TransLayer: A Root Cause Identification Tool for Analyzing Cellular Abnormal Behaviors

Haokun Luo[†], Z. Morley Mao[†], Kranthi Sontineni[♮], Jie Hui[♮], Kevin Lau[♮]
[†]University of Michigan, [♮]T-Mobile USA Inc.

ABSTRACT

Computer Network is originally designed in layers. The protocol decision in a single layer does not depend on the circumstances of other layers, and the scope of optimization is usually within that layer. However, single layer optimization is not necessarily the same as overall performance optimization. Because of the portability of the network protocol, same upper layer protocol could be reused under different lower layer settings. For example, TCP tends to perform worse in the cellular network compared with wired and wireless network because of distinct lower layer protocol designs and network settings. We propose a root-cause identification tool, *TransLayer*, which provides a transparent view across different network layers in the cellular network. We design and implement a real-time user feedback tool to help us allocate the QoE (Quality of Experience) problems in the network layer. After pinpointing the performance problems in network layers, we could directly correlated the lower layer features from QxDM (Qualcomm eXtensible Diagnostic Monitor) logs. We demonstrate the root cause analysis with abnormal RRC (Radio Resource Control) state transitions, and identify the root cause to be the inference between control plane and data plane traffics.

General Terms

Mobile, Measurement, QoE, Tool

Keywords

LTE, WCDMA, RRC, RLC, Cross-layer Analysis, Root Cause Analysis, Real-time User Feedback

1. INTRODUCTION

3G WCDMA and 4G LTE techniques are widely deployed around the world-wide service providers. However, a number of abnormal cellular performance behaviors due to non-adaptive upper layer protocol have been identified. For example, Bufferbloat [19] indicates that large queueing buffer at the Node B [1] or eNB [2] could cause significant latency downgrade. Measurements from [21] shows TCP RTO (Retransmission TimeOut) could under estimate the RTT (Round Trip Time), which causes a substantial delay during congestion control. Both claim that the RLC (Radio Link Control) layer retransmission could be a big contributor to those abnormal behaviors. Unfortunately, all of them, including similar work [16, 17], did not have access to the ground truth information about the data plane in the lower layer, so they could not provide quantitative analysis to prove their claims.

QxDM [7] is a PC based cellular network monitoring and diagnosis tool that records ground truth information in the data and control plane information across different network layers. We build, a cross-layer mapping component in *TransLayer*, to leverage such valuable information to perform cross-layer analysis in order to quantize the impact of RLC layer retransmission and estimated first-hop latency, and also correlate upper layer behaviors with fine-grained context information. In the case study, we reproduce RRC inference algorithm from [16]. However, we found an substantial transport layer latency during the RRC state transitions. After correlating the upper layer information with that in lower layer, we pinpoint the root cause the performance issues in the RLC layer to be the interference from control plane messages.

State-of-art mobile researches were heavily focus on improving bandwidth and reduce latency to allow a better QoS (quality of service) to mobile users. Researchers have actively collected real user-traces (i.e. ARO [17], MyExperience [20]) combining various of context information to identify the performance problems. However, even though resolving their recognized issues, the service provider could only claim a “best effort” service for end users. The fundamental problem for previous approaches is that we are not sure whether users are happy with those “best effort” solutions, and not successful to identify the problems when “best effort” fails to meet user satisfaction. In order to acquire the ground truth application performance problem, we design and implement a real-time user feedback collection tool. It locates all the user dissatisfaction periods, which enables us to use *TransLayer* to find root causes in the lower layer.

We show the overall design of *TransLayer* in Figure 1. The goal of *TransLayer* is to provide a transparent view to mobile researchers, and assist them to identify the root causes of application performance problems in the cellular

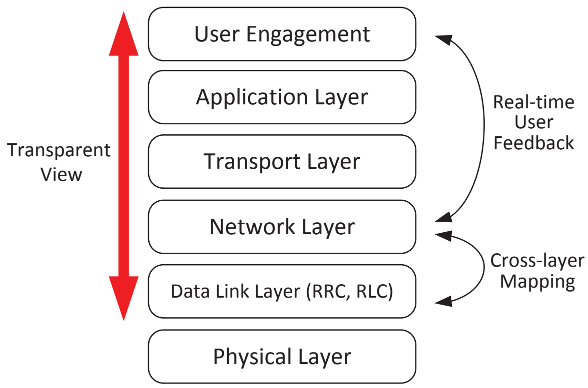

Figure 1: Overview architecture of *TransLayer*

network. There are two essential parts in *TransLayer*: a real-time user feedback collection tool and a cross-layer mapping tool. The real-time user feedback collection tool correlates user identified performance issues with the actual IP packets. The cross-layer mapping tool would utilize lower layer features to pinpoint the root causes.

Our contributions in this paper are:

- We implemented a cross-layer analysis tool, as a part of *TransLayer*, to correlate the upper layer performance problems with ground truth lower layer information. Our tool could achieve up to *99.8%* successful mapping from network layer to RLC layer PDU (Packet Data Unit), which is the smallest data transmission unit in RLC layer. To the best of our knowledge, the tool is the first implementation that enables cross-layer mapping between ground truth information from upper layer to the lower layer.

- We designed a simple real-time user feedback tool, as another part of *TransLayer*, to allocate the ground truth application performance issues. The tool is generalized to apply to any applications to assist people studying QoE problems in cellular network.

- We performed root cause analysis to identify interference between control and data plane problem that could potentially downgrade application performance. The root cause could be a direction for ISP (Internet Service Provider) to improve the cellular performance by modifying the implementation in Node B or eNB.

The paper is organized as follows. § 2 takes a brief overview on background knowledge of cellular network. § 3 highlights the potential challenges in developing *TransLayer*. § 4 explains the cross-layer mapping mechanism, and § 5 describes the design of the real-time user feedback collection tool. § 6 presents a case study of root cause analysis on abnormal RRC state transition problem. § 7 lists the future work for *TransLayer*, and § 8 summarizes the related work in cross-layer analysis and QoE study.

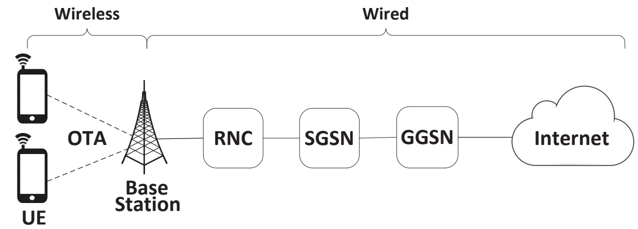

Figure 2: The general cellular network architecture

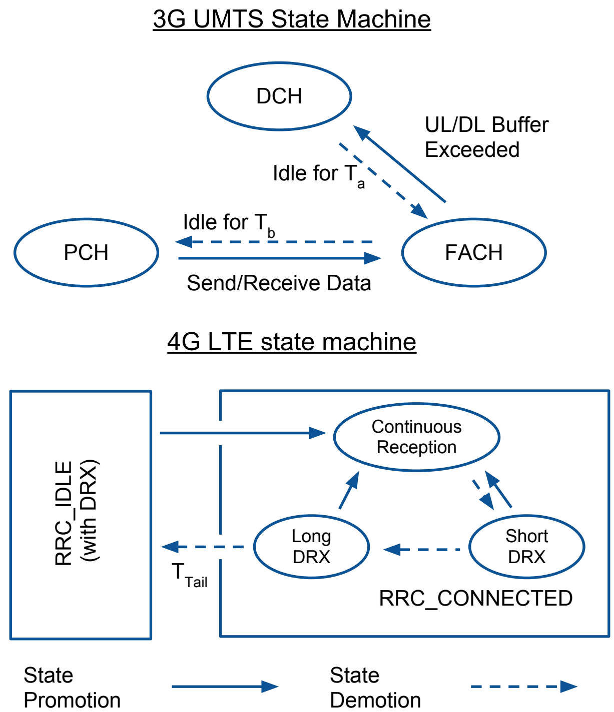

Figure 3: Possible 3G and 4G State Machines as Specified by 3GPP.

2. BACKGROUND

As illustrated in Figure 2, in both 3G UMTS [1] and 4G LTE networks [2], data is transmitted from *user equipment* (UE), i.e. mobile devices, to the base station (Node B in UMTS, eNB in LTE), then to the *Serving GPRS support node* (SGSN) and *Gateway GPRS support node* (GGSN), and ultimately to the server [18]. The link between the UE and the base station is known as the *over-the-air* (OTA) link, and it is the only wireless link in the network topology. The rest of the links from the base station to the internet are all wired.

2.1 Radio Resource Control

Mobile cellular networks use the RRC protocol as the

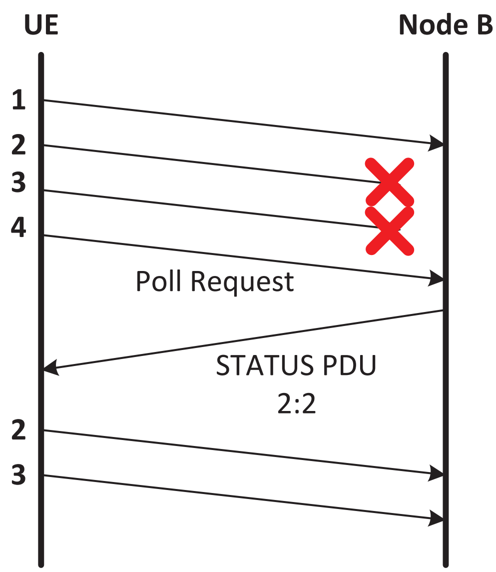

Figure 4: RLC PDU transmission in Acknowledged Mode. When the sender sends out a polling request, the receiver will respond with a "group" of acknowledge in terms of the starting PDU sequence number and the consecutive PDU loss length.

control plane signaling for Layer 3 to allocate resources to mobile devices. Handsets transition between different RRC states, which vary in power consumption and bandwidth, and an individual RRC state machine is maintained for each handset. Transitions between different states occur due to traffic patterns between the device and base station. In general, more traffic will cause a higher-power and higher-bandwidth state to be entered. The RRC protocol for these network types has been defined by 3GPP [9, 10].

For 3G UMTS [9], there are three main states: DCH, which is high-power and high-bandwidth, FACH, which is low power and low bandwidth, and PCH, where no transmission is possible. If a higher-bandwidth state is needed, there is a promotion delay. Some carriers may always go directly from PCH to DCH. An example RRC state machine can be seen at the top of Figure 3.

For 4G [10], the state machine is more complicated, and is summarized in the bottom half of Figure 3. We are concerned mainly with transitions between RRC_CONNECTED, a higher-power state, and RRC_IDLE, a lower-power state where no data is transmitted. The other states have timers on the orders of tens or hundreds of milliseconds are are not practical to measure on end-user devices, as tools such as a power monitor are required.

2.2 Radio Link Control

RLC protocol is a packet fragmentation protocol that used as the data plane transmission for cellular air interface [8]. There are three different modes that configured by the upper layer: TM (Transparent Mode), UM (Unacknowledged Mode), and AM (Acknowledged Mode). RLC in TM directly pass the packet from PDCP (Packet Data Convergence Protocol) layer to MAC layer, and it usually applies for voice transmission. RLC in UM sends the PDU (Packet Data Unit), which is smallest data fragmentation unit, without waiting for any response from the receiver (in this case, Node B or eNB). RLC usually transmit VoIP traffic with UM. RLC in AM deploys an ARQ (Automatic Repeat reQuest) mechanism. As show in Figure 4, RLC sender periodically piggybacks the polling request in the RLC header, and the receiver will respond with a RLC STATUS PDU which contains groups of not received PDU sequence numbers. We call this particular ARQ mechanism as *group acknowledgement*. Since RLC transmit both TCP and UDP with AM, we will assume the rest of RLC transmission mechanism to be AM.

3. CHALLENGES

QxDM is essentially a real-time monitoring tool that connect the device with a PC (requires a USB authentication dongle). Each message or event from different layers is stored using a single log entry that has consistent format. All the analysis process is offline after we exported the filtered log information from QCAT (QualComm Analysis Tool). The first step of building cross-layer mapping of *TransLayer* is to parse the trace and extract meaningful information from each log entry. Unfortunately, due to the defects of the QxDM design, incomplete RLC PDU information, loose RLC RTT estimation, and unnecessary IP packet fragmentation and duplication should be taken into consideration. For the real-time user feedback part, the most challenging component is to accurately record all the user feedback and minimize the number of incorrect signals.

3.1 Incomplete RLC PDU Logging

Due to a bad design decision, QxDM merely provides the first four bytes information in the RLC PDU for both uplink and downlink in 3G (only two bytes in LTE). Since the first two bytes PDU is constantly reserved as header, i.e. sequence number, PDU type (control PDU or data PDU), and etc, we only have two types as payload for IP packet mapping. This leads to potential unsuccessful cross-layer mapping. We provide some degree of confidence on the accuracy of mapping using uniqueness analysis 4.3, which will be covered later.

3.2 Loose RLC RTT Estimation

In TCP, we could easily calculate the RTT (Round Trip Time) for each data packet, because each packet will be received an ACK (acknowledgement) packet except for packet loss. However, as we talked about in the § 2.2, RLC layer uses *group acknowledgement* which implies not every RLC PDU could get ACK feedback from the receiver. RTT information in the lower layer is more coarse-grained compared with that of transport layer. In the future root cause analysis, we always estimate the lower layer RTT using the nearest available polling request PDU's RTT value.

3.3 IP Fragmentation and Duplication

Although packets will be fragmented in the RLC layer because of the protocol, QxDM surprisingly breaks IP packets into units of 256 bytes without a reason. It turns out that such unnecessary fragmentation brings lots of trouble on cross-layer mapping accuracy. Another issue with IP packets is the duplication. That comes from the same IP packet payload logging twice but from different interfaces, even though you enforce the interface to be set as one of them. Therefore, pre-processing on IP log entries is quite necessary, and we evaluate the accuracy improvement to be *15.24%* in our standard *Browsing Dataset* from § 6.1.

3.4 Feedback Tool

The smartphone is a very compact device that allows multiple ways for user to input their requests. However, many of the input components are well defined. Simple functionality overwrite will lead to block the normal user queries. If we create complicated input mechanism, then we might not capture all the user feedback signals. Selecting the proper interface to allow user provide feedback becomes the most challenging part of the design.

4. CROSS-LAYER TOOL DESIGN

TransLayer is a cross-layer mapping and root cause analyzing tool based on QxDM traces. Even though it requires QxDM to gather information, we claim that such detailed lower layer information empowers us to gain a transparent view across different layers. The primary use case for *TransLayer* is to identify performance issues. We talk about how to collect ground truth data transmission information in IP and RLC layer using QxDM in § 4.1. § 4.2 talks about the core algorithm to enable cross-layer mapping between IP and RLC layer. § 4.3 verifies the cross-layer mapping accuracy by showing the unique existence in the whole trace. § 4.4 lists the major lower features that we utilize to perform root cause analysis.

4.1 QxDM

QxDM provides the the IP and lower information as input to *TransLayer*. It is a real-time data collection and diagnostic logging tool for measuring mobile-based RF (Radio Frequency) performance [7]. It is a Windows based monitoring application. When we perform control experiments and real application measurements, we plug in the device to the desktop or laptop with QxDM software installed. Once the experiments finishes, we filtered out the real-time monitoring information related to IP packets, RLC PDUs, and RRC states in Table 1, and dump the results into a log file. The 0x11EB log entry includes IP headers, IP payloads, and its customized header. Since large IP packets will be fragmented into smaller segments, the customer header could indicate the segment index of the whole IP packet. The 0x4132 and 0x4133 unveil the RLC AM configurations, i.e. the polling function timers, the retransmission limit for a single PDU, and etc. The 0x413B, and 0x418B provides RLC PDU header and first byte payload information for both data PDUs and control PDUs (or STATUS PDUs) in both uplink and downlink directions. We wrote a QxDM log parser to aggregate the filtered entries in *TransLayer*, and apply cross-layer mapping to understand the correlation between different layers. We conduct all the experiments on Galaxy S3 with Android OS 4.1.1.

QxDM Log ID	Description
0x11EB	IP data packets
0x4132	WCDMA RLC downlink acknowledge mode configuration
0x4133	WCDMA RLC uplink acknowledge mode configuration
0x413B	WCDMA RLC uplink acknowledge mode PDU
0x418B	WCDMA Flexible RLC downlink acknowledge mode PDU
0x4125	WCDMA RRC states

Table 1: QxDM log entries used in cross layer analysis

4.2 Cross-layer Mapping

QxDM tool provides fine grained lower layer RLC layer transmission information. If we could correlate the transport layer packets with the RLC layer transmitted PDUs, we will have a transparent view of the link layer behaviors, especially the RLC retransmission. As mentioned in § 3.1, one of the fundamental limitation of QxDM is the partial logging issue. For example, only the header and first byte data payload will be logged for each RLC PDU. It is also possible that a small fraction of RLC PDUs cannot be captured, which lead to a unnecessary sequence number gap. Our mapping algorithm will handle all the limitations we have mentioned.

The cross-layer mapping algorithm is the core technique for *TransLayer*. It is essentially a map between the complete IP packets (also known as SDU) and corresponding fragmented RLC payload data bytes (known as PDU). Due to the partial logged information in QxDM, only the first data byte is captured in the log. Thus, we have to skip over the rest of the PDU, and try to match for the first data byte in the next PDU as shown in Figure 5, which we call the *long jump mapping*. The problem at this point is to determine the end of IP packets while we iterate through the consecutive RLC PDUs. Since each PDU could either contain the payload data dedicated to a single SDU or belongs to two SDUs. If

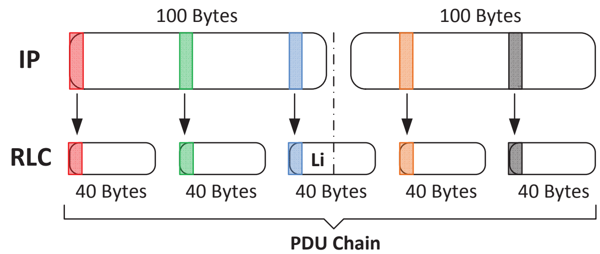

Figure 5: Cross-layer mapping from IP packets to RLC PDUs chains, where are a sequence of consecutive RLC PDUs. Due to the partial logging, we could only map the limited bytes for every certain length (*long jump mapping*). It is possible that one PDU could concatenate the last part of the first SDU and the first part of the second SDU. Therefore, the five RLC PDU forms a unique RLC PDU chain.

the reminder size of the SDU cannot fulfill the largest size of PDU, then RLC protocol will concatenate the part of the next SDU to fill the rest of space [8] using LI (Length Indicator) in the RLC PDU header. Ultimately, if the accumulative mapped index equals the size of SDU, we claim to find a mapping successfully; otherwise no mapping discovered.

There is a corner case in the mapping algorithm such that the QxDM cannot capture the some of the SDUs. Similar to TCP protocol, the sequence number in RLC PDUs could uniquely distinguish between every PDUs. If there are some missed PDUs, then we cannot map the first byte data for every PDU size. In that case, we could even skip over the missed PDUs and add up multiple of PDU size to hunt for a match. Because of the limitation of the QxDM tool, the *long jump mapping* mechanism cannot fully recover the corner case, especially when the missed PDUs were either the beginning or the end part of the mapped RLC list. We evaluate the improved mapping algorithm by checking the percentage of mapped IP packets in our standard *Browsing Dataset* from § 6.1, and the average mapping ratio is *99.52%* for uplink and *88.83%* for downlink. One reason for a lower the downlink ratio is that the size of PDU is flexible in WCDMA downlink, and the average size is around 17 times greater than that of uplink in our traces. Considering the partial logging information, that actually implies fewer information is available in downlink RLC layer, which leads to a smaller successful mapping ratio.

4.3 Uniqueness Analysis

The cross-layer mapping in § 4.2 only claims the existence of cross-layer, but it does not guarantee the mapped RLC PDUs exist once in the trace. We perform uniqueness analysis to prove the uniqueness of the mapped RLC PDUs in the whole trace. One important observation is that RLC SDU concatenation is fairly common as shown in Figure 5, especially during bulk transmission in transport layer. The concatenation actually introduces some degree of randomness of the partial RLC PDU payload locations. In other word, the cross-layer mapping does not always start from the very first byte in SDU within the same RLC PDU chain. For example, the two SDUs in Figure 5 have different byte placements that mapped to lower layer.

Type	Percentage of Unique Mapping (%)
TCP uplink	91.42
TCP downlink	88.70
UDP uplink	30.86
UDP downlink	93.94

Table 2: Evaluate the uniqueness of RLC PDU chains in the standard web browsing trace

We analyze the uniqueness of RLC PDU chains, which are essentially a sequence of data bytes in RLC PDU payloads. In Table 2, after classifying all the PDU chains into guarantee uniqueness and not guarantee uniqueness, we apply cross-layer mapping on our standard *Browsing Dataset* from § 6.1, and examine the percentage of mapped RLC PDU sequences residing inside an unique RLC PDU chains. We notice that UDP uplink is relative lower than the other three categories. The reason is that the majority of UDP uplink traffic is DNS lookup. The size of the UDP packet is usually less than 70 bytes. Since each uplink RLC PDU is fixed 40 bytes, that implies two PDUs are sufficient to carry a single DNS lookup request. The only information that is useful mapping is the first two bytes, and the 41 and 42 bytes. For a IP packet, the first two bytes is always “45 00” in IPv4 without exception for DNS lookup. The 41 and 42 byte are actually the first two bytes of the request URLs, which are most likely to be the first two bytes of string “www”. To make it worse, the mapped RLC PDUs are usually stand alone without help from SDU concatenation. Therefore, UDP uplink is much worse than the other three categories.

4.4 Lower Layer Features

4.4.1 RLC Retransmission

As mentioned in § 2.2, RLC protocol is implemented with ARQ mechanism. Once the sender received the STATUS PDU from the receiver, any unreceived RLC PDU will be preempted to be retransmitted from the retransmission buffer [8]. In Figure 4, the payload of STATUS PDU is a sequence of number pairs. The first number is the starting index of unreceived RLC PDU sequence number, and the second number indicates the length of consecutive unreceived RLC PDUs from that starting sequence number. In this case, the second and the third PDU are not received by the receiver.

Due to the header size limitation of RLC PDU, sequence number is assigned with period of 4096 in both uplink and downlink (rare exceptional cases are possible). Because of the partial logging information, STATUS PDU payload is not complete, so we cannot directly allocate the retransmitted RLC PDUs from those control message. Instead, *TransLayer* follows the periodical sequence number reuse rule to capture the smaller period for sequence number reuse as a sign of RLC PDU retransmission. *TransLayer* labels the retransmitted RLC PDUs as a pre-process so that we could determine the number of retransmitted RLC PDU within a mapped RLC PDU chain as RLC retransmission count. The RLC retransmission ratio is defined as RLC retransmission count divided by the total number of PDU in a RLC PDU chain.

4.4.2 First-hop Latency

The first-hop latency is defined as combination of transmission delay of all the RLC PDUs that belongs to the same IP packet, and the OTA (over the air) RTT in Figure 2. *TransLayer* considers the two latency values separately. Transmission delay is the time difference between the first transmitted PDUs and the last one. We would also want to normalize the transmission delay by dividing the number of mapped RLC PDUs. OTA RTT is essentially the one-way delay from the device to Node B or eNB. As mentioned in § 3.2, it is not feasible to accurately calculate the specific RTT for every RLC PDU (we can for those header enables the polling request bit). *TransLayer* uses the nearest available RTT as an approximation for OTA delay.

4.4.3 Power

There are two important power feature that could infer the traffic load in the cellular network: UE received signal strength and SNR (Signal-to-Noise Ratio) [12]. Signal strength is the received signal power from Node B or eNB, which could be used a relative value to indicate the channel quality. It is usually be referred as RSCP (Received Signal Code Power) in WCDMA and as RSRP (Reference Signal Received Power) in LTE. SNR is a indicator value that combines the pilot power with the inference from other subcarriers. In WCDMA, it is called ECIO (Energy / Interference) in WCDMA and RSRQ (Reference Signal Received Quality) in LTE. *TransLayer* collects all those context information and assign to each transport layer packets.

5. FEEDBACK TOOL DESIGN

The goal the real-time user feedback is to collect user dissatisfaction moment with minimized user interruption, accurately logging feedback, and no extra cellular traffic. § 5.1 talks about various approaches to design the tool, and compare the trade-offs. § 5.2 explains the realization of the real-time user feedback tool.

5.1 Approach

There are two perspective of achieve feedback logging accuracy. First, any unintended feedback signals should not be recorded. Statistically, the false positive error should be minimized. Second, whenever users generate the feedback, we must capture it. Since user feedback is extremely important information, we have zero tolerance on false negative error.

There are several ways to collect user feedback: screen interaction, device sensors, and hardware inputs. The touch screen is the main innovation of smartphone compared with feature phone. It accepts the majority of the user inputs to the device. That also implies either the screen interaction is too simple to be unclassified from other interactions, or the interaction is over complicated to interfere user behaviors. The sensors on the smartphone are not very accurate. It is desirable for inconsistent unstable sensor behaviors due to software implementation bugs or hardware issues [5]. Hardware button is relative more accurate than device sensors to avoid any false negative errors. Although we may not distinguish the original button functionality with the feedback signals, the false positive error is also not avoidable in the screen interaction approach. We decide to use the volume up and down buttons to capture all the user feedback signals.

5.2 Implementation

The feedback tool is essentially another Android application that runs in the background to minimize the user interruption. Whenever the user presses the volume up or volume down button, it receives a volume change broadcast signal, then logs the timestamp onto the disk. Notice that no extra network traffic is generated during the process. Later, we could dump the feedback file from the device and cooperate the QxDM traces to perform the cross-layer analysis as we discussed in § 4.2.

6. EVALUATION

6.1 Dataset

We collect two different datasets to show the usage of *TransLayer*. One dataset is a control experiment that reproduces the RRC inference algorithm from [16], called *Control Dataset*. The duration for *Control Dataset* is 5 hours. Another dataset is collected from real-time web browsing in WCDMA network for an hour, called *Browsing Dataset*. We have evaluate the percentage of mapping in § 4.2, uniqueness analysis in § 4.3, and IP fragmentation and replication elimination in § 3.3 using *Browsing Dataset*.

6.2 Case Study

6.2.1 Abnormal RRC State Transition

[16] tries to infer RRC state using RTT differences. As we mentioned in § 2.1, after the device promoted into DCH state, it could demote to a lower power state, i.e. FACH or PCH, due to timeout on no data transmission. The RRC inference mechanism measures different RTTs based on different inter-packet timing intervals. For any substantial dif-

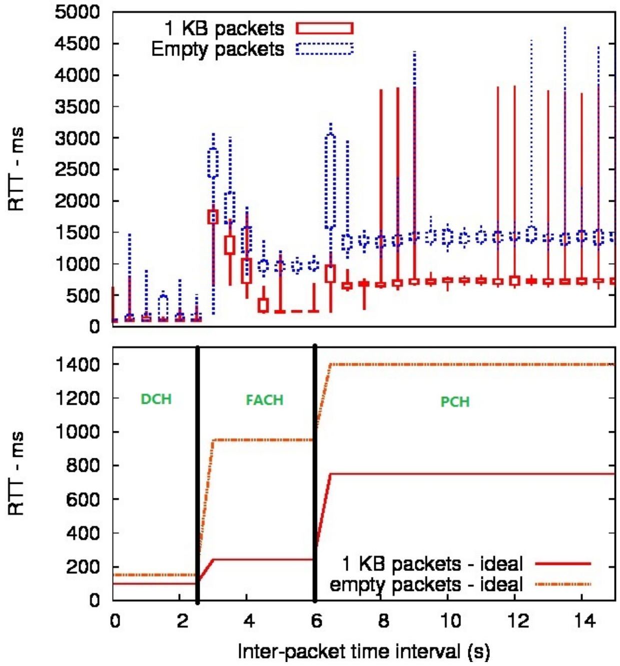

Figure 6: Reproduced RRC inference algorithm results. Unexpected RTT around inter-packet timing 3.5s.

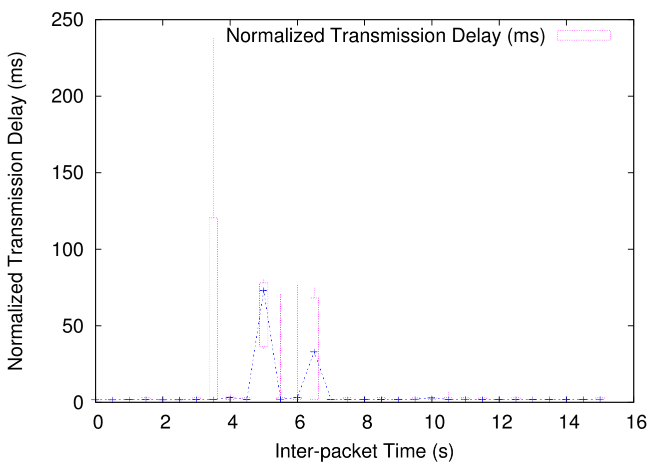

Figure 7: In normalized transmission delay feature, we observe the same delay around 3.5s which matches the observation from the *Control Dataset*

ference in the two adjacent inter-packet timings, we could approximate the demotion timer to be in between the two inter-packet timing intervals. We reproduce results as in Figure 6.

We apply the cross-layer mapping, and extract the lower layer features as mentioned in § 4.4. Among all the feature, we find the normalized transmission delay matches our control experiments results as in Figure 7. Then we take a deeper dive in the QxDM, we figure out that large amount of control messages arrived in the middle of RLC PDU data transmission as shown in Figure 8. Data fragmentation as

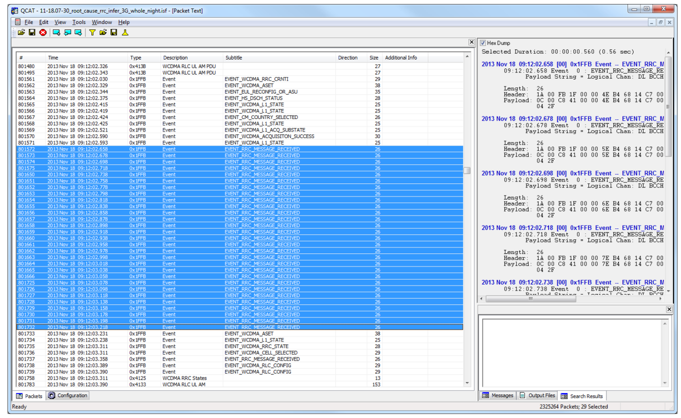

Figure 8: QxDM logging information around the inter-packet timing 3.5s. Substantial amount of PRACH control messages from Node B block the data plane transmission

in WCDMA uplink would increase the possibility of unnecessary control message interference. We do not observe similar problems in the downlink because the average PDU size is around 17 times larger than the uplink. That implies much less fragmentation occur in WCDMA downlink, and less chance to get interrupted by control messages.

7. FUTURE WORK

7.1 User Study

We would like to evaluate our real-time user feedback tool with main stream user interactive applications. One of the recent post suggested that the 53% of the smartphone users listens to Internet radio [3]. There are two popular Internet radio applications: TuneIn [6] and Pandora [4]. We will conduct a user study to collect real-time feedback from users when they are listening to Internet radio. Then we could use the cross-layer mapping to identify the root causes for bad QoE periods.

7.2 Comprehensive Analysis

We would like to compare the performance across different network carriers and under various of network conditions. Utilizing the cross-layer mapping tool, we will have a better understanding of the implementation difference between carriers, and the trade-off of those design decisions. Previous studies suggested link quality and traffic load could affect cellular network performance [12, 13]. We are also thinking that the actual implementation of lower layer protocol could also be a contributor. That requires to perform a comprehensive analysis on various network carriers and distinguished network conditions.

8. RELATED WORK

Cross-layer Analysis: ARO [17] is passive measurement tool that correlates application layer, transport layer, and RRC layer to identify the unnecessary energy consumption due to improper application design. ARO only refers RRC state using network latency measurement results due to lack of accessibility to ground true lower layer information. *TransLayer* not only has access to the lower control plane information, but also the data plane. That benefits us from performing fine-grained lower information collection and analysis, and we are more confident about the accuracy of cross-layer mapping results.

Rilanalyzer [22] enables ground truth RRC information on Intel/Inneon XGold chipsets with root access, and it does not requires to connect to a personal computer. Although RRC state allocation could affect the energy consumption on the devices, it has a much less impact on the device performance. QxDM provides much detailed lower layer information that allows us to extract tons of lower layer features to pinpoint the root causes of the performance issues.

The survey [24] summarized the cross-layer design for wireless network, and highlighted several innovative design for non-layered network topology. It indicates that many performance problems come from the blocked communication mechanism between different network layers. The performance issue from a single layer could propagate to others due to not inform other layers in time. In our study, we would like to support that introducing communication between layers could potentially benefit the overall performance in the cellular context.

Quality of Experiences: In recent years, people started to shift focus from "best effort" level QoS performance improvement study to a more realistic QoE user-oriented analysis. Moorsel defined quality of experience as the combination of reliability with performance, or in his words "performability" [14]. Ickin, et al., proposed various QoE metrics for mobile applications from a Human Computer Interaction (HCI) perspective [23], which covers application performance, battery, phone features, app and data cost, user's routine, and user's lifestyle.

[15] studied the impact of video quality on user engagement, and used measurement results to approve that the QoS performance metrics do not linearly related to user engagement metrics. [11] proposed a predictive model that utilized actionable performance metrics to improve the user engagement using decision tree from machine learning. Those study collect user feedback information passively.Our active real-time feedback collection mechanism is actually the opposite approach. *TransLayer* focuses on locating the ground truth information about bad user engagement.

9. CONCLUSION

In this paper, we design and implement *TransLayer*, a root cause analysis tool that helps people identify the root causes of cellular network performance problem. There are two parts in the tool: real-time user feedback collection and cross-layer mapping. Because of the partial logging challenges in QxDM, we provides uniqueness analysis to guarantee the correctness of mapping. The mapping accuracy is up to 93.94% in our *Browsing Dataset*. We also design and realize the real-time user feedback collection tool. We log the feedback directly to disk without generating any extra network traffic, when we detect user pressing the volume up or volume down button. We present a case study of root cause analysis on abnormal RRC state transition problem, and find the problems comes from the interference between control plane and data plane traffic.

10. ACKNOWLEDGMENTS

We would like to thank Yihua Guo, Qi Chen, Sanae Rosen, and Z. Morley Mao for their valuable comments on this report. We would also like to appreciate the support from TMobile.

11. REFERENCES

[1] 3GPP LTE. `http://www.3gpp.org/LTE`.
[2] 3GPP UMTS (3G). `http://www.3gpp.org/UMTS`.
[3] Internet radio becoming more mainstream, survey says. `http://goo.gl/wNs31W`.
[4] Pandora. `http://www.pandora.com/`.
[5] The iPhone 5S Motion Sensors Are Totally Screwed Up. `http://goo.gl/UGmN0e`.
[6] TuneIn. `http://www.tunein.com/`.
[7] QxDM Professional QUALCOMM eXtensible Diagnostic Monitor. `http://goo.gl/DOmGz1`, 2012.
[8] 3GPP TS 25.322: Radio Link Control (RLC) - UMTS, 2013.
[9] 3GPP TS 35.331: Radio Resource Control (RRC) - UMTS, 2013.
[10] 3GPP TS 36.331: Radio Resource Control (RRC) - LTE, 2013.
[11] A. Balachandran, V. Sekar, A. Akella, S. Seshan, I. Stoica, and H. Zhang. Developing a predictive model of quality of experience for internet video. *Proc. ACM SIGCOMM*, 2013.
[12] A. Chakraborty, V. Navda, V. N. Padmanabhan, and R. Ramjee. Coordinating cellular background transfers using loadsense. *Proc. ACM MobiCom*, 2013.
[13] A. Schulman, V. Navda, R. Ramjee, N. Spring, P. Deshpande, C. Grunewald, K. Jain, and V. N. Padmanabhan. Bartendr: a practical approach to energy-aware cellular data scheduling. *Proc. ACM MobiCom*, 2010.
[14] A. V. Moorsel. Metrics for the internet age: Quality of experience and quality of business. In *In Fifth International Workshop on Performability Modeling of Computer and Communication Systems*, 2001.
[15] F. Dobrian, V. Sekar, A. Awan, I. Stoica, D. A. Joseph, A. Ganjam, J. Zhan, and H. Zhang. Understanding the impact of video quality on user engagement. *Proc. ACM SIGCOMM*, 2011.

[16] F. Qian, Z. Wang, A. Gerber, Z. M. Mao, S. Sen, and O. Spatscheck. Characterizing Radio Resource Allocation for 3G Networks. In *Proc. ACM IMC*, 2010.

[17] F. Qian, Z. Wang, A. Gerber, Z. M. Mao, S. Sen, and O. Spatscheck. Profiling Resource Usage for Mobile Applications: A Cross-layer Approach. In *Proc. ACM MobiSys*, 2011.

[18] H. Holma, and A. Toskala. *WCDMA for UMTS: HSPA Evolution and LTE.* John Wiley and Sons, Inc., 2007.

[19] H. Jiang, Y. Wang, K. Lee, and I. Rhee. Tackling bufferbloat in 3G/4G networks. In *Proc. ACM IMC*, 2012.

[20] J. Froehlich, M. Y. Chen, S. Consolvo, B. Harrison, and J. A. Landay. MyExperience: a system for in situ tracing and capturing of user feedback on mobile phones. In *Proc. ACM MobiSys*, 2007.

[21] J. Huang, F. Qian, Y. Guo, Y. Zhou, Q. Xu, Z. M. Mao, S. Sen, and O. Spatscheck. An in-depth study of LTE: Effect of network protocol and application behavior on performance. In *Proc. ACM SIGCOMM*, 2013.

[22] N. Vallina-Rodriguez, A. Auinas, M. Almeida, Y. Grunenberger, K. Papagiannaki, and J. Crowcroft. Rilanalyzer: a comprehensive 3g monitor on your phone. In *Proc. ACM IMC*, 2013.

[23] S. Ickin, K. Wac, M. Fiedler, L. Janowski, J. Hong, and A. K. Dey. Factors influencing quality of experience of commonly used mobile applications. In *IEEE Communication Magazine*, 2012.

[24] V. Srivastava, and M. Motani. Cross-layer design: a survey and the road ahead. In *IEEE Communication Magazine*, 2005.