A Smartphone-Based Platform to Test the Performance of Wireless Mobile Network and Preliminary Findings

Xinli Geng1,4, Hao Xu2, Xiaowei Qin3
1Department of Automation, University of Science and Technology of China, Hefei, Anhui 230027, China;
2Department of Civil and Environmental Engineering, University of Nevada, Reno, Reno, NV 89557, USA;
3Department of Electronic Engineering and Information Science, University of Science and Technology of China, Hefei, Anhui 230027, China;
4Institute of Applied Technology, Hefei Institutes of Physical Science, Chinese Academy of Science, Hefei, Anhui 230088, China;

ABSTRACT

During the last several years, the amount of wireless network traffic data increased fast and relative technologies evolved rapidly. In order to improve the performance and Quality of Experience (QoE) of wireless network services, the analysis of field network data and existing delivery mechanisms comes to be a promising research topic. In order to achieve this goal, a smartphone based platform named Monitor and Diagnosis of Mobile Applications (MDMA) was developed to collect field data. Based on this tool, the web browsing service of High Speed Downlink Packet Access (HSDPA) network was tested. The top 200 popular websites in China were selected and loaded on smartphone for thousands times automatically. Communication packets between the smartphone and the cell station were captured for various scenarios (e.g. residential area, urban roads, bus station etc.) in the selected city. A cross-layer database was constructed to support the off-line analysis. Based on the results of client-side experiments and analysis, the usability of proposed portable tool was verified. The preliminary findings and results for existing web browsing service were also presented.

Keywords: QoE, wireless network services, web browsing, HSDPA.

1. INTRODUCTION

With the rapid development of wireless networks and the fast evolution of smartphone, more and more Web Applications (WebApps, such as Microblogging, Wechat, and Navigation map applications, etc.) have been integrated into mobile terminals. This phenomenon caused the mobile traffic growth fast. As Cisco predicts [1], the global mobile devices and connections will increase to 10.2 billion by 2018 by a Compound Annual Growth Rate (CAGR) of 8 percent. At the same time, this tendency will increase the competition between mobile phone manufacturers, network operators, WebApps developers. However, the corporations between aforementioned service providers can definitely improve services’ performance. For example, when the application accesses to the wireless network and transports data with the server, the receive power and transmission power will play a significant role in the battery consumption, which will influence the quality of experience (QoE) directly. The application developer will pay more attention to the operability and interface of the application instead of battery consumption, which is important for phone manufacturers and network operators. A study that can cover both aspects will benefit all the service providers. So it would be valuable to collect and analysis cross-layer variables, which contains high-layer (e.g. QoE) metrics and also lower-layer (e.g. physical layer of the network.) indictors.

As an example, for the web browsing service, studies have found that the long Round-Trip-Time (RTT) is the lethal factor of slowing down the pages’ load speed [2]. Existing studies overcome the long RTT either by client-side optimizations (such as local caching, page pre-fetching, browser parallelization and speculative loading, etc.) or cloud-based optimizations (such as cloud-based preprocessing, cloud-based parallelization and cooperating in the mobile cloud computing, etc.) [3]. However, any one of the existing methods cannot solve the problem thoroughly, and also may lead to negative impacts. For example, it may waste the network resource if the user do not read the pages that pre-fetched by the previous page. In [4], the authors studied the limitations of the client-only approaches. Based on their study, the upper bound of delay reduction for the client-only solutions was just 1.4 second in current 3G network. For the cloud-based optimizations, complex systems and high-cost is required. Generally, there are both shortages for client-based or...
server-based methods. It may be very valuable to explore the wireless link performance between the server and the client, not just consider the client or the server side.

As data mining and big data related technologies attract more and more attention, field data become more and more important for performance analysis and on-going pattern recognition. As the start point to this goal, this study developed a smartphone based network diagnosis tool, which was named Monitor and Diagnosis of Mobile Apps (MDMA). Based on literature review, MDMA is the first tool that can record the applications’ operational state comprehensively. The high-layer metrics (such as, QoE metrics, the speed of the phone, etc.) and the lower-layer (e.g. Transmission Control Protocol (TCP), Radio Resource Control (RRC) state, etc.) were both included. For the truth that Hypertext Transfer Protocol (HTTP) traffic is still the dominating flow in the Internet, in this study, web browsing service was selected to verify the performance of MDMA first. Field data were collected at one selected city of China during three months. A cross-layer database was constructed and the off-line analysis was conducted. The experiments and results verified the usability of the tool, which also can further be used to test other services (e.g. video stream service, on-time chatting service, etc.).

This paper documents the construction of the proposed data collection platform and the variables that can be captured by this tool. Some preliminary findings for web browsing service are also summarized. The rest of the paper is organized as following: Section 2 illustrates the construction of MDMA system. Data collection and cross-layer database also be introduced in this section. Section 3 introduces the cross-layer analysis for web browsing service. In section 4, the study is concluded and future work is summarized.

2. PLATFORM CONSTRUCTION AND FIELD DATA COLLECTION

2.1 Platform construction

MDMA is a synthesized test system (as been shown in Fig. 1) to estimate the performance of WebApps. It is constructed by three modules: User Equipment (UE) based collector, network signaling collector, and off-line analysis server. The first two parts are on-line data collectors. Off-line sever in charge of data storage and post data analysis.

2.1.1 UE-based collector

UE based unit works as an Android application to record environmental data variables and the states of smartphone, including Global Positioning System (GPS) information, velocity of the UE, memory usage, Internet Protocol (IP) address and so on. All this variables will be stored in the light-weight database named SQLite. tcpdump was also be applied to capture the data of network transport layer. The communication packets were stored in “.pcap” files. In order to simulate the user's behavior, a series of script files were edited manually, which can control the phone do the pre-defined actions automatically, such as opening the browser. These scripts were stored on the off-line sever after editing. The installed Application on the smartphone will request specific script according to the requirement of the experiment. Then, downloaded scripts control the smartphone to act human-like activities. At the same time, all related information will be recorded. Snapshots of screen were also captured, which can be used to replay the “user” behavior.

2.1.2 Network signaling acquisition

Network signaling collector was constructed based on Qualcomm Extensible Diagnostic Monitor (QxDM) software [5]. QxDM software provides a rapid prototyping platform for diagnose network performance based on protocol packets. The uplink and down-link radio signaling packets can be captured. The Radio Resource Control (RRC) states, Channel Quality Indicator (CQI) of the wireless network, Radio Link Protocol (RLC) connection state can be studied. Usually, we cannot find any functions or methods for accessing all relative variables directly, such as the RRC state, which is a critical indicator of power consumption and radio resources utilization. In [6], the author presented an algorithm to infer RRC states and states transitions from packet traces, but the accuracy and real-time capability cannot be ensured. For benefits of QxDM, these kinds of variables can be collected directly.

2.1.3 Off-line server

The off-line server contains the control scripts and the data collected by UEs or QxDM. After the server received the data uploaded by the UE-based collector and Network Signaling collector, data variables from multiple sources were extracted and aggregated. The data were extracted from four types of files, i.e. “.db3” file (produced by “SQLite”, on the UE), “.pcap” file (produced by tcpdump, on the UE), “.isf” file (produced by QxDM), and the pictures shoot by the mobile phone (on UE). Then, corresponding KPI (Key Performance Indicators) parameters, environmental factors (e.g.
UE moving speed, internal memory usage, etc.) were identified and stored into the final database. The final database is valuable for network traffic modeling and the network diagnosis.

2.2 Data Collection

In this study, we choose the web browsing service to verify the reliable of the proposed platform. Field data were collected at one selected city in China. Three modes of smartphone were used as UE in this study. Thousands of events were tested in various scenarios, including static scenarios (such as parks, train stations, residential areas) and moving scenes (such as taxis, high speed trains). We selected the target web sites by referring the statistical results on Alexa [7], which provides the popularity of web sites based on traffic data from millions of internet users using over 25,000 different browser extensions. The top 200 web sites in China were selected finally. Based on the statistical results of pages alive time schedules (such as pages' present time period on the screen), we compiled 200 scripts. These scripts can control the cell phone do specified maneuvers automatically. Such as one web browsing event, the scripts can control the phone to open web browser first like a human. Then, the commands "click" one pre-defined web site (e.g. "m.sohu.com", "www.google.com", etc.). In this study, the browser had 30 seconds to load the clicked page. Once the page was loaded successfully, the commands of the scripts will move to the next maneuver (select one page and open it) until the last page. Finally, the browser was closed and this event was done. And at the end of each event, the local cache is cleared. We use this method to imitate the real user's web browsing behavior. At the end of experiments during three months, more than thirty thousand events for each scene were collected. The detailed data variable and results of post data processing are presented in the following sections.
3. CROSS-LAYER ANALYSIS

3.1 Cross-layer Data Construction

Analysis the performance of the wireless mobile internet is not a novel topic. However, how to analyze the features of applications clearly and comprehensively is still need further studies. We classified all collected variables and features into four layers (Fig. 2). The highest layer is QoE metrics layer (L4). The parameters that directly affect the quality of experience (QoE) were studied at L4. The second layer is the packet-based layer (L3). At L3, the distributions and patterns of the HTTP traffic can be studied. For example, the distribution of arrival interval time of packets and the properties of the objects in loaded pages. The third layer is the wireless network signaling layer (L2). This layer was used to describe the network characteristics, which can help us to find the relationship between the cell handover, RRC connection state, and other network state with the higher layer parameters. The results can be applied to diagnosis existing shortages of current network and optimize the existing wireless network. The lowest layer is "user"/UE feature layer (L1). This layer aims at studying the pattern of UE's environmental states, which can the surplus memory of the system, and moving speed of the user and other variables.

For each event, a cross-layer analysis map was depicted based on the final database. From these maps, the data transactions and general states can be identified easily. As Fig. 2 shows, we can find that the load delay of embedded objects is the major delay for the appearance of the whole page. Based on L3, there are two HTTP request failures and one DNS lookup failure can be detected. For L2, three cell handovers were happened and RRC states were changed as IDLE->CONNECTING->DCH->IDLE. Besides, the power consumption can also be estimated by the instantaneous receive/ transmit power. The lowest layer shows that the user's speed changes from 16.3m/s to 20.7m/s during the session. The browser occupies 82 megabytes memory when open the page, then less than 10Mb after loading all the data. There are 20 processes running totally but just have 4 processes access to the network. In the following section, the detailed variables in each layer were illustrated.

3.2 Cross-layer Analysis

3.2.1 QoE metrics analysis (L4)

In table I, QoE metrics for web browsing were listed. Based on these parameters, the quality of user's experience can be estimated. The distributions of these parameters can be the high-level description. Besides aforementioned variables, the factors in the lower layer will also influence the final QoE, such as screen brightness, screen size, environment brightness, etc. All the getable variables were recorded by the various on-board sensors of smartphone.

<table>
<thead>
<tr>
<th>X</th>
<th>QoE metrics</th>
<th>Description</th>
<th>Normalized</th>
<th>Average Times</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>DNS lookup delay</td>
<td>The first DNS request to the valid DNS reply</td>
<td>0.0225 (0.47%)</td>
<td>9.71</td>
</tr>
<tr>
<td>2</td>
<td>TCP Connection delay</td>
<td>The SYN packet to the valid SYN ACK packet</td>
<td>0.0129 (0.11%)</td>
<td>36.76</td>
</tr>
<tr>
<td>3</td>
<td>HTTP Response delay</td>
<td>The HTTP request to the HTTP ok response</td>
<td>0.0430 (0.39%)</td>
<td>32.72</td>
</tr>
<tr>
<td>4</td>
<td>Main objects load delay</td>
<td>The HTTP request to the appearance of main page</td>
<td>2.6284 (36.86%)</td>
<td>3.36 (13 KB)</td>
</tr>
<tr>
<td>5</td>
<td>Embedded objects load delay</td>
<td>The embedded file request to the appearance of these file IDs</td>
<td>4.3405 (63.91%)</td>
<td>26.8 (18KB)</td>
</tr>
</tbody>
</table>

The QoE metrics in table I were normalized by equation (1). \( T_{\text{total}} \) is the page elapsed time, \( T_{x}^{*} \) is the normalized time, where \( x \) is the metric ID in the first column of Table I. As tested, most of the browsers support concurrent multiple TCP connections to one single web domain. However, not all the TCP connections were used to transport data. So these connections were divided to two categories, i.e. connections with data and connections without data. \( T_{\text{up_data}} \) is the total duration of connections that have data, and \( T_{\text{total}} \) is the total duration of all the connections.

The first three QoE metrics in table I could be used to estimate the RTT from UE to DNS server or web service server. DNS lookup, TCP connections, and HTTP connections performance were plotted in Fig. 3. For the truth that Content Delivery Networks (CDNs) and HTTP proxies were usually applied in the data transmission, the data for one web page may be downloaded from more than one web server, which located at multi locations around the world. For this reason, the number of DNS lookup may increase. The average number of DNS lookups for one page is 9.71 and the normalized delay is 22.5ms for the top 200 sites of China (Table I). The proportion of DNS delay in the total delay is 0.47%, which is even greater than the TCP connection delay and HTTP response delay. From some sense, DNS lookup delay should be considered when modeling network traffic and performance analysis.
More details, the RTT to the first hop and RTT to the web server were tested by sending ICMP packets and be compared in Fig. 3. For DNS service, the Round-Trip-Time ($RTT_{udh}$) between the UE to the first hop and Round-Trip-Time ($RTT_{us}$) between the UE to the DNS server have small difference. However, the Round-Trip-Time ($RTT_{wh}$) between the UE to the first hop and Round-Trip-Time ($RTT_{ws}$) between the UE to the web server have significant difference. The $RTT_{wh}$ occupies 70%~90% of $RTT_{ws}$. The average delay of DNS lookup is around 580ms. However, the median RTT to the first hop of the 3G network is just 100ms~150ms, and the medium TCP connection delay is about 318ms. There statistical results suggest that there is a large room to reduce the DNS lookup time for performance optimization. The medium HTTP response delay is about 760ms, which much larger than the RTT to the web server, there is also a large room to improve the performance of the web server. These values also imply that the network performances will be various when collect data at different network equipment (e.g. central network nodes, Node B, UE, etc.).

### 3.2.2 Packets-based layer (L3)

At this layer, more fine-grained analysis for the transport layer traffic can be conducted. TCP header was parsed and used to classify the packets into three categories, which were labeled with different properties (Table II). Besides the normal types of packets, abnormal packets were also identified by parsing TCP header. The normal packets refer to communication packets between smartphone and cell station, which captured when the network was stable and no adverse conditions were detected. The abnormal packets imply the service may be influenced by traffic congestion, server vulnerability, packets reordering or packets loss. RST packets was usually sent by the receiver when the receiver detected packets coming from the closed connections or half-open connections. ACK_DUP was caused by out-of-order data packets or retransmitted data packets. DATA_DUP was usually caused by fast retransmission or retransmission.
timeout. Http_3xx, HTTP_4xx and HTTP_5xx were caused by redirection, client error and/or server error. By parsing the HTTP header, Uniform Resource Indicator (URI), content length (when Chunked-Transfer-Coding not be used), Host name, Referrer name and content type (".htm", ".jpg", ".png", etc.) can be identified. For the reason that data packets captured by tcpdump include packets form all the network-accessed applications, specific strategies must be developed to filter out the packets belonging to one given application. In our proposed platform, the tuple {Host name, Referrer name} was applied to gain this goal. The HOST name refers to the server that stores the resource files. Internet Protocol (IP) address of HOST server and client was identified. Then use the tuple {source IP, destination IP, source port, destination port} to distinguish the traffic packets related to web browser (or other applications). Referrer name specifies the web pages linked to given object. So it can be used to divide different pages among the overall web browsing traffic. Besides these two parameters, object length (refers to the total size of one object) can be used to find the boundary of each object.

### Table II: Different Properties of the Communication Packets

<table>
<thead>
<tr>
<th>Category</th>
<th>Properties</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>DNS_REQ, DNS_RES</td>
<td>The packets to gain the IP address of the web resource server</td>
</tr>
<tr>
<td></td>
<td>SYN, SYN_ACK, FIN, FIN_ACK</td>
<td>Packets to establish or close to the TCP connection</td>
</tr>
<tr>
<td></td>
<td>HTTP_GET, HTTP_2xx</td>
<td>HTTP request packets to get objects and the valid response from the server</td>
</tr>
<tr>
<td></td>
<td>DATA(DATA-ACK)</td>
<td>Normal data packets and the ACK contain the same sequence number</td>
</tr>
<tr>
<td>Abnormal</td>
<td>DATA_DUP, ACK_DUP</td>
<td>The duplicate data and the duplicate ACK contain the same sequence number</td>
</tr>
<tr>
<td></td>
<td>DNS_REQ_DUP, DNS_RES_DUP</td>
<td>Duplicate DNS request packets and duplicate DNS response packets</td>
</tr>
<tr>
<td></td>
<td>RST</td>
<td>Response packets to receiving a packet for a closed socket</td>
</tr>
<tr>
<td></td>
<td>HTTP_3xx, HTTP_4xx, HTTP_5xx</td>
<td>The packets indicate the bad response for the HTTP request</td>
</tr>
<tr>
<td>Others</td>
<td>OTHER</td>
<td>Other packets captured by MDMA</td>
</tr>
</tbody>
</table>

*node: xx, represents an identification number. For example, HTTP_404 indicate the requested web site cannot be found [8].

Based on the aforementioned analysis and statistical results, the existing four-layer on/off model [8] for HTTP service was revised. Based on previous study, one web browsing event can be modeled as on/off process (Fig. 4) at different layers, i.e. session level, page level, object level, packets level. When the user requests the web service, the session is started. Pages' arrival/leave time are described by the page level. Usually, a web page includes one main object and few embedded inline objects. The type of embedded objects can be picture, video and audio. For the importance of DNS lookup in the web browsing service (Fig. 3), it was incorporated into the object level. In this way, the UE gets the IP addresses of the content servers just like acquiring the objects from the content server. According to HTTP 1.1 standard RFC2616 [9], after downloading the main object, there are up to 2 TCP connections can be established in parallel for downloading the embedded inline objects. This phenomenon was also verified by field data collected in this study. So we modify the object layer to parallel structure, like Fig. 4 shows.

#### 3.2.3 Network signaling analysis (L2) and “User” feature layer (L1)

For the benefits of QxDM, the network signaling messages can be collected. Such as the network signaling messages to setup the Radio Access Bearer (RAB) or reconfigure the existing bearers periodically. By parsing these messages, the performance and actual conditions of the physical layer can be estimated.

MDMA can also be applied to record external environmental factors. Such as the speed, GPS locations, system timestamp, temperature, screen lightness. Although these kinds of parameters have no direct relationships with network performances, QoE of users would also be influenced significantly by these external factors indirectly [10]. Besides aforementioned benefits, locality and time information can also be considered when modeling higher-level network traffic flows.

As an example, Fig. 5 shows the cumulative distributions of total load delay ($T_{load}$) of "m.sohu.com" for different stations. It's easy to find that the total delay at railway station and bus station is much longer than other scenarios. This may be caused by the high number of users at these areas. Based on L2 data, the Common Pilot Channel (CPICH) power ($Ec/lo$) for the serving cell was calculated for each scenario (Fig. 4). This measurement also imply that railway station, bus station area have been covered with poor network performance. The combination of different layers can imply more interesting findings for given scenarios. It's should be emphasized that the mentioned findings presented in this paper are used to show the potential capabilities of our proposed platform. Much more applications can be tested.
4. CONCLUSION

This paper presented a smartphone-based cross-layer field data collection platform, which can be applied to acquire wireless network traffic data generated by WebApps automatically. The construction of the platform was illustrated. As a testing example, web browsing service was selected to verify the performance of our proposed platform. Communication data collected by two modules (smartphone-based unit, Network signaling acquisition unit) of the platform were grouped into four layers, i.e. QoE metrics, packets-based layer, network signaling layer and scenario feature layer, respectively. Based on off-line data analysis, for the existing HSDPA network, DNS lookup delay should be considered in the future studies. Detailed variables and potential capabilities of the proposed platform were presented hierarchically. Based on the up-to-date results, the platform can be used to capture high level parameters and also low level factors, including network performance related factors and environmental factors. This platform has the ability to test existing wireless networks performance by client-side mode. Future work will expand our database by collecting extra field data at different time period and for different wireless networks (such as the fourth generation of mobile telecommunications technology, 4G).

5. ACKNOWLEDGMENTS

This study is supported by the National High Technology Research and Development Program of China (863 program: 2015AA01A706).

REFERENCES


Proc. of SPIE Vol. 9902 99020F-7
Downloaded From: http://proceedings.spiedigitallibrary.org/ on 08/14/2017 Terms of Use: http://spiedigitallibrary.org/ss/termsofuse.aspx