. Second, trending network metrics is invaluable for capacity planning, as they provide objective benchmarks to measure the relationship between investment in infrastructure, such as base stations, and user experience. Third, cellular networks are extremely complex to manage and optimize, involving a huge amount of parameter tuning in the RAN and CN. The adjustment of these parameters often involve implicit trade-offs between different aspects of network performance, such as average capacity vs. peak latency. Monitoring web QoE and understanding how various network factors influence it provide an objective way for operators to perform such parameter optimizations.

# 3. RELATED WORK

**Web traffic modeling:** The most widely used technique to model web traffic and identify web pages from network traces is based on idle time [27]. It has been used extensively for characterizing web traffic in several works [16, 34]. This approach works well for simple static web pages. However, it does not work well for most web pages today since they include dynamic content (shown in Section 4). To overcome this limitation, a page detection algorithm that works for dynamic content was proposed [24]. However, it only identifies clicks resulting in new web pages and does not identify clicks within a page. We propose and evaluate a text classification-based mechanism that has high accuracy in identifying user clicks in Section 4.

**Web Performance Studies:** There have been several efforts made in previous works towards improving web performance. These include developing better browsers specifically for mobile devices [28], techniques to optimize webpages [3, 4], and detecting inefficiencies in HTTP [15, 30]. More recent work has characterized how web site complexity can affect user experience [18]. Unlike these past works, the focus of our work is on understanding the impact of cellular radio characteristics on mobile web browsing sessions with the aim of helping network operators make informed choices on improving web QoE.

**Performance of network protocols over cellular networks:** Past work has also looked at the performance of TCP and HTTP on LTE network highlighting the need to develop more LTE-friendly transport and application protocols [12], characterized the physical and MAC layers in CDMA and its impact on TCP performance [10], studied how large buffers in cellular networks cause TCP queing delays [11]. These efforts have helped understand and improve transport layer and application performance over cellular network, and hence user experience indirectly. In this work understanding the impact of transport layer protocols on user experience is not our immediate focus—the goal of this paper is on understanding how radio network parameters impact user experience.

**Measures of web browsing user experience:** User experience studies in the past have shown that a complete page load time has an impact on user satisfaction [17, 20, 21]. These works are primarily based on controlled studies with few users, and they involve logging page load times and user feedback using client-side instrumentation techniques. However, since network operators do not have access to client-side logs, it is challenging to exactly measure the page load time. However, in our traces we observe that large fraction of the pages are only partially downloaded and we define the partial download ratio metric to capture user experience. Similarly, past work has also extensively used several metrics related to user browsing behavior to quantify user satisfaction including user clicks, dwell time and scrolling [23, 26]. We also use metrics related to user click behavior. However, since we do not have client-side instrumentation, we are unable to capture other behavior such as dwell time and scrolling and incorporate them in our study. Moreover, our work takes a step forward by analyzing the impact of radio network factors on these different user experience metrics.

**QoE in other domains:** Several past efforts study the impact of network factors on user experience and user satisfaction in other applications. Measured impact of bitrate, jitter, and delay on VoIP call duration is used with a machine-learning approach to derive a user satisfaction metric [13]. Past works have employed machine-learning algorithms to develop predictive models for Internet video user engagement [14, 32]. Radio network factors, such as signal strength and handovers, are used to quantify their impact on video viewing experience [32]. Our work focuses on performing a similar analysis for mobile web browsing experience.

# 4. EXTRACTING USER EXPERIENCE METRICS

Network operators cannot access detailed server-side or client-side logs of user browsing patterns. Hence they need to reconstruct web browsing sessions and estimate user experience metrics from network traces. However, over the years, webpages have evolved from serving relatively simple static objects such as hypertext to serving dynamic and even personalized content. This makes it even more challenging to reconstruct mobile web sessions and extract user activities from network traces alone.

Previous work identified *engagement* as a key measure of user experience because more satisfied users tend to stay around longer and use an application more [13, 14, 25, 36]. For web browsing, two central engagement metrics recognized in the web analytics industry are *session length* (i.e., the number of pages a user clicks through) and *abandonment* or bounce rate (i.e., if a user leaves the website after only visiting the landing page) [19]. Unfortunately, both of these metrics necessitate the identification of user *clicks*, which is non-trivial from the perspective of an operator. The difficulty comes from lack of access to client-side or server-slide logs and HTTP records do not readily distinguish requests that are initiated automatically by the browser (e.g., embedded objects) and those that are initiated by user activity.

In Section 4.1, we present and evaluate a novel click detection algorithm based on machine learning, which achieves higher accuracy than the best known approach. Then, in Section 4.1.2, we extract different user experience metrics from the dataset using our algorithm. Based on our findings, we propose *partial download ratio* as a more fine-grain metric that more precisely captures user

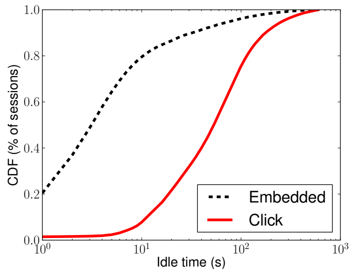



*Figure 1: CDF of arrival time for clicks vs embedded objects*

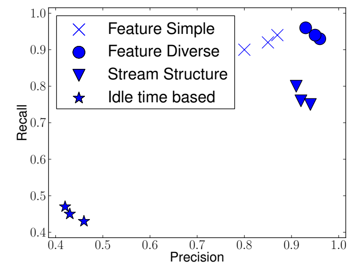



*Figure 2: Our approaches have higher precision and recall compared to previous approaches*

experience impairments due to network conditions, as opposed to user interest.

## 4.1 Detecting Clicks

### *4.1.1 Limitations of Previous Techniques*

The most common approach for differentiating between clicks and embedded objects using network traces is to use the idle time between requests [16, 27, 34]. This approach is based on the assumption that the idle time for requests for embedded objects will be extremely short since they are automatically generated by the browser, whereas requests generated by clicks would typically have higher idle time since they require manual intervention. Therefore, these techniques use a pre-defined threshold and classify a request as embedded object if and only if the idle time is shorter than the threshold. However, we find that in modern mobile web pages, a non-trivial fraction of embedded objects have idle times as long as many user clicks (e.g., requests generated by periodic beacons from third-party analytic services [24]). For example, Figure 1 shows the distribution of arrival times for next click and next embedded objects. This figure is based on web requests from around 50,000 web sessions on the three websites. We labeled each of the web requests in these sessions as clicks or embedded objects manually. An idle time threshold approach would select a point on the x-axis and classify all objects to the left as embedded and those to the right as clicks. We see that there is no idle time threshold that we can select that achieves lower than 20% error on at least one of the two classes.

To improve on this approach, StreamStructure [24] exploits the structure of "desktop" web pages to detect requests for new webpages. However, we show in the next section that it is not as adept at identifying clicks in mobile web pages. Moreover, it is a page detection algorithm that is used to identify clicks resulting in new pages. Other client-side interaction (e.g., clicking to play a video within a page) are not identified by this algorithm.

### *4.1.2 Our Approach*

Our approach to differentiate between clicks and embedded objects is based on our observation that most of the embedded objects are hosted by third party services such as advertising agencies, Content Distribution Networks (CDNs) and analytics services [18]. This opens up an opportunity to distinguish embedded objects from clicks by inspecting request URLs. For example, a request to google-analytics.com is *very likely* to be an embedded object, while a request to news.google.com is *very likely* to be a click. Hence we can employ text based classification algorithms that have been extensively used in other domains (such as spam filtering [31] and sentiment analysis [22]) to classify requests. We would need to learn the classification model separately for each website/domain. In the remainder of the section, we explain four steps in our approach.

**Step 1: Grouping Sessions:** We first filter our sessions to a specific website. We only study traffic originating from web browsers, and hence filter out traffic of native mobile apps using User Agent HTTP header. Similar to the approach in StreamStructure [24], we group requests into different sessions using the anonymized IMEI and Referer header information. The IMEI information helps us separate sessions from different devices. The Referer field identifies the address of an object from which the new request came from. The first request in a session has an empty Referer field. The Referer field is further used to build the request chain within a session. It also helps us separate simultaneous requests from multiple browser instances from the same user equipment.

**Step 2: Extracting Features:** In order to perform text classification, we extract features from the requested URLs. We extract two sets of features to represent the URLs:

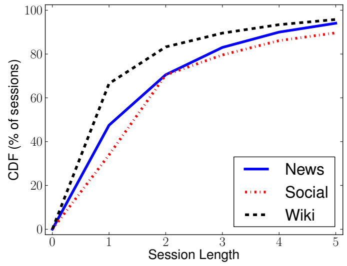
- Feature simple: We extract a bag of words [29] from the domain name. For example, the feature set for the URL www.blog.xyz.com/my/blog/abc.html is <blog, xyz, com>.
- Feature diverse: In addition to domain name features, we include features from the URN and type of content. Hence, the feature set for the above URL would be domain = <blog, xyz, com>, urn = <my, blog, abc.html> and type = html.

**Step 3: Automatic Labeling to Obtain Training Set:** In order to create a classification model to label the URLs as clicks and embedded objects, we need a training set labeled with the ground truth. To obtain the training set, we inspect only the very first 10 seconds of every web session. We assume that only the first request during this time frame was a click and the remaining requests are for embedded objects and collect both feature simple and feature diverse along with the ground truth based on the assumption. We pick 10 seconds because based on the ground truth in Figure 1, almost all user clicks have an idle time of more than 10 seconds and almost 80% of the embedded objects are requested with this time frame. This automatic labeling technique enables our approach to be applied to any website without any manual labeling or ground truth.

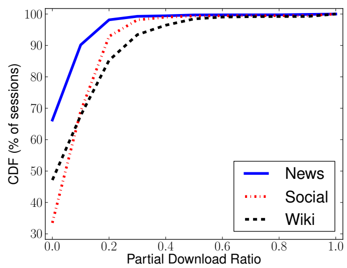
**Step 4: Running Classification Algorithm:** We first learn the classifier model using the training set, and then input the entire dataset and classify each request as the click or embedded object. After testing with multiple machine learning algorithms (such as decision trees, logistic regression, Support Vector Machines [29]), we found that Naive Bayes performs the best compared to other approaches. This is not surprising given that Naive Bayes has been found to perform the best in other text classification problems as well [7].

### *4.1.3 Validation*

To validate our approach, we apply it to the three web sites we study. For each web site, we manually inspect its structure in detail




*(a)* Session Length




*(b)* Partial Download Ratio

*Figure 3: CDF of user experience metrics*

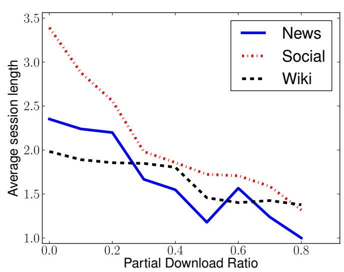



*Figure 4: Session length decreases with increasing partial download ratio*

in order to label each HTTP request as either a click or a as an embedded object. We manually label one day of trace data with roughly 50,000 sessions.

We then compare the idle-time threshold based approach, StreamStructure and our approach (both using Feature simple and Feature diverse) and then estimate the performance in terms of both precision and recall. Precision is defined as the number of correct clicks identified by the total number of clicks identified, and recall is defined as the number of correct clicks identified by the total number of clicks. Figure 2 shows the precision and recall using each the three websites. Our Feature simple and Feature diverse have higher recall than the previous baselines. Feature diverse has higher precision than Feature simple because some embedded objects are hosted on the main domain. Feature simple will incorrectly classify this object as a click since it just uses features from the domain name. In the remainder of this paper, we hence use the Feature diverse approach to identify clicks.

## 4.2 Measuring User Experience

As described earlier, session length and abandonment rate are two important metrics that the web industry recognizes as representative of user engagement. Our click detection algorithm presented above enables operators to estimate session length and abandonment rate using only HTTP traces collected from the network. However, session length and abandonment rate are relatively coarse engagement metrics because they are also influenced by user interest, which is a confounding factor that is difficult to measure for both network and website operators. Indeed, we find that many web sessions are only one click (and thus, by definition, abandoned). These metrics do little to distinguish satisfied and dissatisfied users of these single-click sessions. In this section, we show that *partial download ratio*, i.e., the fraction of HTTP objects that are not completely downloaded in a session, is strongly correlated with session length and abandonment rate, so we can use it as a proxy to estimate user experience, even for sessions lasting a single click.

To better understand the session length, abandonment rate, and partial download ratio metrics, we extract web sessions and estimate number of clicks for each session for the three different websites from the entire 1 month HTTP record trace. Figure 3(a) shows the distribution of session lengths. We observe that all sessions have length less than 10 on all the three websites. A significant fraction the sessions on all the three websites have a length of 1 (47% for News, 33% for Social, 67% for Wiki). The overall abandonment rate is 35% for the three websites. These observations highlight the need for a user engagement metric that can highlight network problems in sessions of length one.

One such candidate measure in common use is the web page load time (i.e., the time it takes to load a page). However, it is known that web page load time is difficult to measure from HTTP traces because the traces do not capture the browser's rendering pipeline [35]. Moreover, without a priori knowledge of web page structure, operators can not easily distinguish complete vs. incomplete page loads. Therefore, naïvely using download time to approximate page load time would incorrectly suggest that abandoned pages have low load time.

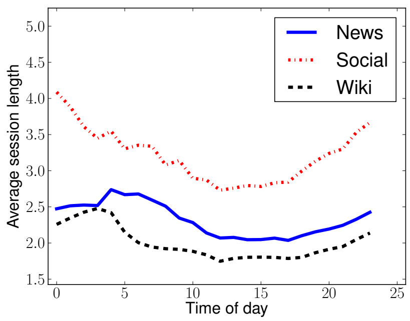
Instead, we propose that partial download ratio is a useful proxy metric for user engagement. Figure 4 shows the average session length as a function of the partial download ratio. We see that there is roughly a negative linear relationship between the partial download ratio and session length, supporting our hypothesis that users are less engaged when more objects on the page fail to load completely (or do not load completely before the user moves on). The linear coefficient is different for each website, as website design likely influences how much partially downloaded content effects the user experience, but the coefficient can be easily learned using regression. For example, using a linear fit to determine session length in terms of partial download ratio, the partial download ratio co-effieceints for the News, Social, and Wiki websites are -1.6, -2.36, -0.85 respectively. Figure 3(b) shows the distribution of partial download ratios for each session. We also observe that over 60% of the sessions have objects that are partially downloaded on each website.

Figure 5 shows the average session length, abandonment rate, and partial download ratio by time of day. We observe strong temporal patterns in the user engagement metrics. Lower session lengths, higher abandonment and higher partial download ratio occur during peak hours (10 am - 6pm) compared to the rest of the day.

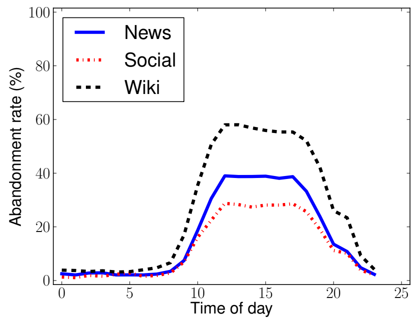
Corresponding to the different linear coefficients we see in Figure 4, we observe that the web QoE metrics are different across different websites. This is likely because, as previous work has showed [18], user experience is dependent on factors other than network quality, such as how mobile-friendly the website is, the number of objects, type of objects etc.

# 5. ANALYZING NETWORK FACTORS

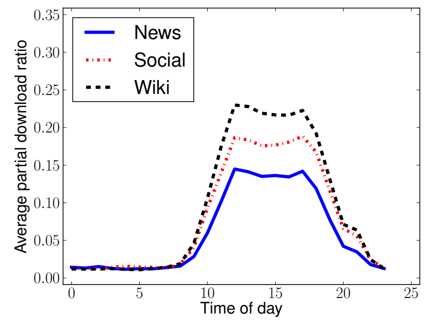
Our first goal is to understand the relationships between individual network factors and web QoE, with the end goal of building




*(a)* Session Length

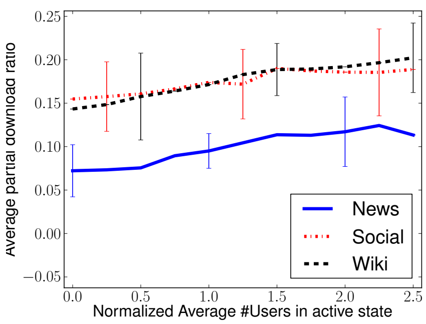



*(b)* Abandonment Rate

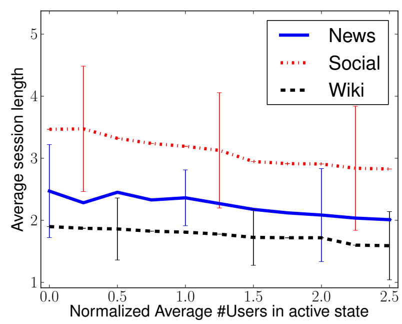



*(c)* Partial Download Ratio

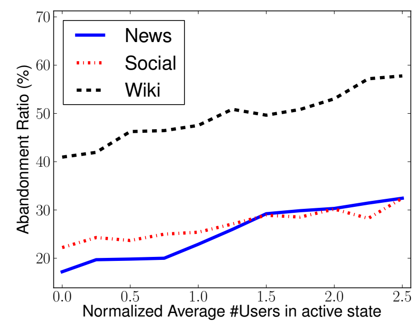
*Figure 5: Time of day effects on the experience metrics*




*(a)* Partial Download Ratio




*(b)* Session Length

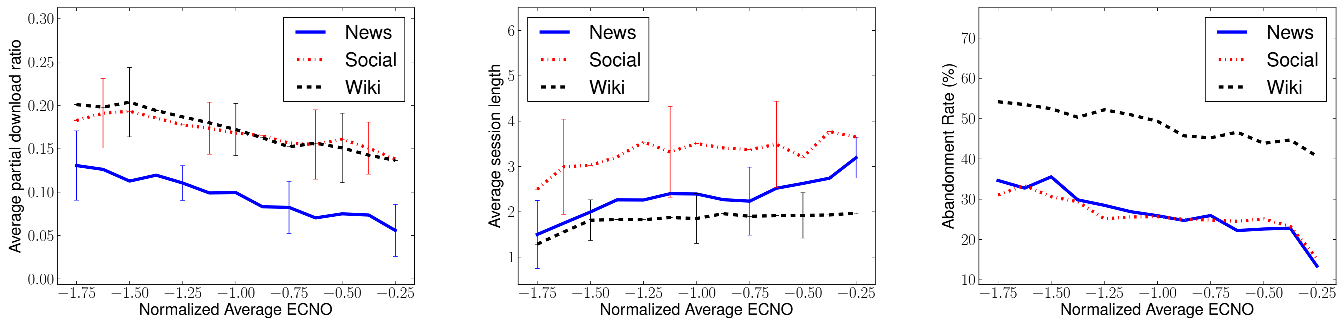



*(c)* Abandonment Rate

*Figure 6: Higher load in the cell (measured in terms of number of active users) leads to worse web QoE. Session length has higher variance since it is a more "noisier" metric as explained in Section 4.2*

models that can be used to improve web QoE by tuning various network parameters. We first itemize all radio network factors that we study, which may effect web QoE:

- *Number of soft handovers (SOHO):* A soft handover occurs when a cell sector is added or removed from the set of cell sectors that a UE is connected to. A SOHO is a "make-before-break" handover in that the link to the new sector is established before an old link is removed. From radio network data that we collected from the RNCs, we count the total number of soft handovers during a session.
- *Number of inter-frequency handovers (IFHO):* An IFHO occurs when the UE switches to a cell sector that is operating on a different frequency. An IFHO is a "break-before-make" handover because a device can only listen on one frequency at a time; thus, it must break the link with all old sectors before establishing the link with the new one. We count the number of inter-frequency handovers during a web session.
- *Number of inter-radio access technology (IRAT) handovers:* An IRAT handover happens when a UE switches between different radio access technologies (e.g., UMTS to GPRS or EDGE). These do not include handovers to and from WiFi since our data collection apparatus does not capture such handovers. An IRAT handover is also a "break-before-make" handover because the device must disconnect entirely from the current radio network before connecting to the new one. This process involves a significant amount of network signaling and can take several seconds.
- *Number of admission control failures (ACF):* We count the number of times the UE fails to complete the admission control procedure during the web browsing session. These events mostly occur when the radio network is overloaded.
- *Number of RRC failures (RRC):* An RRC failure occurs if the RNC is overloaded and it cannot allocate a request from the UE for more radio resources. We count the number of RRC failures within a web session.
- *Average Received Signal Code Power (RSCP):* This is the downlink power received by the UE receiver on the pilot channel. It is measured in dBm.
- *Average received energy per chip of the pilot channel over the noise power density (ECNO):* It is expressed in dB and it measures how well a signal can be distinguished from the noise in a cell. It is measured in dB. Note that ECNO is measured on the pilot channel and thus may be different from the SINR of the traffic channel.
- *Average received Signal Strength Indicator (RSSI):* Expressed in dBm, it is the wide-band received power within the relevant channel bandwidth. It is related to RSCP and ECNO as follows: RSSI = RSCP - ECNO. Note that RSSI is measured on the pilot channel and thus may be different from the received power of the signal on the traffic channel.
- *Average uplink and downlink radio data throughput:* We compute the average uplink and downlink data rates for the UE when it is in active state during the web session in Kbps. Note that the radio data rate is not equivalent to the long-term through-




*(a)* Partial Download Ratio *(b)* Session Length *(c)* Abandonment Rate

*Figure 7: Higher signal energy to interference (ECNO) leads to better web QoE*

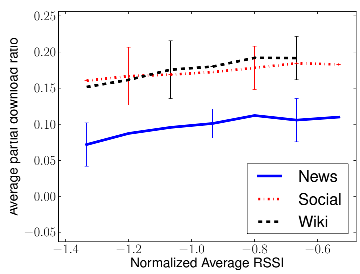



*Figure 8: Surprisingly, higher received signal strength leads to higher partial download ratio*

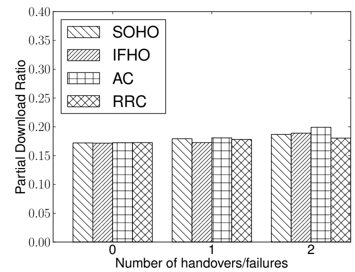



*Figure 10: The impact of soft handovers, inter-frequency handovers, access control failures and RRC failures on web QoE is minimal*

put because it is only measured when the device is scheduled to send/receive data (the radio link is time and code division multiplexed). The radio data rate does not count the non-scheduled time slots in the denominator of the throughput calculation. The number of users in active state (see below) serves as an estimate of the number of competing flows, as a sector schedules each user in a proportionally fair manner.

- *Average number of users in active state:* We measure the number of active users served by each cell at a minute-level granularity. Using this information we compute the average number of users that are served by the cell that the UE is connected to during the web session. This is an indication of the load in the cell.

We report the normalized the value of RSCP, ECNO, RSSI, average uplink and downlink throughput and number of users in active state as a fraction of the mean of the metric. For example, instead of reporting the absolute value of RSSI, we report RSSI/mean(RSSI) and hence the plots can be read as x% above or below the average.

## 5.1 How network factors impact web QoE

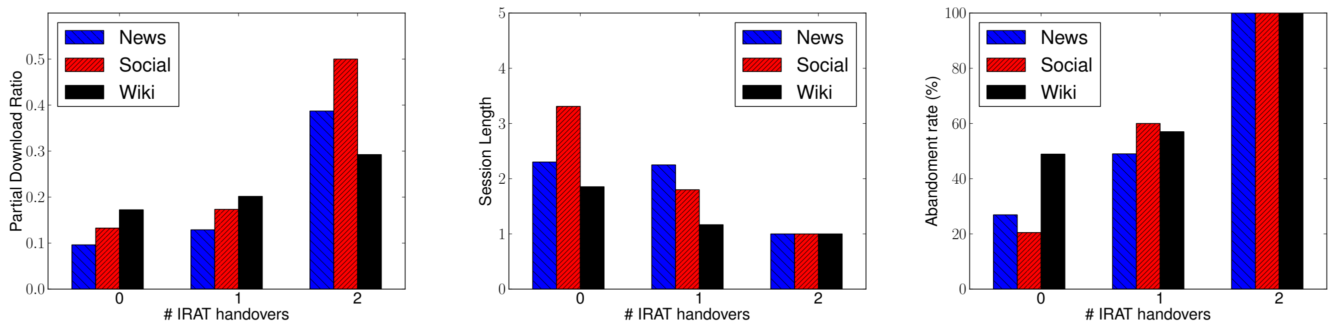
To understand how each network factor individually affects web QoE metrics, we plot web QoE metrics against measured values of network factors from our radio data. The main takeaways are as follows:

***1. Higher network load results in worse web QoE.*** Number of users in active state in a cell is an indication of load in the network. As Figure 6 shows, there is a linear relationship between the load and various web QoE metrics. For instance, adding 25% more users than the average can increase abandonment by 2 full percentage points. Increasing cell load also leads to lower session lengths and higher number of partially downloaded objects on average. This relationship between load and web QoE metrics holds even when conditioned by time-of-day, though cell load is significantly higher during peak hours (not shown due to space constraints). The results suggest that cellular network operators can improve web QoE by decreasing cell load by deploying more radio cells or re-distributing users across cells.

***2. Higher signal strength (RSSI) does not necessarily correlate with better user experience, but higher signal energy to interference (ECNO) does.*** As Figure 7 shows, increasing the signal energy to interference (ECNO) by 10% above average reduces abandonment rate by about 2 percentage points, increases average session length between 2.6% and 9.4% and improves partial download ratio by 0.7 percentage points. In contrast, Figure 8 shows that sessions with higher RSSI have higher partial download ratio on average.

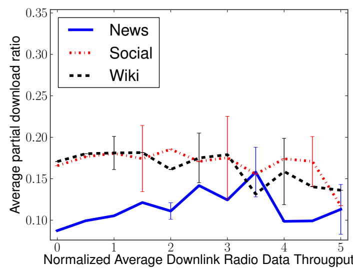
These results confirm that, similar to recent WiFi findings, ECNO (an analogous measure to the SINR of WiFi beacons) is a better indicator of channel quality than RSSI because RSSI does not exclude the power of noise and interference. This finding suggests that web QoE is interference and noise limited, not power (i.e., coverage) limited. We did not observe any impact of RSCP on user experience metrics (not shown).

***3. IRAT handovers lead to worse web QoE.*** IRAT handovers had the strongest impact on user experience, as seen in Figure 9. Sessions with IRAT handovers are much shorter than those without IRAT handovers. Also, all sessions with more than 1 IRAT handover were abandoned. The impact of other handovers (soft handovers, inter-frequency handovers) and failure events (access control failure, RRC failure) on web QoE were negligible. Figure 10 shows that increasing number of such handovers and failures leads to minimal increase in partial download ratio. This indicates high

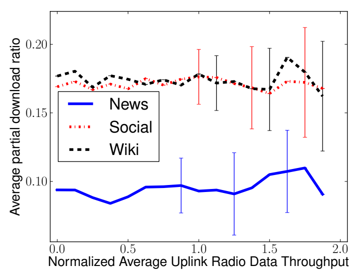



*(a)* Partial Download Ratio *(b)* Session Length *(c)* Abandonment Rate

*Figure 9: IRAT handovers have a strong impact on web QoE—all sessions with 2 handovers are abandoned.*




*(a)* Downlink




*(b)* Uplink

*Figure 11: Radio data link rate does not impact partial download ratio*

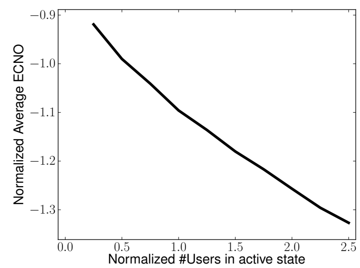
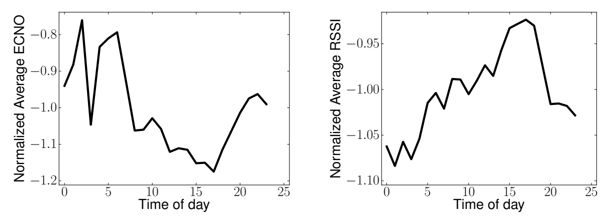



*Figure 12: Number of users vs ECNO*




*(a)* ECNO *(b)* RSSI

*Figure 13: Time of day effect on signal strength parameters*

robustness against these types of events, hence they should not be used to assess web QoE and their management would likely yield insignificant improvement.

***4. Higher radio data rate does not necessarily lead to better web QoE.*** Figure 11 shows the impact of radio data rate on partial download ratio. As web objects are primarily downloaded onto the mobile device, we start by looking at the downlink direction and find that higher data rates do not improve partial download ratio (Figure 11a). As expected, uplink data rate shows no impact (Figure 11b). We find similar relationship between data link rates and other web QoE metrics (not shown). While it may not be intuitive that data rate and web QoE metrics have weak relationship, it has been shown that web browsing traffic is more latency-limited than throughput-limited [5, 9].

## 5.2 Analysis on Other Websites

To test whether the observations we made above hold for other websites, we analyze one day's worth of HTTP records and radio data for five other leading mobile websites (*Shopping*, *Marketplace*, *News2*, *Social News* and *Blog*) that consistently appear in the top 100 [6].

In Table 2, we characterize these websites based on two key metrics of website complexity that past work has identified [18]—namely, (1) the average number of objects requested per click and (2) the average number of domains from which requests are served. We found that these websites represent a varied range both for complexity and for user behavior. For example, *News*, *News2*, *Social* and *Blog* have the highest complexity these metrics, whereas

| # IRAT handovers | Normalized ECNO | Normalized RSSI |
|---|---|---|
| 0 | -1.09 | -0.97 |
| 1 | -1.53 | -1.15 |
| 2 | -1.84 | -1.21 |

*Table 1: We observed lower average ECNO and RSSI for sessions with IRAT handovers*

*Shopping* and *Marketplace* are less complex. Moreover, users tend to have different browsing behavior on these websites: *Shopping*, *Marketplace*, and *Social News* sites understandably tend to have higher session lengths, while *Wiki* and *Blog* tend to have low session lengths.

Table 2 shows that our result hold for this varied set by correlating the impact of increasing each of the radio network factors on partial download ratio. For each radio network factor (e.g., RSSI, ECNO, etc.), we tabulate the slope of the partial download ratio vs. the network factor. For each of RSSI, ECNO, # IRAT handovers and # Users, the partial download ratio graphs exhibit the same trend on increasing the radio network factor even across the varied set of websites. For example, increasing ECNO decreases partial download ratio (i.e., negative slope across all sites).

## 5.3 Comparison with Other Mobile Applications

Our findings that higher load (in number of users) and lower signal to noise ratio (ECNO) correlate with lower web QoE is not entirely surprising and confirms previous findings on the relationship between these network factors and QoE of mobile video streaming [32]. Interestingly, as in the case of video streaming, the relationship between these two network factors and abandonment rate are both linear and have roughly the same slope.

In contrast to findings on video streaming, however, we observed that only IRAT handovers were disruptive to web QoE and that web QoE metrics were uncorrelated with SOHOs and IFHOs. This finding suggests that web browsing is more tolerant to minor handover disruptions than video streaming. IRAT handovers are much more disruptive because changing radio technologies can take several seconds to complete, which is long enough to influence user perceived latency. Moreover, we find that, unlike mobile video streaming, the radio data rate is uncorrelated with web QoE metrics. This may be because video streaming is a more bandwidth intensive application, whereas web browsing is more sensitive to latency.

In summary, our findings complement previous work on cellular mobile video [32], demonstrating that reducing load and improving ECNO are equally important for both applications. However, carriers need not optimize handovers (except IRAT) or radio throughput rates if they only want to improve web QoE.

## 5.4 Dependencies and Other Factors

We found that many network factors under study are not independent of each other. An obvious example is that RSSI is related to ECNO and RSCP (as we mentioned earlier). We also found several other dependencies between the radio network factors. Some examples are:

- The number of users in active state in a cell and ECNO are dependent on each other [8]. As shown in Figure 12, there is a linear relationship between the two—adding more users into the cell steadily decreases ECNO.
- Table 1 shows that sessions that experience IRAT handovers also experience lower signal strength (RSSI) and lower signal energy to interference (ECNO).

Further analyzing the radio network factors, we also observed significant time of day effects. Figure 13 shows the average value of RSSI and ECNO observed per hour of the day over the entire one month dataset. We observe that average signal strength to interference (ECNO) is lower during peak hours compared to non-peak hours. On the other hand, average signal strength (RSSI) is higher during peak hours compared to non-peak hours. In Section 4, we also observed strong temporal effects on the various user experience metrics (Figure 5). These could also be caused by external factors/reasons—for example, users are less likely to engage in long browsing sessions during working hours pointing to the need for including external factors such as time of day into the analysis.

| Model | Avg. Accuracy (%) |
|---|---|
| Radio factors alone | 73.02 |
| Radio factors + time of day | 79.25 |
| Radio factors + time of day + website | 83.95 |

*Table 3: Adding time of day and learning a separate decision tree for each website improves accuracy.*

In summary, complex interdependencies between network factors as well as external factors (e.g. time of day) make it very challenging to understand and quantify the true impact of each network factor using correlation analysis. This points to the need to use more systematic techniques, including machine learning algorithms, to capture the complex relationships in order to quantify the impact of network factors.

# 6. MODELING WEB QOE

Our end goal is to develop models that can be used by cellular network operators to improve web QoE by tuning network parameters. For example, our model should help answer the question "how much can we improve the partial download ratio if we increase ECNO by 1 dB?". The model should also help network operators monitor web QoE metrics using only radio network characteristics. To achieve this goal, the QoE models that we build should be *intuitive*, *accurate*, and must be able to *predict web QoE from network factors alone*.

Building an accurate QoE model is challenging because of the complex relationships between network factors and web QoE metrics, interdependencies between various network factors, and also due to external factors (e.g., differences between websites, time of day effects). To tackle these challenges, we use machine learning to capture the dependencies and relationships, and develop and evaluate models that can predict web QoE metrics. The model we derive will express web QoE metrics as a function of radio parameters; specifically, we wish to capture the relationship:

$$WebQoE = f(RadioNetworkParameter_{1..n})$$

where $WebQoE$ denotes one of the user experience metrics (partial download ratio, abandonment, or session length), and $RadioNetworkParameter_i$ denotes the $i$th observed radio network parameter listed in Section 5.

## 6.1 Evaluation

We predict: (1) partial download ratio, (2) session length, (3) if the session includes partially downloaded pages or not (*part* or *full*), and (4) if the user will abandon a session or not (*abandoned* or *not-abandoned*). The choice of the machine learning algorithm is important because the model it learns should be expressive enough to capture all the complex relationships and dependencies. After experimenting with different regression, tree and bayes algorithms (such as linear and logistic regression, variations of decision trees and naive Bayes) we found that linear regression worked best for (1) and (2), and C4.5 decision tree algorithm was able to predict the binary classification most accurately for tasks (3) and (4). We use 10-fold cross-validation to evaluate all our models [29].

**Evaluating Linear Regression Models:** Since most network factors have a linear relationship with web QoE metrics (session length and partial download ratio) and since they are linearly dependent on

| | Website complexity | | Average webQoE | | | Impact of increasing radio factor on Partial Download Ratio (P.D.R.) | | | |
|---|---|---|---|---|---|---|---|---|---|
| Website | # Domains | # Objects | Average P.D.R | Average Session Length | Abandon Rate (%) | RSSI | ECNO | # IRAT | # Users |
| News | 13.4 | 23.1 | 0.1 | 2.2 | 21 | 0.0006 | -0.007 | 0.15 | 0.0004 |
| Social | 13.1 | 20.2 | 0.15 | 3.1 | 23 | 0.0004 | -0.004 | 0.06 | 0.0003 |
| Wiki | 3.69 | 13.31 | 0.16 | 1.8 | 41 | 0.0011 | -0.005 | 0.18 | 0.0003 |
| Shopping | 4.6 | 7.5 | 0.12 | 8.5 | 10 | 0.0015 | -0.004 | 0.07 | 0.0003 |
| Marketplace | 3.2 | 3.6 | 0.04 | 12.3 | 5 | 0.0007 | -0.001 | 0.06 | 0.0002 |
| News2 | 15.7 | 29.9 | 0.09 | 3.4 | 15 | 0.0010 | -0.008 | 0.17 | 0.0004 |
| Social News | 11.65 | 7.38 | 0.03 | 10.9 | 8 | 0.002 | -0.005 | 0.12 | 0.0003 |
| Blog | 11.89 | 20.40 | 0.17 | 2.65 | 30 | 0.0003 | -0.009 | 0.19 | 0.0005 |

*Table 2: Observations made in Section 5.1 hold for a varied set of websites*

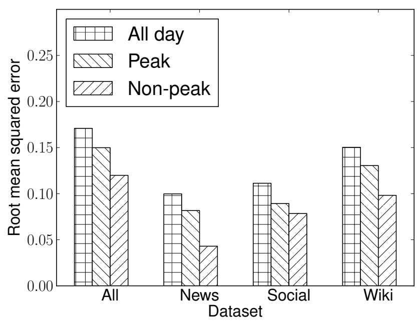
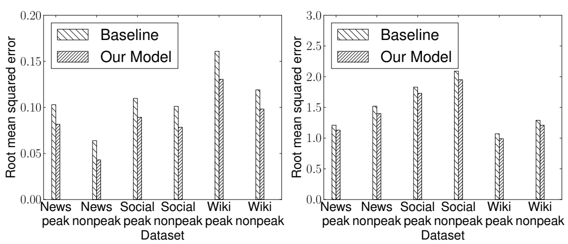



*Figure 14: Learning a separate regression models for each website and time of day (peak/non-peak) improves accuracy.*




*(a)* Partial Download Ratio *(b)* Session Length

*Figure 15: Our models are more accurate than the baseline in predicting partial download ratio.*

each other (Section 4), linear regression is well-suited for capturing these relationships. To measure the "goodness-of-fit" of linear regression, we use the standard measure of root mean squared error (RMSE), where lower RMSE values indicate better prediction. Figure 14 shows RMSE when using linear regression to predict partial download ratio. We show the results for each website separately on the x-axis, as well as the results for data from all web sites taken together ("*All*"). The three bars in each cluster represent RMSE for the entire day, for peak hours only, and for non-peak hours only—each of these have separately learned models due to significant time of day effects. We see that separate models for time of day and individual web sites results in significant prediction improvement, as indicated by lower RMSE values.

We repeated these experiments for session length and found similar improvements from splitting data and learning models for each split (not shown). Hence, our final linear regression models for both partial download ratio and session length are *each* a collection of six linear regression models: one each for each combination of web site (*News*, *Social*, *Wiki*) and time of day (*Peak*, *Non-peak*).

We compare the performance of our models with a baseline model that always predicts the mean (using the ZeroR classifier [2]) for the particular dataset in Figure 15. In the case of partial download ratio, our model has around 20% lower RMSE compared to the baseline. For session length, our model has up to 10% lower RMSE. Predicting session length is not as accurate as predicting partial download ratio because session length is more affected by external confounding factors (e.g. user interest), which are very difficult to capture from network traces.

**Evaluating Decision Tree Models:** We also develop models that make binary predictions for users' web sessions, such as "will this session have partially downloaded pages?" and "will the user abandon this session?". The C4.5 decision tree algorithm performed the best in predicting both these classifications. Table 3 shows the accuracy for predicting partial downloads. Refining the model by inputting the time of day (in terms of hour of day (0-23)) along with learning a separate model for each website led to around 11% improvement in accuracy. We observed this for the decision tree that predicts abandonment as well. Essentially, our models for predicting abandonment and partial downloads are both a collection of 3 decision trees, one for each website. They take both the radio factors as well as time of day as input.

Further, we compared our model against a baseline model that predicts the majority class using the ZeroR classifier [2] for each dataset in Table 4. Our partial download model achieves up to 20% higher accuracy compared to the baseline while our abandonment model achieves up to 10% more accuracy. Again, smaller improvement for abandonment is due to confounding factors like user interest that we cannot measure, but can significantly influence user abandonment.

| | Partial | | Abandonment | |
|---|---|---|---|---|
| Dataset | Model | Baseline | Model | Baseline |
| News | 80.4 | 59.7 | 78.6 | 75.6 |
| Social | 87.7 | 72.5 | 82.0 | 78.6 |
| Wiki | 80.6 | 70.3 | 62.3 | 53.3 |

*Table 4: Our models are more accurate than the baseline in predicting partial downloads and abandonment.*

## 6.2 Insights and Discussion

Fortunately, both linear regression and decision tree algorithms that gave the highest accuracy also generate very intuitive models. They can hence provide insights to network operators towards tuning network factors to improve web QoE. For example, Table 5 shows the regression co-efficients for our partial download ratio model. We observe that the features that the models learnt (number

| Dataset | # active users | RSSI (dBm) | ECNO (dB) | # SOHO | # IRAT | Constant |
|---|---|---|---|---|---|---|
| News Nonpeak | 0.0002 | 0.0005 | -0.0043 | 0 | 0 | 0.0411 |
| News Peak | 0.0002 | 0 | -0.0032 | 0 | 0 | 0.0976 |
| Social Nonpeak | 0.0002 | 0.0005 | -0.0037 | -0.0007 | 0.0639 | 0.0485 |
| Social Peak | 0.0002 | 0.0005 | -0.0047 | -0.0005 | 0.0627 | 0.1367 |
| Wiki Nonpeak | 0.0002 | 0.0003 | -0.0042 | -0.0005 | 0.0871 | 0.0848 |
| Wiki Peak | 0.0002 | 0.0004 | -0.0037 | -0.0004 | 0.0799 | 0.2022 |

*Table 5: Linear regression coefficients of the model that predicts partial download ratio.*

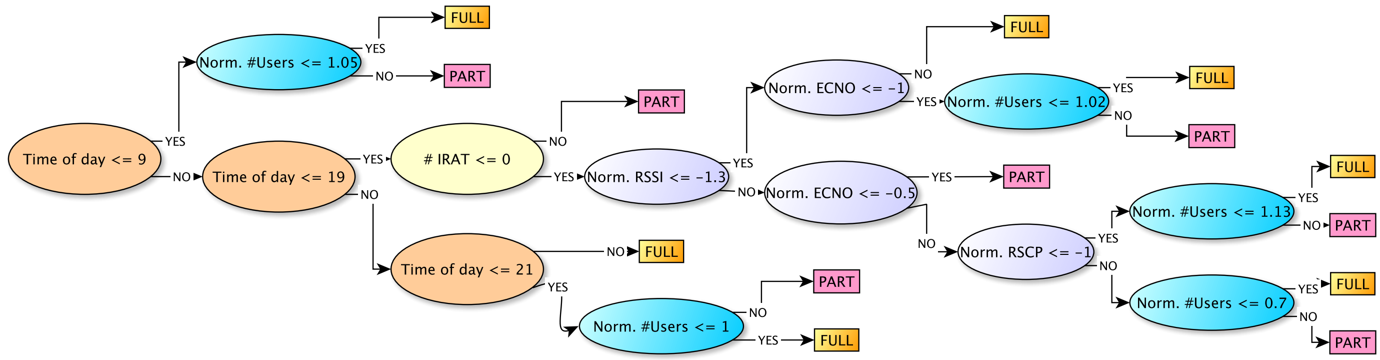



*Figure 16: Pruned decision tree that predicts partial downloads*

of users, ECNO, RSSI etc.) are the same as those that we found to be impactful in Section 5. Moreover, the model also ignores factors such as downlink and uplink throughput that we found to not have an impact on web QoE metrics.

Interestingly, the value of the regression co-efficients are similar across the different datasets. This implies that irrespective of the time of day and the website, tuning a particular network parameter has the same impact on partial download ratio. For example, improving ECNO by 1 dB decreases partial download ratio by roughly 0.004 across all times of day and websites. Network operators can hence use this model to understand the true impact of a parameter. For example, comparing the co-efficients, decreasing IRAT handovers and improving ECNO has the highest impact on improving partial download ratio. We also found similar conclusions from analyzing the regression co-efficients for session length (not shown due to space constraints).

Figure 16 shows the pruned decision tree that we learnt for predicting partial download for the *Wiki* website. Again, consistent with our analysis in Section 5, the model picks parameters such as Number of users, ECNO, IRAT etc. to branch on, reconfirming the impact of these factors. Further, the decision tree rules separate the data based on time of day into a similar classification that we made for peak vs. non-peak (e.g., Time of day $<= 9$, Time of day $> 9$ and Time of day $<= 19$, Time of day $> 19$). We also observe that the feature splits conform with several of our observations. For example, during non-peak hours the partial downloads are lower (if Time of day $> 21$, predict *full*). Also if load is higher partial downloads are higher (if Normalized Num User $<= 1.05$, predict *full* otherwise *part*).

# 7. DISCUSSION

**Other websites and native apps:** In this paper, we developed and analyzed QoE models primarily for three different websites. However, preliminary analysis in Section 5.2 indicates that the general observations also hold for other websites with different complexities. We also found that (1) the same machine learning algorithms (decision trees and linear regression) performed the best across different websites, (2) the model parameters are similar across websites, and (3) the *ordering* of the importance of network parameters is the same across different websites. Furthermore, our methodology is completely automated and can be easily applied to other websites. In this paper we focused on traffic from web browsers by filtering based on the User Agent string; this analysis does not include traffic from native mobile apps, and applying our techniques to this type of traffic is an interesting direction for future work.

**New technologies:** Our study is based on traffic collected from a UMTS network. However, we expect our methodology and results to generally apply to newer technologies such as 4G Long Term Evolution (LTE). Radio network parameters in UMTS have analogous parameters in LTE. For example, the number of users in active state in UMTS is related to the number of users in CONNECTED state in LTE. Similarly, number of RRC failures in LTE is related to the number of RRC failures in UMTS. We observed that the number of IRAT handovers has the highest impact on user experience. This for instance might have a lesser impact when switching between UMTS and LTE since the handovers are expected to complete faster. Similarly since web browsing is more latency-limited than throughput-limited, higher throughput offered by LTE may not have a significant impact on web QoE.

**Encrypted traffic:** Our click detection algorithm uses URL to classify if a web request is a click or an embedded object. With SSL/TLS traffic, the entire user payload including HTTP URL is encrypted and hence we cannot apply the current methodology. However, a network operator can still identify the domain to which an encrypted web request is sent by correlating the IP address of the web request from DNS server logs—the IP address will likely correspond to a prior DNS request/reply pair. Our click detection technique can be tailored to use this information in the future. Also, encryption would not affect our QoE prediction methodology, since it is based on radio statistics and completely independent of traffic encryption.

**Limitations:** One of the main constraints faced by a network operator is the lack of client-side instrumentation. This makes it difficult to differentiate between the abandonment caused by the lack of user interest from the ones caused by network issues. For example, a user could potentially have abandoned the session due to lack of interest, and yet the network would have delivered all the data. It is impossible to identify such a situation from network logs alone. Similarly, network operators cannot identify cellular-to-WiFi handovers from cellular network traces alone, and would mistakingly mark such handovers as abandonments. Nonetheless, operators are typically interested in aggregate performance trends and changes that signify network issues or improvements. A few false positives or negatives introduced by these limitations are unlikely to significantly alter the behavior of aggregate metrics.

# 8. CONCLUSION

In this paper, we presented a large-scale study that analyzed web QoE, such as session length, abandonment rate and partial download ratio, from a "cellular network operator" point of view. Understanding web QoE from a network operator perspective is challenging due to lack of visibility or instrumentation at clients and servers and a priori knowledge of web site structure. We developed and evaluated text-classification-based mechanisms to extract mobile web browsing sessions and accurately estimate various web QoE metrics from network traces. Our classification approach has 20% higher precision than previous state-of-the-art. Further, we analyzed the impact of various radio network factors on web QoE. We observed that web QoE is particularly sensitive to IRAT handovers, ECNO and load in the cell. Moreover, we identified radio metrics that are often considered important, but show negligible impact on web QoE, such as average radio link data rate, soft handovers, and inter-frequency handovers. Finally, we developed *accurate* and *intuitive* machine learning models that can predict various web QoE metrics from *network factors alone*. Our models can be used by network operators to monitor web QoE using standard radio network metrics alone and prioritize improvement of network factors that have the highest impact.

## Acknowledgments

We thank our anonymous shepherd and reviewers for their feedback that helped improve our paper. This work is supported in part by NSF FIA Grant CNS-1040801.

# 9. REFERENCES

[1] Cisco visual networking index: Global mobile data forecast update 2013-2018. `http://goo.gl/fjHsga`.

[2] Data mining with weka. `http://goo.gl/YD9Awt`.

[3] Firebug. `getfirebug.com`.

[4] Google page speed. `http://goo.gl/p1biC0`.

[5] Impatient Web Users. `http://goo.gl/4jPDkv`.

[6] Quantcast website ranking. `http://goo.gl/QH7te2`.

[7] Text classification. `http://goo.gl/iN6syK`.

[8] Wcdma - network planning and optimization. `http://goo.gl/Alivjh`.

[9] Using Predictive Prefetching to Improve W.W.W. Latency. In *SIGCOMM Computer Communivation Review*, 1996.

[10] Experiences in a 3G Network: Interplay between the Wireless Channel and Applications. In *MOBICOM*, 2008.

[11] Tackling bufferbloat in 3G/4G networks. In *IMC*, 2012.

[12] An In-depth Study of LTE: Effect of Network Protocol and Application Behavior on Performance. In *SIGCOMM*, 2013.

[13] K. Chen, C. Huang, P. Huang, C. Lei. Quantifying Skype User Satisfaction. In *Proc. SIGCOMM*, 2006.

[14] A. Balachandran, V. Sekar, A. Akella, S. Seshan, I. Stoica, and H. Zhang. Developing a Predictive Model of Quality of Experience for Internet Video. In *SIGCOMM*, 2013.

[15] H. Balakrishnan, V. N. Padmanabhan, S. Seshan, M. Stemm, and R. H. Katz. TCP Behavior of a Busy Internet Server: Analysis and Improvements. In *INFOCOM*, 1998.

[16] P. Barford and M. Crovella. Generating Representative Web Workloads for Network and Server Performance Evaluation. In *SIGMETRICS*, 1998.

[17] A. Bouch, A. Kuchinsky, and N. Bhatti. Quality is in the Eye of the Beholder: Meeting UsersâĂŹ Requirements for Internet Quality of Service. In *CHI*, 2000.

[18] M. Butkiewicz, H. V. Madhyastha, and V. Sekar. Understanding Website Complexity: Measurements, Metrics and Implications. In *IMC*, 2011.

[19] B. Clifton. *Advanced Web Metrics with Google Analytics*. John Wiley and Sons, 2012.

[20] H. Cui and E. Biersack. On the Relationship between QoE and QoS for Web Sessions. In *Eurecom Research Report*, 2012.

[21] D. F. Galletta, R. Henry, S. McCoy, and P. Polak. Web Site Delays: How Tolerant are Users? In *Journal of the Association of Information Systems*, 2004.

[22] N. Godbole, M. Srinivasaiah, and S. Skiena. Large-Scale Sentiment Analysis for News and Blogs. In *ICWSM*, 2007.

[23] A. Hassan, Y. Song, and L. wei He. A Task Level Metric for Measuring Web Search Satisfaction and its Application on Improving Relevance Estimation. In *CIKM*, 2011.

[24] S. Ihm and V. S. Pai. Towards Understanding Modern Web Traffic. In *IMC*, 2011.

[25] S. S. Krishnan and R. K. Sitaraman. Video stream quality impacts viewer behavior: Inferring causality using quasi-experimental designs. In *IMC*, 2012.

[26] C. Liu, R. W. White, and S. Dumais. Understanding Web Browsing Behaviors through Weibull Analysis of Dwell Time. In *SIGIR*, 2010.

[27] B. A. Mah. An Empirical Model of HTTP Network TrafiňĄc. In *INFOCOM*, 1997.

[28] L. A. Meyerovich and R. Bodik. Fast and Parallel Webpage Layout. In *WWW*, 2010.

[29] T. Mitchell. *Machine Learning*. McGraw-Hill.

[30] J. C. Mogul. The Case for Persistent-Connection HTTP. In *SIGCOMM*, 1995.

[31] M. Sahami, S. Dumais, D. Heckerman, and E. Horvitz. A Bayesian Approach to Filtering Junk E-Mail. In *AAAI*, 1998.

[32] M. Z. Shafiq, J. Erman, L. Ji, A. X. Liu, J. Pang, and J. Wang. Understanding the Impact of Network Dynamics on Mobile Video User Engagement. In *SIGMETRICS*, 2014.

[33] M. Z. Shafiq, L. Ji, A. X. Liu, J. Pang, and J. Wang. Characterizing Geospatial Dynamics of Application Usage in a 3G Cellular Data Network . In *INFOCOM*, 2012.

[34] F. D. Smith, F. H. Campos, K. Jeffay, and D. Ott. What TCP/IP Protocol Headers Can Tell Us About the Web. In *SIGMETRICS*, 2001.

[35] S. Sundaresan, N. Feamster, R. Teixeira, and N. Magharei. Measuring and Mitigating Web Performance Bottlenecks in Broadband Access Networks. In *IMC*, 2013.

[36] T. Tullis and W. Albert. *Measuring the User Experience: Collecting, Analyzing, and Presenting Usability Metrics*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2008.