Traffic Measurement and Analysis in Cellular Data Charging System

Yong Piao, Xiaolan Liu and Jiayun Li
Tsinghua National Laboratory for Information Science and Technology
Dept. of Computer Science and Technology, Tsinghua University, Beijing, China
{poy12, xl-liu12, lijiayun12} @mails.tsinghua.edu.cn

ABSTRACT

In recent years, the development of 3G/4G networks makes people pay more attention to the traffic accounting in data-based charging cellular networks. Although this charging system has been worked many years and updated frequently, we find that there are still some points which can be improved. First, we measure the traffic in poor signal between end devices and base station by different operators, discovering some billing loopholes and overcharge. Our experiment shows the overcharge can be 28.4% of the data we want. Then, we analyze the traffic through phones in one extreme case, and detect that the retransmission only take place no more than 10% of the whole traffic. While, the reconnection volume can be three times by the goodput. At the end, we propose feasible solutions and discuss possible future directions.

KEYWORDS

Charging, TCP, Cellular Networks, Badput, Reconnection.

1 INTRODUCTION

People gradually access to data services by mobile devices instead of traditional means, thanks to the development of 3G/4G networks. Statistics from mobiForge[1] shows that, during the whole 2014, worldwide smartphone shipments are predicted to reach 1.3 billion units. This figure is expected to grow to 1.91 billion next year and to 2.16 billion in 2016. In terms of traffic, US users' average data usage is 2GB per month, while this number is 149MB only 6 years ago which is almost the same with today's China that owns the largest number of mobile users[2]. These two reasons lead to the world traffic exploding.

While the bill we pay is also growing and not cheap even now. Although the average cost of data has dropped down, Chinese total cost annually increase 40%[2] which is much larger than user growth. Most operators charge users based on their monthly data bill. People get data through buying traffic package. This bill would be cleared at the first day of a month. Because of the large cardinal number of users and amount of traffic, even 1% of the monthly waste in China will result in the charge of million dollars. Therefore, saving traffic is getting more and more important.

<table>
<thead>
<tr>
<th>Time</th>
<th>Type</th>
<th>Traffic</th>
</tr>
</thead>
<tbody>
<tr>
<td>2014-10-05 15:30 pm</td>
<td>Goodput of the Web page</td>
<td>1.8MB</td>
</tr>
<tr>
<td>2014-10-05 15:30 pm</td>
<td>UE side</td>
<td>6.9MB</td>
</tr>
<tr>
<td>2014-10-05 15:12 pm</td>
<td>BS side</td>
<td>7.5MB</td>
</tr>
</tbody>
</table>

Figure 1. Three kinds of traffic for HTC Web page tracked by the browser, Bob’s phone and official sites

However, existing charging technique is not fair to the users sometimes. This charging module working between the base station(BS) and Internet is completely independent of user equipment(UE). All operators charging vastly different[3] and may hide some deadly loopholes[4]. So it is obvious that what we pay is not what we get in same cases[5]. For example, Bob wants to connect HTC home page which is 1.8MB. He step to a weak signal area and click on the button. At the same time, TrafficStats[6] is used to record the data in Bob's phone. What he would pay is exhibited on the operator online business hall[7]. When Bob has completely get the whole resources...
from HTC service, three kinds of traffic has been tracked and shown in Figure 1.

He repeated this experiment in different place and found that: these three kinds of traffic of Web page usually differ, the traffic bill is much larger than what he finally get in some specific conditions(e.g., poor signal power and rapid mobility) and the official data is not that 'official' sometimes. So we naturally come up with three questions:

(a) Which one of the three should we pay? (A\textsuperscript{1}, B\textsuperscript{2}, C\textsuperscript{3} or)
(b) What causes the user to pay 5.7MB more?(C-A)
(c) How to reduce or avoid these factors that cause extra traffic accounting?(minimize C-A)

Question 1 includes mass of user-operator conflicts: (1) Users tend to pay for the data they really want(Traffic A) instead of the data they get(Traffic B). While the operators charge them by the Traffic C. (2) It is evident that some of the Traffic (B-A) is also helpful to transport the goodput(Traffic A), like TCP retransmission and three-way handshake, some are gratuitous because of poor signal and handoff. Here goodput refers to the data user want to get from application layer excluding the MAC overhead, the retransmission overhead, the acknowledgment (Ack) transmission and so on. (3) The part of traffic (C-A) wears down BS's equipment and costs lots of energy\cite{8}. On another hand, the operators can't easily distinguish Traffic A from C in an ordinary connection. If Traffic(C-A) is not charged, operators have to bear the risk of traffic attacks\cite{9, 10} and pay for the replacement of devices and energy power. (4) In this paper, we find that the error of (C-B) is mainly caused by the leak of accuracy in charging system and the improper setting of BS. Because of this, users feel that what they get is not what they pay. In conclusion, this paper suggest user to record the traffic they get from the application layer and operators to keep track of Traffic C which passes by the BS. What's more, BS is obliged to reduce the gap by B and C by optimizing system to make its charging accurate and impartial.

Actually, Question 2 and 3 can be divided into two parts: (C-B) and (B-A). If the operators have achieved a precise Traffic C which is approximate to B, Question 2 and 3 is transformed into what causes (B-A) and how to reduce it. In order to decrease the needless part of (B-A), we must need to know the structure of B in all kinds of situation. About the mobility, \cite{11} has conduct series of mobility experiments and proposed that buffered packets in BS would be lost when handoff is performed. We just focus on the situation of static data waste.

Our study yields two main findings. First, we observe that there are some problems between UE and BS. At the UE side, two primary traffic monitor tools(TrafficStats and TCPdump) can't trace accurate data. Each of them failed in different scenarios. At the BS side, Existing charging based on flow have time and space errors. BS could make mistakes at the junction of the end and beginning of the month or the junction of idle and busy time of a day. It can also overcharge people for 15% of their goodput traffic. Second, we have measured many apps' traffic, and find out TCP retransmission only take a little place of the goodput data volume(less than 15%). Among this, spurious retransmission account for less than 10%, and packets loss due to bufferbloat\cite{12} is no more than 5%. During the experiment, there is a more serious problem that may waste data far more than retransmission. It is reconnection. We discover that if reconnection starts, the data which has transferred before would be cleared both on the UE side and the server side. But this data has already been counted in the user bill. In extreme cases, it cost 316% or more than the size of what we want.

The rest of the paper is organized as follows. Section II simply introduces 3G/4G charging.

\textsuperscript{1} In this paper, A represents the goodput traffic(first item in the Figure 1)
\textsuperscript{2} B represents the traffic measured by UE side(second item in the Figure 1)
\textsuperscript{3} C represents the traffic measured by BS side(third item in the Figure 1)
background. Section III shows our experimental methodology. Section IV and Section V describe the two questions (Question 1 and 2) and analyze the reasons. Section VI provides some discussions and solutions about those two questions. Section VII compares with the related work.

2 BACKGROUND

There are two ways of charging in the world. One is based on the traffic, the other is based on time. Almost all the countries record bill by how much data users consume in one month. So we only study the architecture of data-based method in this article.

In this section, we simply present how the charging system works. First, it is necessary to know the structure of the operator networks. We focus on the most widely used 3G technology: Universal Mobile Telecommunications System (UMTS) [13]. The 3G High-Speed Packet Access (HSPA) and 4G Long Term Evolution (LTE) [14] charging systems are similar to UMTS. Second, we present how charging system works in the core networks.

![3G/4G network architecture and charging components](image)

2.1 Charging Architecture

The 3G UMTS network architecture consists of two parts. From Figure 2, we can see the wireless part and the wire one. The wireless part Terrestrial Radio Access Network (RAN) contains the User Equipment (UE), the base station ((NodeB for 3G, eNodeB for LTE)), and the Radio Network Controller (RNC). Before surfing the Internet, UE must hold a connection with 3G base station in order to get IP address, base band, and so on. But the most important units we talk about is the wire one between the RAN and Internet, core network (CN). It bears charging function and can process users’ and servers’ data.

When the base station receives a packet from UE, it pass the data to the CN’s Serving GPRS Support Node (SGSN) or Serving Gateway (S-GW). In 3G cellular networks, SGSN and Gateway GPRS Support Node (GGSN) are the major components of CN. They work like interface between CN and extra network.

The charging system contains three charging components: the Billing Domain (BD), the Online Charging System (OCS), and the Charging Gateway Function (CGF). Now, the cellular networks support both online and offline charging modes [15]. When a packet come to the SGCN, BD records the time, ID, and other important charging information in order to support OCS and CGF to process different models of charging. In LTE networks, the Mobility Management Entity (MME) is in charge of tracking the UE and call the appropriate S-GW to serve the UE.

2.2 Accounting Process

CN use Charging Data Record (CDR) to identify every data bill. The CDR keeps necessary information such as the user’s ID, the network elements, and the session. The user sends a Packet Data Protocol (PDP) Context Request message to the SGSN to trigger the PDP context setup process before the CDR creation [16]. Because of this activation, the smartphone is allowed to link the Internet by the SGSN and GGSN. At the same time, the charging program is launched, and GGSN assigns a sole charging ID to the activated PDP context. The CDR is created to be ready to record the traffic through the CN.

For data-based charging method, there are two main mechanisms: offline and online. In

---

4 PDP contexts provide all the required information for IP packet data connections in cellular networks.
the offline charging condition, the charging information is collected concurrently with the data itself. This method don't need to ask whether this user has enough paid resource left. However, online charging must get the authorization for the network resource like duration or the data volume. During online charging, OCS's Online Charging Function(OCF) get a charging event from Charging Trigger Function(CTF), and feedback the authorization for the resources. When the available credit decrease to zero, the base station would cut down the connection.

3 PROBLEM AND METHODOLOGY

In this section, we enumerate some possible problem in cellular networks. To actually find and solve these problems, we have taken some experimental methods.

3.1 Possible Problems

In the second section, we introduced the charging process of cellular networks. These features of charging process and TCP/IP protocol inevitably leads to the following questions:

(1) Traffic mismatch between UE and BS (B!=C)

The charging process of user is independent of CN, there is no feedback mechanism between the two, which would lead to the actual resource usage and what they pay for do not match.

Although all the 3G or 4G network use the same charging mechanism, different ISP have nuances. These differences could lead to some charge trap.

(2) Badput\(^5\) in traffic (B-A)

Bob has to pay 5.7MB more than what the size of the Web page. If the volume measured by BS is accurate (B=C), whatevless could make users pay more than what they get. If badput represent this part of data, how we minimize it to make the users obtain these resources with less cost.

3.2 Experimental Methodology

We conduct these experiments in two major Chinese operators, denoted as OP-I and OP-II, for privacy. These two ISPs cover more than 80% Chinese mobile market[17]. We use 5 phone models of HTC g10, xiaomi Mi3, Huawei X1, apple iPhone6 and 6 plus, containing two kinds of mobile OS: iOS and Android. The applications(e.g., UC browser, Baidu cloud) we select owns hundred millions of users in China. For further performance analysis, we set up a FTP server and measure both the UE side and the server side.

Before the experiment, we collect and develop several tools to help: (1) TrafficStats: This is an Application Programming Interface(API) provided by Google to record the volume of every application working on the Android system. (2) TCPdump: We use it to capture the traffic on the Android and iOS phones. (3) Wireshark: We use it on our serve for packet capture traces, and analyze the .pcap file cached by TCPdump on the smartphones. (4) Tcpprobe: We transplant traditional Tcpprobe to Android in order to measure the smartphone TCP states on kernel level. (5) Others: Some apps are provided to help users to know the signal power, the Network Type, and so on.

To test different situation, we conduct our experiments in various applications and places. From the bottom layer to the first floor in the east of the main building of Tsinghua University, the scope of the signal from negative infinity to -60dbm. First we have to choose a fixed position, measure signal strength, and then begin to experiment.

4 UE-BS CHARGING

In this section, we pay attention to the traffic between UE and BS. Because BS record the traffic, and don't feedback to the mobile device, the UE side can't know whether the data bill is exactly what they have transmitted. However, we have found that there are some problems

\(^5\) In this paper, badput represent the traffic of (B-A)
Figure 3. Nine kinds of experiments to show the goodput and traffic traced by TCPdump, TrafficStats and operator online business hall.

between these two sides: (1)Space-time Error (2)Overcharge.

4.1 Space-time Error

As shown in Section II, BS side keep a CDR for every connection. But it is not sure that the BS records are accurate. In this part, we try to download a file or visit a fix Web page under different signal, and find that the records have space and time errors.

In order to make reliable experimental data, we use tree different phones and repeat more than 5 times in one place, and record the experimental parameter concurrently. We use TrafficStats and TCPdump to keep track of the traffic on the UE side, and search the operator online business hall for the bill on the BS side. At the end, we use Wireshark to analyze the packets captured by TCPdump.

Figure 3 is the result of these experiments. Horizontal axis represents different kinds of tests. The number 1,2,4,7 of tests stand for uploading and downloading a 7.1MB file by Baidu cloud. The number 5,6,8,9 of tests stand for refreshing the HTC home page(1.8MB). All of these tests are made in poor signal less than -70dmb. From the figure, we can see TrafficStats is not that precise in the poor power. It even overestimates more than 100% in the extreme situation. We propose that TrafficStats is not a reliable traffic analysis tool, so packet capturing become our main means in the next measurement.

Figure 4. We download two files with same size started by point B and C. The yellow lines represent the points recorded by operator. The black lines represent the actual value.

No.9 experiment was done immediately following the eighth experiment in time. But the bills are not slightly higher than the TCPdump like what we expect. The No.8 on BS side is much higher than the UE, but the No.9 bill is much smaller. However, the total values of them are close. We think the bill in No.8 has "stolen" some data from No.9. In a series of experiments, the BS side sometimes records the wrong beginning and end of a flow. Figure 4 shows this process. We start download file 1 and file 2 at point B and point C, but the record of the BS bill shows the start points are point A and point D. Red line represent the real
traffic. So the bill of fist file (10.2MB) is larger than the second (4.3MB). It is not important in the usual time, but can cost troubles at the junction of the two charging modes like busy-leisure time of a day and end-beginning of a month.

4.2 Overcharge

In Figure 3, No.1 represents upload by Baidu cloud, and No.2 represents download. Unusually, No.1 and No.2 tests' bills are both a little bigger than the TCPdump. Through expand experiments, we fond this overcharge would be larger if the power become worse. It is mainly because the lack of communication between the BS side and the UE side. Even if the TCP protocol can ensure the reliability of the transmission, BS will have overcharge users, even beyond the 28.4%.

5 BADPUT IN TRAFFIC(B-A)

We already study the gap of Traffic B and C above. In this section, we assume that the BS bill is accuracy as the UE side, and measure what would make us pay more and how much we pay for this badput. For example, Figure 5 shows the bills of a Web page of 1.8MB in different environment. In strong power LTE network, the bill is only 2.012MB, but it is almost four times in poor signal 3G network. What on earth caused this big different?

![Figure 5](image)

**Figure 5.** The traffic captured by TCPdump when browsing a 1.8MB HTC home page.

Let's analyse the last item by Wireshark in Figure 6. We count all the 7.1MB traffic and put them into different classes. In this picture, we can see that the goodput (1.8MB) takes only 34% of the bill. Contrary to our expectation is that the part of spurious retransmission is only 4% of the whole traffic, and the biggest part comes to be the data wasted by multiple reconnections. We examine the reasons for these parts and produced their share of the range below.

![Figure 6](image)

**Figure 6.** The structure of traffic traced by browsing HTC Web page.

5.1 Spurious Retransmission

Spurious retransmission mainly caused by out-of-order of packets. When radio channel quality is poor, large sequence number packet arrives earlier than the small sequence number packets, fast retransmission mechanism can’t tell it is out-of-order or packet loss, so it would sent Dup ACK to ask retransmission. So, in the poorer signal environment, the more out-of-order packets are. But in the measurement, this proportion of goodput is still small even in extreme environments.

<table>
<thead>
<tr>
<th>NO.</th>
<th>Network Type</th>
<th>power (+/-1dBm)</th>
<th>ISP</th>
<th>Smart- phone</th>
<th>traffic (KB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4G TD-LTE</td>
<td>101</td>
<td>OP1</td>
<td>HUAWEI X1</td>
<td>2012</td>
</tr>
<tr>
<td>2</td>
<td>4G TD-LTE</td>
<td>106</td>
<td>OP1</td>
<td>HUAWEI X1</td>
<td>5190</td>
</tr>
<tr>
<td>3</td>
<td>3G UMTS</td>
<td>80-85</td>
<td>OP1</td>
<td>HUAWEI X1</td>
<td>4280</td>
</tr>
<tr>
<td>4</td>
<td>2G GPRS</td>
<td>69</td>
<td>OP1</td>
<td>HUAWEI X1</td>
<td>4980</td>
</tr>
<tr>
<td>5</td>
<td>3G UMTS</td>
<td>120-inf</td>
<td>OP2</td>
<td>HTC g10</td>
<td>7169</td>
</tr>
</tbody>
</table>

**Figure 7.** The average retransmission rate in different signal environment measured by four popular applications.

We conduct many experiments at the same place using three phones (HTC g10, Huawei X1, and iPhone 6 plus). Figure 7 shows the average rate in different applications including FTP download and upload, Web page, Baidu cloud, and pptv. In the normal power, retransmission...
rate is 0.5%. This number is 5% in almost no signal place.

Obviously, this value is little compared to other badput. We originally think when the power gets very poor, lots of packet loss and out-of-order would occur in the wireless link. However both these two rates are less than 5% of goodput.

![Figure 8. Queue length in bufferbloat. The process last 600 seconds, and the loss rate is 0.17%.](image)

5.2 Bufferbloat

If BS has big buffer, it could throw packets when the buffer if full. We use NS2 to simulate the relationship between packet loss and buffer size, and discover that the packet loss rate is still very small (less than 2%). If we adopt bufferbloat avoidance mechanism, the rate would be close to 0. Figure 8 shows the queue length in a long flow connection after without any avoidance mechanism. X axis is connecting time. Y axis is queue length. We can see the size of buffer can hardly reach full size.

When the buffer is small (less than 10), the packet loss rate is over 5%. That's mainly because the buffer would be easily filled and after then throw next arriving packets. But in reality, this situation is impossible. In addition, we can set buffer a proper size to reduce the packet loss rate as little as possible.

![Figure 9. Four tests on Baidu cloud. The red line means the goodput is 7.1MB](image)

5.3 Reconnection

The traffic captured by TCPdump reveals that the retransmission rate is poor in the cellular networks. However the BS bill is quiet big. We check every connection and find some connections break and reconnect frequently and previous data retransmit many times. This part of data looks normal in the Wireshark, but would be charged by BS side.

Network quality certainly affects the reconnection probability. We try to download a file of 7.1MB by Baidu cloud in Figure 9. The first pillar is the size of file. Then, next three times shows the file has been completely downloaded. But the fourth attempt is failed and produces 43.8% badput traffic. It also shows that Baidu cloud doesn't support breakpoint resume.

In addition, connecting number is important to the badput caused by reconnection. HTC home page has more than thirty resources. HTTP protocol would establish a connection for each resource. Multi-connections increase the fail probability in the same quality of network. If one failed, it would initiate new connection instead of stopping like Baidu cloud. Figure 10 shows the badput in four trials. We
can see the badput rate caused by reconnection play a decisive role in the whole transmission.

According to the above features, we can set up a method to reduce the rate of failed connection. When the phone asks to get a Web page with many resources, we can pack these resources up to a file, and transmit it in one connection. Because one connection's failure probability is much lower than multi connections. When this connection breaks, users can choose whether the task goes on.

6 DISCUSSIONS

In this section, we discuss some feasible ways to reduce error between UE and BS and unnecessary badput. First, we must know what should we be charged for? The proposal bill is Traffic C. But it has sometimes space-time error and overcharge. Space-time only make sense when changing the billing pattern. On the side of user, we can temporarily stop data transmission. On the side of BS, the operators are obligated to reduce billing errors, and avoid the charging trap. The reasons of overcharge is not clear now, so we suggest reduce time of Internet when signal is really poor.

Reconnection is the biggest part of badput. It is affected by power and connecting number. UE is able to get this information, then reminds user that this task could lead to much badput traffic. Another way is build a proxy between the Internet and the BS. The proxy can provide users a mechanism to avoid repeated reconnection. When a user wants to get a Web page with so many resources, the proxy can download the whole resources and pack up to one file. Then, the page can be transmitted as one resource. This method could greatly reduce reconnection rate.

7 RELATED WORK

[4, 9] has measured and analyzed two loopholes between UE and BS about UDP protocol. (1)They move to a no-signal scenario with ongoing downlink UDP transmission from an Internet server to the UE. Three hours later, they were charged for 450MB. (2)They are able to transfer any amount we specify for free in an extreme case. But UDP only take place a very small part of total traffic, and the case's condition is very rigors.

Then, [10] researched TCP retransmission in different countries, and proposal that do not charge for TCP retransmission. In that paper, the Retransmitted data is created by the server they built, but there is not that much retransmission in the real environment. If we don't charge this part, the operator could bear great cost, like packet processing and electric power for the retransmitted data[8]. However, there is no research on the aspect of how to help users to save traffic.

REFERENCES


[16] 3GPP.TS23.060, “General packet radio service (gprs); service description; stage 2.”