Concise Paper: TCP Performance over Mobile Networks in High-speed Mobility Scenarios

Qingyang Xiao†, Ke Xu†, Dan Wang‡, Li Li‡, Yifeng Zhong*†
* Tsinghua National Laboratory for Information Science and Technology (TNList)
† Department of Computer Science and Technology, Tsinghua University, Beijing, China
‡ Department of Computing, The Hong Kong Polytechnic University, Hong Kong, China
Email: xqy13@mails.tsinghua.edu.cn, xuke@tsinghua.edu.cn, csdwang@comp.polyu.edu.hk,
ll-12@mails.tsinghua.edu.cn, zhong-yf11@mails.tsinghua.edu.cn

Abstract—Recently, the performance of mobile data networks has been evaluated from many aspects, e.g., TCP/IP protocols, comparison with WiFi or even satellite communication, under different movements within a metropolis area. Nevertheless, the result is still unknown in high-speed mobility scenarios and in a scale that crosses different metropolis and geographic areas. To fill in this blank, we carry out a comprehensive measurement study on the performance of mobile data networks under high-speed mobility, i.e., 300 km/h or above. Such speed is the current de facto standard of the China Railway High-speed (CRH) network, the largest commercial high-speed railway network in the world so far. We first present an overview on the TCP performance over LTE networks. We observe that decent throughput may exist under high-speed mobility. However, comparing to the stationary and driving (100 km/h) scenarios, the throughput and RTT not only are worse, but also have a large variance. We then take an in-depth investigation into two key factors affecting the performance, i.e., the wireless channel and handoff. We believe our study on these factors is useful not only for TCP, but also for other upper-layer protocols.

I. INTRODUCTION

Mobile data networks are now expanding their domains to every part of our life. In daily life, mobility is a basic requirement, including walking, driving, taking the subway and other public transport. Recent advancements, e.g., 3G/LTE, in mobile data networks try to provide high data rates and seamless mobility support for most of these scenarios.

Along with these development and deployment, researchers have investigated the performance of mobile data networks from many aspects. Tan et al. analyze the performance of UMTS networks in the stationary environment [1]. Tso et al. conduct a large-scale test on HSPA networks to study low-speed mobility impact [2]. Huang et al. carry out a measurement study to quantify the impact of mobile speed for a commercial LTE network [3]. Nevertheless, little is known about the performance of mobile data networks under high-speed mobility scenarios. With the increasing expansion of high-speed trains, especially in China, it is an important task to understand the performance of mobile data networks in such scenarios. High-speed mobility has unique features distinguishing it from other low-speed ones, requiring some existing topics to be reexamined. This motivates us to investigate the network performance and user experience under high-speed mobility.

In this paper, we carry out a measurement study on the performance of mobile data networks in high-speed mobility scenarios, i.e., 300 km/h or above. Such speed is the current de facto standard of the China Railway High-speed (CRH) network, the largest commercial high-speed rail (HSR) network in the world so far. In particular, we focus on the Beijing-Tianjin Intercity Railway (BTR). In our experiments, we study the performance over LTE networks of China Mobile, one of the largest carriers in the world. To obtain specific information of network usage, we develop a measurement tool on the Android platform, MobiNet. For comparison, we also conduct experiments for the stationary and driving (100 km/h) scenarios. To have a fair comparison, our driving follows the same path of BTR. We seek to uncover the speed limit for current cellular networks.

Note that, we use TCP as a working example to study the performance of LTE networks under high-speed mobility. Clearly, TCP is the most important protocol, which supports tons of application protocols. We believe understanding certain key parameters of TCP performance is of crucial importance for the design and optimization of upper-layer protocols. However, we strategically try to avoid protocol-specific conclusions. We believe our results are useful for TCP, as well as the design of other upper-layer protocols, for example, the Real-time Transport Protocol (RTP) over UDP, or the protocol used in Skype [4]. We consider that a more TCP-related study, which may include the impact of high-speed mobility on TCP handshake, TCP performance of different flow sizes, and TCP break-down operations, worth a separate study.

Based on our measurements, we first present an overview of the TCP performance over LTE networks under high-speed mobility. We show the throughput and RTT of TCP. Compared to the stationary and driving scenarios, the throughput is worse. Still, we observe that there can be a decent throughput of 1000 kbps. This is a fine throughput for almost all applications to date. However, there is a large variance in both the throughput and RTT. In many occasions, the throughput drops to zero. The RTT can easily become greater than 100 milliseconds, whereas in most driving scenarios, the RTT is only 10 milliseconds.
We then look into two key factors underlying the performance, i.e., the wireless channel and handoff. We find that for the wireless channel, the signal strength becomes worse. We conjecture that this is due to the Doppler effect. There are even coverage holes in 2G EDGE networks only. We further investigate the handoff, and find that there are both inter-system handoff and intra-system handoff. In addition, the handoff frequency is much greater. In particular, 45.1% of handoffs can be less than 10 seconds. This may seriously affect TCP retransmission, etc. As a matter of fact, we observe that there are periods when TCP transmission suspends, i.e., no packet sent or received.

To the best of our knowledge, this is the first public large-scale empirical study on network performance in high-speed mobility scenarios. Our empirical study results are useful for network operators to provide extensive support for high-speed mobility. It is also meaningful for network protocol and application designers to take high-speed mobility condition into consideration. Our work makes a first step to explore these issues focusing on high-speed mobility. We also add driving forces to this study from two aspects. One is the popularity of HSR, and the is the popularity of mobile broadband services (i.e., data access).

II. RELATED WORK

There is great interest in the performance of mobile data networks. We categorize the related work into network performance, mobility and handoff.

Network performance: Tan et al. investigate the performance of UMTS networks in terms of throughput, latency, video, and their ability to provide service for different traffic classes under various loading conditions [1]. Huang et al. analyze general performance with various impacts, including signal strength, time of day and carriers [5], [6]. In addition, to measure network usage and performance, MobiPerf [7] tests network throughput and latency.

Mobility: Tso et al. present an empirical study on the performance of HSPA networks in Hong Kong via extensive field tests, including many possible mobile scenarios in urban areas, i.e., subways, trains, off-shore ferries and city buses [2]. A. Schulman et al. show that signal strength varies across locations and has a non-negligible impact on the performance [8]. P. Deshpande et al. show that over a wide geographic region and under vehicular mobility, networks exhibit very different throughput and coverage characteristics [9]. Measurements have been carried out at different mobile speed [3]: stationary, 35 mph and 70 mph. It shows that there is no major performance downgrade at different speed for LTE networks, and RTT remains stable at different speed with small variation. This greatly differs from our observations.

Handoff: Tu et al. analyze how mobility affects mobile data accounting and discovers the root cause of handoff [10]. Studies [11] and [12] analyze the handoff influence in detail based on the ns-2 simulator and prove handoff has a bad impact on network performance. K. Jang et al. examine the impact of high-speed mobility and show that mobile nodes have far worse performance than stationary nodes in the same condition [13]. Our work shows that variance (instability from the upper-layer point of view), rather than the sheer throughput, is the key problem. We investigate the root causes on wireless channel and handoff.

III. EXPERIMENT METHODOLOGY

In this section, we present our methodology for measuring mobile network performance over 4G LTE networks. Our dataset is based on a series of network tests at different mobile speed.

A. Measurement Tool

To conduct an in-depth study and collect packet traces, we develop an Android application named MobiNet for measuring cellular network performance. MobiNet achieves a series of network tests on throughput and latency, as well as other useful network metrics. MobiNet also records the variation of signal strength, the information of base stations, fine locations based on GPS, mobile speed, data connection state, and handoff occurrence. We trace communication route from clients to local servers and record the ping latency of each hop.

Our test algorithm is based on iperf [14], a famous tool to measure the maximum TCP bandwidth. Our work improves the packet-sending method in order to control the TCP flows by limiting the transmission rate and TCP window size. In our experiments, we generate four types of traffic: TCP long-lived traffic, TCP intermittent traffic with different flow sizes, concurrent TCP traffic, and UDP traffic with constant bit-rate.

Further, we have shared MobiNet to the public1.

B. Measurement Setup

To explore the impact of high-speed mobility, we carry out a comprehensive measurement study on TCP performance over LTE networks using MobiNet. We conduct three types of evaluations: downlink only, uplink only, as well as simultaneous downlink and uplink. At the same time, packet traces are collected from both sides of the client and server. We take throughput and latency as our main metrics. Through a fine-grained packet analysis, we calculate the throughput and RTT. In addition to packet traces, we collect real-time data connection state, network type, and the variation of signal strength. We also record the mobile speed, the information of base stations, cell identifier (CID), and location coordinate based on GPS (i.e., latitude and longitude).

During the test, we let smartphones communicate with the benchmark server directly via TCP/UDP connections. Our deployed server can provide a sufficient bandwidth of 100 Mbps. Thus we ensure the server bandwidth is not the bottleneck in our tests. In the experiments, we use four popular Android phone models as main devices, including Samsung Galaxy Note 3, Samsung Galaxy S4, Google Nexus 4, and HTC Desire, with the same runtime environment. During the experiments on trains, all the devices are set on the table close

---

1Our measurement tool is available at: http://www.thucsnet.org/projects. html
TABLE I

<table>
<thead>
<tr>
<th>Test location</th>
<th>Stationary</th>
<th>Driving</th>
<th>HSR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mileage (km)</td>
<td>20</td>
<td>115</td>
<td></td>
</tr>
<tr>
<td>Duration (min)</td>
<td>5</td>
<td>12</td>
<td>33</td>
</tr>
<tr>
<td>Speed (km/h)</td>
<td>0</td>
<td>100</td>
<td>300</td>
</tr>
<tr>
<td>Test times</td>
<td>6</td>
<td>6</td>
<td>42</td>
</tr>
</tbody>
</table>

We plot the results of TCP throughput in Fig. 2, where Fig. 2 (a) (b) (c) show the stationary, driving and high-speed mobility cases respectively. The vertical dotted lines represent where a handoff occurs. Clearly, we can see that the throughput of the stationary case is the best, whereas under high-speed mobility, the throughput is much worse. TCP downlink throughput on HSR is only 575.8 kbps on average. For comparison, we observe a throughput of 1663.4 kbps in the stationary scenario and 1509.5 kbps in the driving scenario. We can also see that under high-speed mobility, the throughput can reach 1000-1500 kbps from time to time. As a matter of fact, such throughput, if stable, can actually satisfy most applications to date. Yet, there is a large variation in the throughput. Most notably, we see that there are many cases that the throughput drops to nearly zero. Such variation, if not understood or handled appropriately, may significantly affect the performance of upper-layer protocols.

We next show the RTT under high-speed mobility. In Fig. 3, we show the CDF of RTT. We see that 11.4% of RTT can be as long as 0.4 seconds. In such scale, the RTT can be beyond the tolerance of TCP. For comparison, in Fig. 3(b), 99.6% of the RTT in the stationary scenario is less than 30 milliseconds and 97.3% of the RTT in the driving scenario is less than 30 milliseconds.

We next show the RTT jitter in Fig. 4. We see that for stationary and driving at the speed of 100 km/h, for 90% or above, the RTT is only 7.6 milliseconds, yet for high-speed train, 56.6% of the RTT is greater than 100 milliseconds. In addition, the RTT variance is 1636 times greater than the other two scenarios. Note that an average RTT of 200 milliseconds may still be tolerable (we measure a connection between Beijing and New York can be 214.9 milliseconds), yet the large RTT jitter may seriously affect the performance.

Our conclusion is that the variance, rather than the sheer throughput or RTT is damaging for the wireless data network performance under high-speed mobility. In this paper, we are not trying to provide solutions, rather, we are looking into the root causes of such performance. We discuss wireless channel and handoff respectively.

V. ROOT CAUSE I: WIRELESS CHANNEL

We first present the wireless signal strength, which is the underlying support for wireless communication quality. There are many metrics for signal strength, such as RSSI, Ec/Io, SNR, RSRP (Reference Signal Receiving Power), and RSRQ (Reference Signal Receiving Quality). Each of these metrics is used for different scenarios, in particular, RSRP is used primarily for LTE networks [15]. In our work, we focus on LTE services; as such, we use RSRP as our major signal strength measurement metric. There are three levels in RSRP, Good (-85 dBm), Average (-95 dBm), and Weak (-105 dBm).

Fig. 5 shows the signal condition of the BTR line. Fig. 5(a) shows an experiment from 6:50 am to 7:23 am and Fig. 5(b) shows an experiment from 8:31 am to 9:04 am, both from Beijing to Tianjin. The three red lines in the figures show the Good, Average, Weak indices of RSRP.

We also conduct measurement by driving a car. To have a fair comparison, we choose Beijing-Tianjin Freeway (BTF). This freeway is in parallel with the BTR line for a subset area, as shown in Fig. 1. The total time for the high-speed train to complete this journey is 33 minutes. Besides the starting and ending phase, the speed of BTR reaches 300 km/h.

We next show the RTT under high-speed mobility. In Fig. 3, we show the CDF of RTT. We see that 11.4% of RTT can be as long as 0.4 seconds. In such scale, the RTT can be beyond the tolerance of TCP. For comparison, in Fig. 3(b), 99.6% of the RTT in the stationary scenario is less than 30 milliseconds and 97.3% of the RTT in the driving scenario is less than 30 milliseconds.

We next show the RTT jitter in Fig. 4. We see that for stationary and driving at the speed of 100 km/h, for 90% or above, the RTT is only 7.6 milliseconds, yet for high-speed train, 56.6% of the RTT is greater than 100 milliseconds. In addition, the RTT variance is 1636 times greater than the other two scenarios. Note that an average RTT of 200 milliseconds may still be tolerable (we measure a connection between Beijing and New York can be 214.9 milliseconds), yet the large RTT jitter may seriously affect the performance.

Our conclusion is that the variance, rather than the sheer throughput or RTT is damaging for the wireless data network performance under high-speed mobility. In this paper, we are not trying to provide solutions, rather, we are looking into the root causes of such performance. We discuss wireless channel and handoff respectively.

V. ROOT CAUSE I: WIRELESS CHANNEL

We first present the wireless signal strength, which is the underlying support for wireless communication quality. There are many metrics for signal strength, such as RSSI, Ec/Io, SNR, RSRP (Reference Signal Receiving Power), and RSRQ (Reference Signal Receiving Quality). Each of these metrics is used for different scenarios, in particular, RSRP is used primarily for LTE networks [15]. In our work, we focus on LTE services; as such, we use RSRP as our major signal strength measurement metric. There are three levels in RSRP, Good (-85 dBm), Average (-95 dBm), and Weak (-105 dBm).

Fig. 5 shows the signal condition of the BTR line. Fig. 5(a) shows an experiment from 6:50 am to 7:23 am and Fig. 5(b) shows an experiment from 8:31 am to 9:04 am, both from Beijing to Tianjin. The three red lines in the figures show the Good, Average, Weak indices of RSRP.
acceptable, signal strength is dynamically changing. Since the
within the Good level. Although the general signal quality is
that data transmission is possible. And only 3 of 17 minutes are
approximately 2.15 minutes. Third, we can see that the signal
coverage holes. Most notably, we observe that there is no
effects in the other experiments as well. Second, we observe
are consistent. As a matter of fact, we have seen similar
First, we can see that the results of the two experiments
Fig. 2. TCP downlink throughput comparison with different speed.
Fig. 3. RTT comparison with different speed.
(a) RTT of HSR
(b) RTT comparison

First, we can see that the results of the two experiments
are consistent. As a matter of fact, we have seen similar
effects in the other experiments as well. Second, we observe
coverage holes. Most notably, we observe that there is no
4G LTE coverage between the 462th second and the 591th
second in the BTR line. This translates to a coverage hole of
approximately 2.15 minutes. Third, we can see that the signal
strength is primarily between Good and Average. This assures
that data transmission is possible. And only 3 of 17 minutes are
within the Good level. Although the general signal quality is
acceptable, signal strength is dynamically changing. Since the
experiments at different time are consistent, in the remaining
part of the paper, we use the data of case 1, 6:50 am - 7:23
am.

We next compare the signal strength between the high-
speed train and driving scenarios. In Fig. 6, we plot the signal
strength between the 700th second and the 1050th second on
HSR. We choose these two time-spots because the driving road
is perfectly in parallel with the high-speed rail, as shown in
Fig. 1. In addition, we can also see from Fig. 5 (a) that there
are the locations with the consistent signal.

Fig. 6(a) shows the signal strength of high-speed train and
Fig. 6(b) shows the signal strength during driving. Note that
since the speed of high-speed train is three times faster than
that of a car, the total duration of Fig. 6(a) and Fig. 6(b) is
350 seconds and 1200 seconds, respectively. We can make two
observations from Fig. 6. First, the signal strength in driving
scenarios is mostly above Good, usually between -80 dBm and
-75 dBm, which is much better than high-speed train. Second,
the variance in driving is also consistently smaller than that
of the high-speed train. These two observations show that the
high-speed mobility does negatively affect the performance on
signal strength.

After understanding the signal strength, we then investigate
the impact of signal strength on throughput under high-speed
mobility. Fig. 7 illustrates the correlation between TCP down-
link throughput and signal strength on HSR, where the three
lines represent the Good, Average, Weak signal conditions.
We find that a Good signal condition leads to good TCP
throughput and a Weak signal condition leads to much worse
TCP throughput. In addition, we observe that the throughput in
the Average and Good signal conditions are reasonably close,
yet the Weak signal condition leads to a significantly worse
performance. This shows that it is necessary to increase the
signal strength to an Average condition.

VI. ROOT CAUSE II: HANDOFF

A good wireless signal strength alone does not guarantee
good performance for the throughput of upper-layer protocols.
We next consider another key factor, handoff.

Current mobile networks are based on the cellular archi-
tecture and each base station can only cover a certain range.
occurs in every 248.2 seconds on average. In comparison, driving at the speed of 100 km/h, the handoff translates to a handoff in every 25.4 seconds on average. For 65 intra-system handoffs and 6 inter-system handoffs. This uses different radio access networks [10].

A new base station, yet for an inter-system handoff, the MD will choose intuitively, for both inter and intra handoff, the MD will choose a new base station, yet for an inter-system handoff, the MD will choose a better cell with the improvement of infrastructure, such coverage hole can be ultimately fixed. As such, we focus on intra-system handoff in the remaining part of this section.

A. Handoff Types

In our experiments, we observe multiple types of handoffs. Intuitively, there are handoffs between base stations of the same type of network, e.g., 4G to 4G, 3G to 3G, and there are handoffs of different types of network, e.g., 4G to 3G, 3G to 2G. As a matter of fact, 3G networks alone have many different subtypes, e.g., UMTS, HSPA, HSPA+. Nevertheless, these 3G networks are using the same radio access network. Therefore, in this paper, we classify two types of handoffs, inter-system handoff and intra-system handoff. More specifically, for both inter and intra handoff, the MD will choose a new base station, yet for an inter-system handoff, the MD uses different radio access networks [10].

The total number of handoffs we observe is 71, including 65 intra-system handoffs and 6 inter-system handoffs. This translates to a handoff in every 25.4 seconds on average. For comparison, driving at the speed of 100 km/h, the handoff occurs in every 248.2 seconds on average.

When a mobile device (MD) goes through the coverage and the signal strength is too weak, a handoff is triggered by the MD or the serving cell. The MD would select a better cell and switch connection from one base station to another.

With the increasing user mobility, handoff frequency will increase. Since the link is temporarily lost during handoffs, packets will be lost or delayed. If looking from the TCP point of view, either case may cause an increase in its retransmission timer and a decrease in its transmission window size, which finally leads to throughput reduction.

Fig. 11. CDF of suspension duration of throughput.

In our experiments, we observe multiple types of handoffs. Intuitively, there are handoffs between base stations of the same type of network, e.g., 4G to 4G, 3G to 3G, and there are handoffs of different types of network, e.g., 4G to 3G, 3G to 2G. As a matter of fact, 3G networks alone have many different subtypes, e.g., UMTS, HSPA, HSPA+. Nevertheless, these 3G networks are using the same radio access network. Therefore, in this paper, we classify two types of handoffs, inter-system handoff and intra-system handoff. More specifically, for both inter and intra handoff, the MD will choose a new base station, yet for an inter-system handoff, the MD uses different radio access networks [10].

The total number of handoffs we observe is 71, including 65 intra-system handoffs and 6 inter-system handoffs. This translates to a handoff in every 25.4 seconds on average. For comparison, driving at the speed of 100 km/h, the handoff occurs in every 248.2 seconds on average.

When a mobile device (MD) goes through the coverage and the signal strength is too weak, a handoff is triggered by the MD or the serving cell. The MD would select a better cell and switch connection from one base station to another.

With the increasing user mobility, handoff frequency will increase. Since the link is temporarily lost during handoffs, packets will be lost or delayed. If looking from the TCP point of view, either case may cause an increase in its retransmission timer and a decrease in its transmission window size, which finally leads to throughput reduction.

A. Handoff Types

In our experiments, we observe multiple types of handoffs. Intuitively, there are handoffs between base stations of the same type of network, e.g., 4G to 4G, 3G to 3G, and there are handoffs of different types of network, e.g., 4G to 3G, 3G to 2G. As a matter of fact, 3G networks alone have many different subtypes, e.g., UMTS, HSPA, HSPA+. Nevertheless, these 3G networks are using the same radio access network. Therefore, in this paper, we classify two types of handoffs, inter-system handoff and intra-system handoff. More specifically, for both inter and intra handoff, the MD will choose a new base station, yet for an inter-system handoff, the MD uses different radio access networks [10].

The total number of handoffs we observe is 71, including 65 intra-system handoffs and 6 inter-system handoffs. This translates to a handoff in every 25.4 seconds on average. For comparison, driving at the speed of 100 km/h, the handoff occurs in every 248.2 seconds on average.

In our experiments, we observe multiple types of handoffs. Intuitively, there are handoffs between base stations of the same type of network, e.g., 4G to 4G, 3G to 3G, and there are handoffs of different types of network, e.g., 4G to 3G, 3G to 2G. As a matter of fact, 3G networks alone have many different subtypes, e.g., UMTS, HSPA, HSPA+. Nevertheless, these 3G networks are using the same radio access network. Therefore, in this paper, we classify two types of handoffs, inter-system handoff and intra-system handoff. More specifically, for both inter and intra handoff, the MD will choose a new base station, yet for an inter-system handoff, the MD uses different radio access networks [10].

The total number of handoffs we observe is 71, including 65 intra-system handoffs and 6 inter-system handoffs. This translates to a handoff in every 25.4 seconds on average. For comparison, driving at the speed of 100 km/h, the handoff occurs in every 248.2 seconds on average.

In our experiments, we observe multiple types of handoffs. Intuitively, there are handoffs between base stations of the same type of network, e.g., 4G to 4G, 3G to 3G, and there are handoffs of different types of network, e.g., 4G to 3G, 3G to 2G. As a matter of fact, 3G networks alone have many different subtypes, e.g., UMTS, HSPA, HSPA+. Nevertheless, these 3G networks are using the same radio access network. Therefore, in this paper, we classify two types of handoffs, inter-system handoff and intra-system handoff. More specifically, for both inter and intra handoff, the MD will choose a new base station, yet for an inter-system handoff, the MD uses different radio access networks [10].

The total number of handoffs we observe is 71, including 65 intra-system handoffs and 6 inter-system handoffs. This translates to a handoff in every 25.4 seconds on average. For comparison, driving at the speed of 100 km/h, the handoff occurs in every 248.2 seconds on average.
investigate other mobile data networks, such as HSPA+ and EVDO. Besides Beijing-Tianjin Intercity Railway, we also plan to study other railway lines. We believe these understanding can help design better upper-layer protocols in the future.

ACKNOWLEDGMENT

This work has been supported in part by NSFC Project (61170292, 61161140454, 61272464), National Science and Technology Major Project (2012ZX03005001), 973 Project of China (2012CB315803), 863 Project of China (2013AA013302), EU MARIE CURIE ACTIONS EVANS (PIRSES-GA-2013-610524), RGC/GRF PolyU 5264/13E, HK PolyU G-UB72, 1-ZVC2, and 1-ZE9B.

We would like to appreciate all the anonymous reviewers and Dr. Chunyi Peng for shepherding the paper. We also thank all the volunteers who help us collect data traces on HSR.

REFERENCES