

# Do Twin Clouds Make Smoothness for Transoceanic Video Telephony?

Jian Li<sup>1,2</sup>, Zhenhua Li<sup>1\*</sup>, Yao Liu<sup>3</sup>, and Zhi-Li Zhang<sup>4</sup>

<sup>1</sup> School of Software and TNLIST, Tsinghua University

<sup>2</sup> China MinSheng Banking Corporation (CMBC)

<sup>3</sup> Department of Computer Science, Binghamton University, SUNY

<sup>4</sup> Department of Computer Science and Engineering, University of Minnesota  
{upton, lizhenghua1983}@tsinghua.edu.cn, yaoliu@cs.binghamton.edu, zh Zhang@cs.umn.edu

**Abstract**—Transoceanic video telephony (TVT) over the Internet is challenging due to 1) longer round-trip delay, 2) larger number of relay hops, and 3) higher packet loss rate. Real-world measurements of Skype, Facetime, and QQ confirm that their TVT service quality is mostly unsatisfactory.

Recently, when using WeChat to make transoceanic video calls, we are fortunate to find that it achieves stably smooth TVT. To explore how this is possible, we conduct in-depth measurements of WeChat data flow. In particular, we discover that the service provider of WeChat deploys a novel, specially designed “*twin clouds*” based architecture to deliver transoceanic (UDP) packets. Thus, data delivery between two callers is no longer *point-to-point* (used by Skype, Facetime, and QQ) over the best-effort Internet. Instead, transoceanic video packets are delivered through the *privileged backbone* formed by twin clouds, which greatly reduces the round-trip delay, number of relay hops, and packet loss rate. Besides, whenever a packet is found lost, multiple duplicate packets are instantly sent to aggressively make up for the loss.

On the other hand, we notice two-fold shortcomings of twin clouds. First, due to the sophisticated resource provisioning inside the twin clouds, the video startup time is considerably extended. Second, due to the high cost of deploying twin clouds, the capacity of the privileged backbone is limited and sometimes in shortage, and thus WeChat has to deliver data via a detour path with degraded performance. Ultimately, we believe that the twin clouds based data delivery solution will arouse a new direction of Internet video telephony research while still deserves optimization efforts.

## I. INTRODUCTION

As the world becomes more “flat”, more people leave their homelands to seek for better opportunities and/or distinct experiences. As a representative case, in recent years, China and the US (two largest economies in the world) have carried out extensive social, industrial, and educational communications and cooperations, in which transoceanic video telephony (TVT) over the Internet plays an irreplaceable role. TVT over the Internet, *e.g.*, Skype, Facetime, and Tencent QQ, is known to be not only cheap (sometimes even free) but also convenient (the two callers only need to be online at the same time).

Unfortunately, TVT over the Internet is also known to be challenging. A lot of users suffer from its unstable quality of service (QoS), in particular the unstable video telephony *smoothness*. The major reason is three folds:

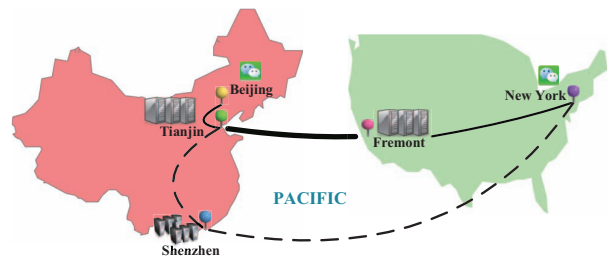


Fig. 1. WeChat’s twin clouds based infrastructure for transoceanic video telephony between Beijing and New York.

- First, the round-trip delay across the ocean is usually considered inevitably long. According to our real-world measurements of Skype, Facetime, and QQ video telephony services, their round-trip delay between Beijing and New York generally stays around 1200 msec. In comparison, their round-trip delay within China or the US is merely 252 msec.
- Second, communication packets often have to traverse a number of (IP-layer) relay hops where each hop can become a bottleneck. The typical number of relay hops between Beijing and New York is 21 for Skype, Facetime, and QQ. In comparison, their typical number of relay hops within China or the US is 12.
- Third, the transoceanic packet loss rate is higher than normal, which further degrades the service quality. The packet loss rate for Skype, Facetime, and QQ between Beijing and New York usually exceeds 10%. More importantly, only 24% of these losses are repaired. In comparison, their packet loss rate within China or the US is about 6%, and 30% of the losses are repaired.

WeChat [1] emerged as a mobile messaging service in Jan. 2011. In the past three years, it has attracted over 600 million users and its functionalities have been greatly enriched. In Jan. 2014, when we use WeChat to make transoceanic video calls, we are fortunate to find that it achieves smooth video telephony under various typical network environments. Quantitatively, considering a typical video call between users in Beijing and New York, the TVT smoothness of WeChat is stable around

\* Corresponding author.

97%, while that of Skype, Facetime, and QQ is only around 91% — note that the 6% increase in smoothness will lead to perceivable improvement in user experience.

Motivated by the better TVT smoothness of WeChat, we conduct in-depth measurements and analysis of WeChat data flow, in order to uncover what essentially distinguishes WeChat from Skype, Facetime, and QQ in terms of TVT smoothness. In particular, we discover that the service provider of WeChat deploys a novel, specially designed “twin clouds” based infrastructure to deliver the transoceanic UDP packets<sup>1</sup>. As depicted in Fig. 1, one cloud is located in Tianjin (China) and spans the two major ISPs in China (*i.e.*, Unicom and Telecom), and the other is located in Fremont (CA, US), hosted by Hurricane Electric ISP. Specifically, the major technical advantages of WeChat are summarized as follows:

- 1) With twin clouds, data delivery between two parties of the call is no longer point-to-point over the best-effort Internet. Instead, the transoceanic UDP packets are delivered through the *privileged backbone* formed by the twin clouds. Thus, both the round-trip delay and the number of relay hops are greatly reduced. Compared with Skype, Facetime, and QQ, the round-trip delay is reduced from 1200 msec to 360 msec, and the number of hops is reduced from 21 to 11.
- 2) In order to mitigate the negative effect of the “ISP barrier” (especially between Unicom and Telecom) in China, WeChat utilizes an ISP-aware cloud server selection mechanism for the Tianjin cloud. In most cases (more than 98%), a Unicom user connects to a Unicom server and a Telecom user connects to a Telecom server.
- 3) WeChat adopts an aggressive packet retransmission strategy to recover from packet loss. Whenever a UDP packet is found lost, multiple (1 to 4) duplicate UDP packets are instantly sent out to make up for the loss, bringing about 8% retransmission traffic increase. As a result, the *eventual* TVT packet loss rate of WeChat is merely 3%.

In addition, the twin clouds based data delivery solution is also used by WeChat when we make TVT calls between Beijing and Pennsylvania (in the US), Milan (in Italy), Dubai (in the United Arab Emirates), and so forth. Of course, as for Milan and Dubai the other cloud does not lie in Fremont, US, but the TVT performance is similar to that between Beijing and New York. In a nutshell, our measurement results shed lights on how TVT service providers can wisely invest their resources to improve the TVT service quality, thus further increasing the user retention rate.

On the other hand, we notice that the twin clouds based TVT service has its shortcoming in two folds, and is therefore far from perfect at the moment:

- 1) Due to the sophisticated resource provisioning mechanism in the twin clouds, the TVT video startup time of WeChat is as long as 10 sec. In comparison, the

<sup>1</sup>Skype, Facetime, and QQ also use the connectionless UDP protocol to transport TVT data.

TVT video startup time of Skype, Facetime, and QQ is generally less than 5 sec.

- 2) Due to the high cost of deploying twin clouds, the capacity of the privileged backbone is limited and sometimes (~5%) in shortage. In case that the capacity of the privileged backbone is temporarily exhausted, WeChat has to deliver data via a detour path (plotted as the dotted line in Fig. 1) with performance degradation (*e.g.*, the TVT smoothness is decreased from 97% to around 94%).

Ultimately, we believe that the twin clouds based data delivery solution will arouse a new direction of Internet video telephony research while still deserves further optimization efforts.

**Roadmap.** The rest of this paper is organized as follows. We first review the related work in § II, and then describe our research methodology in § III. The detailed research results are presented in § IV and § V. Finally, we conclude the paper with some possible future work in § VI.

## II. RELATED WORK

A number of researches have been conducted on video telephony service over the Internet. Most previous works focus on the QoS [2], [3], [4], [5], [6], [7], [8], [9], traffic analysis [10], [11], [12], [13], [14], [15], [16], [17], system architecture [18], [19], [20], [21], [22], [23], [24], [25], and enabling techniques [26], [27], [28], [29], [30], [31] of Internet video telephony within an area or a nation. However, to the best of our knowledge, little is known about whether the mainstream Internet video telephony services are applicable to the extremely challenging transoceanic video telephony scenario. Below we review three recent works that are most related to our research.

Xu *et al.*'s seminal work [21] (IMC'12 best paper) identifies the key differences in design choices between Skype, iChat, and Google+. Their research results reveal that different system architectures, video generation/adaption mechanisms, and packet loss recovery strategies actually lead to distinct qualities of video telephony. We follow their methodology to study transoceanic video telephony. Thereby, our work is complementary to this work, and it identifies several unexpected issues that can hardly be detected under common network environments. In this paper, we investigate QQ rather than Google+ because 1) the Google+ service is currently not available in China, and 2) QQ is the most popular video telephony service in China, and has many users in the US as well. Moreover, we study Facetime rather than iChat though they are both provided by Apple Inc., because Facetime is more often used for transoceanic video telephony.

As a sequel to the above work, Yu *et al.* scrutinize the user-perceived performance of video telephony under various wireless environments [29]. They observe that the conservative video rate selection and packet loss recovery strategies of Skype, Facetime, and Google+ generally achieve smooth video telephony. However, our measurements indicate that Internet video telephony still bears performance bottlenecks under

transoceanic scenarios. Consequently, this topic is worthy of further research efforts.

Prior to this paper, we have studied the QoS of transoceanic video *messaging* service provided by WhatsApp and WeChat [32]. We find that when delivering video messages, WeChat’s small TCP sliding window size yields unexpectedly long delay. It is also observed that cloud-assisted video message delivery is a promising way to increase transoceanic throughput. This paper further studies the transoceanic video telephony of WeChat, and unravels its amazing smoothness and novel architecture. It is reported that WhatsApp will launch its video telephony service this year [33]. Given that WhatsApp also has a considerable part of overseas users [34], its developers may benefit from our research.

### III. METHODOLOGY

This section describes our research methodology of designing proper benchmark experiments and defining relevant metrics, in order to make a comprehensive and comparative study of Skype, Facetime, QQ, and WeChat, in terms of their domestic video telephony (DVT) performance and transoceanic video telephony (TVT) performance. To accommodate the dynamics of the Internet and the four concerned services, the experiments are performed across a range of locations, devices, and time periods.

**Client locations.** First, we make domestic video calls between Beijing and Shanghai (both in China) with 10 Mbps of access bandwidth. In this case, the Beijing and Shanghai clients alternatively use Internet accesses provided by the two major ISPs in China (*i.e.*, Unicom and Telecom).

Next, we make transoceanic video calls between two distant locations: Beijing (in China) and New York (in the US), also with 10 Mbps of access bandwidth. The Beijing client still alternatively uses Internet accesses provided by Unicom and Telecom; the New York client uses the Internet access provided by the TWCI (Time Warner Cable Internet) ISP.

**Hardware and software.** Two iPhone 5S smartphones (running iOS 7.0) are used in our experiments. Other types of client hardware (*e.g.*, Android smartphones, laptops, and desktops) and operating systems (*e.g.*, Android 4.2 and Windows 7) are also tried, and we find that they generally do not affect the video telephony performances. As a result, we would focus on the case of two iPhone 5S smartphones in the rest of this paper. The client software installed on our experiment smartphones include Skype (v4.17.135), Facetime (v7.0), QQ (v4.6.2), and WeChat (v5.2.1).

**Data collection.** Each smartphone is connected to the Internet via a dedicated WiFi AP (access point, or says home router). With such a network setting, we are able to sniff all the communication packets involved in the video telephony process. Specifically, the video telephony packets are captured by a laptop that runs the Wireshark [35] network protocol analyzer in its (in fact the NIC’s) promiscuous mode. We also write a script to extract a series of detailed fields of each packet from the collected packet traces, including the arrival

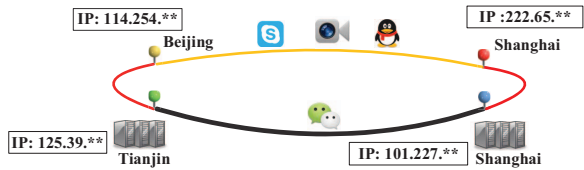


Fig. 2. DVT data transmission paths and the corresponding IP addresses.

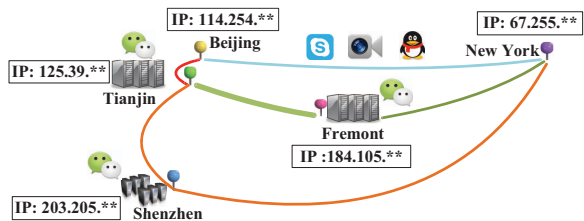


Fig. 3. TVT data transmission paths and the corresponding IP addresses.

time, source and destination IP addresses, retransmission flag, time to live (TTL), and so forth.

**Data transmission paths.** In the DVT experiments, there are two different data transmission paths, as illustrated in Fig. 2. The corresponding IP addresses<sup>2</sup> used in the experiments are labelled along each path:

- 1) The upper, yellow path between Beijing and Shanghai is used by Skype, Facetime, and QQ to directly transmit UDP packets over the best-effort Internet.
- 2) The lower, circuitous path is used by WeChat. It is composed of three sub-paths: Beijing ↔ Tianjin, Tianjin ↔ Shanghai, and Shanghai ↔ Shanghai. The second sub-path (*i.e.*, Tianjin ↔ Shanghai) is the private transmission path (or says the privileged backbone) of WeChat.

In the TVT experiments, there are three different data transmission paths, as plotted in Fig. 3. We also label the corresponding IP addresses along each path:

- 1) The uppermost path is used by Skype, Facetime, and QQ. With this path, every UDP packet is delivered in a point-to-point manner over the best-effort Internet.
- 2) The middle path is composed of three sub-paths: Beijing ↔ Tianjin, Tianjin ↔ Fremont, and Fremont ↔ New York. The second sub-path (*i.e.*, Tianjin ↔ Fremont) is the private transmission path (also the privileged backbone) of WeChat.
- 3) The lowermost path is also composed of three sub-paths: Beijing ↔ Tianjin, Tianjin ↔ Shenzhen, and Shenzhen ↔ New York. This is also a private transmission path of WeChat (for detour use when the middle path is congested).

**Video telephony smoothness.** Following the conventional understanding, we use the *eventual* UDP packet receiving rate

<sup>2</sup>We anonymize the last two segments of each IP address for privacy reason.

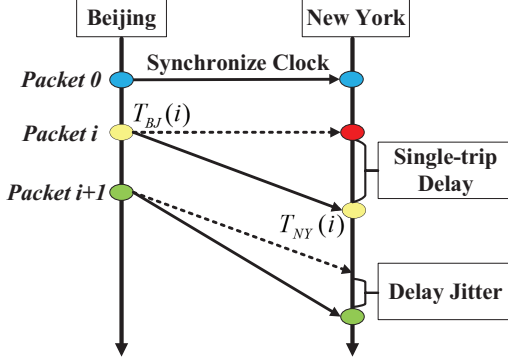


Fig. 4. Delay and jitter.

to measure the video telephony smoothness in this work [21], [29], [36]. That is, we consider both the packets that are received upon first transmission and the packets that are received after retransmission upon packet losses. In other words, a lower packet loss rate or a higher packet recovery rate may result in higher video telephony smoothness. Moreover, for real-time video telephony service, recovering lost packets after a certain period of time (typically,  $T = 2$  sec) generally makes no sense [21]. Therefore, a packet is considered *in-time* if it is received *within two seconds after its first transmission* in our research. In summary, we compute video telephone smoothness as the ratio of the number of packets received *in-time* (at the receiver side) over the total number of packets sent (from the sender side).

**Network metrics.** In order to provide an in-depth explanation of the video telephony smoothness, we also have to understand several other network metrics:

- *Number of relay hops* denotes how many (IP-layer) relay hops a video telephony UDP packet has to traverse to reach its destination over the Internet. In all our benchmark experiments, the UDP packet size is always found smaller than 1400 bytes, and thus every UDP packet is shipped by a single IP packet.
- *Round-trip delay* is the sum of two single-trip UDP packet delays, *e.g.*,  $\text{delay}_{BJ \rightarrow NY} + \text{delay}_{NY \rightarrow BJ}$  (BJ denotes Beijing and NY denotes New York). As illustrated in Fig. 4, to accurately measure the UDP packet delay, we first synchronize the clocks in Beijing and New York by running the `ntpdate` tool. Let  $T_{BJ}(i)$  denote the departure time of packet  $i$  at Beijing, and  $T_{NY}(i)$  denote the arrival time of packet  $i$  at New York. Then, the single-trip delay is calculated as:

$$\text{delay}_{BJ \rightarrow NY} = T_{NY}(i) - T_{BJ}(i). \quad (1)$$

As a result, the round-trip delay between Beijing and New York is:

$$r - \text{delay} = \text{delay}_{BJ \rightarrow NY} + \text{delay}_{NY \rightarrow BJ}. \quad (2)$$

- *Delay jitter* describes the stability (in terms of deviation) of single-trip (or round-trip) delay. Low delay jitter is

desired for achieving high-quality video telephony. As depicted in Fig. 4, the delay jitter is calculated as:

$$\begin{aligned} \text{jitter}_{BJ \rightarrow NY} &= (T_{NY}(i+1) - T_{NY}(i)) \\ &\quad - (T_{BJ}(i+1) - T_{BJ}(i)). \end{aligned} \quad (3)$$

- *ISP awareness* is crucial to China's Internet which suffers greatly from the serious "ISP barrier" [37], [38]. Video telephony service will benefit from ISP-aware data delivery (especially in China) if there is any.
- *Video startup time* denotes how long it takes to set up a relatively stable connection between the two video telephony callers.

#### IV. EVALUATION OF DOMESTIC VIDEO TELEPHONY

In this section, we evaluate the quantitative benefits of the twin clouds based data delivery solution under the scenario of domestic video telephony (DVT). Through packet trace analysis, we quantify several metrics that largely affect video telephony smoothness, such as the number of relay hops, round-trip delay, packet loss rate, and recovery rate.

##### A. Video Telephony Smoothness

To sustain continuous domestic video calls, service providers must maintain stably high video telephony smoothness. We compute video telephony smoothness by matching packet traces collected at the sender side and the receiver side. For Skype, Facetime, and QQ, packets can be matched by their data field even if they pass through a cloud. But for WeChat, packets' data field will be changed after they are delivered by a cloud server. Using the same RTP header analysis technique introduced in [29], we match packets according to their RTP headers.

According to the analysis in [21] and our definition of smoothness, users are satisfied when the video telephony smoothness approaches 98%. Fig. 5(a) and 5(b) show the typical DVT smoothness of the four video telephony services on both sides, *i.e.*, the Beijing user and the Shanghai user. We use each point on a curve to denote the average value of DVT smoothness in one minute. The combined result of Fig. 5(a) and 5(b) is shown in Fig. 5(c), from which we find that the DVT smoothness of WeChat is excellent: 98.4% on average, ranging between 97% and 98.7%. Moreover, the average DVT smoothness of Skype, Facetime, and QQ is also high: 94.3%, 95.8%, and 95.2%, respectively.

Meanwhile, the stability (in terms of variance) of DVT smoothness of the four video telephony services is presented in Fig. 6. We see that WeChat outperforms Skype, Facetime, and QQ in both the absolute value and the stability of DVT smoothness. Although the DVT smoothness for the four video telephony services is already high, we expect their performances to be similarly fine under more challenging scenarios, *e.g.*, transoceanic video telephony.

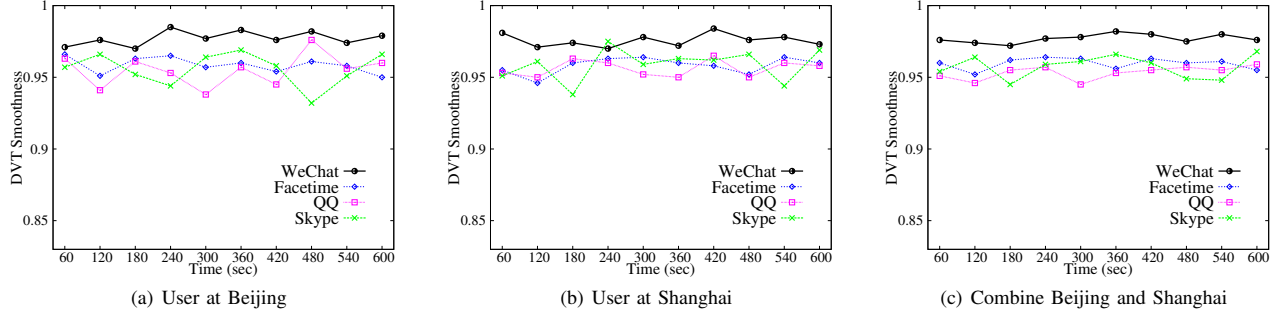


Fig. 5. Domestic video telephony (DVT) smoothness.

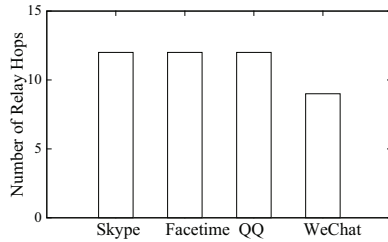


Fig. 7. Number of relay hops.

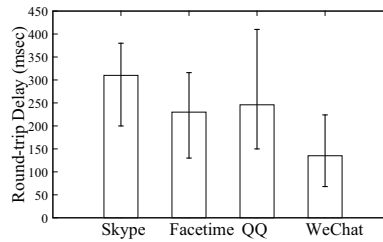


Fig. 8. Round-trip delay.

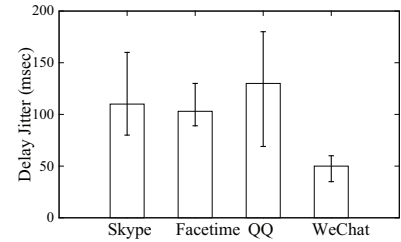


Fig. 9. Delay jitter.

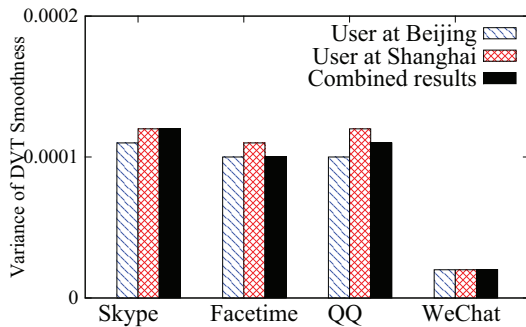


Fig. 6. Variance of DVT smoothness.

### B. Relay Hops and Round-trip Delay

Video telephony smoothness heavily relies on two fundamental network metrics: 1) the number of relay hops and 2) the round-trip delay over the Internet. Of course, the video telephony service providers always want the two metrics to be as small/short as possible. The major reason is two folds:

- 1) A smaller number of relay hops may decrease packet loss rate and round-trip delay. It also improves system reliability since each hop can become a bottleneck.
- 2) Shorter round-trip delay can facilitate real-time interactions between users. It also contributes to the recovery from data losses within a short period of time (typically in 2 seconds) through packet retransmission.

Fig. 7 shows the number of relay hops of the four video telephony services (with confidence interval). The result shows that WeChat decreases the number of relay hops from 12 to

9, compared with Skype, Facetime, and QQ. The reason is that WeChat delivers its UDP packets through its special twin clouds based infrastructure while the others deliver packets through the best effort Internet.

We use the formal model (refer to Equation (2) in § III) to evaluate the round-trip delay. Fig. 8 shows the average DVT round-trip delay of Skype, Facetime, QQ, and WeChat, respectively. It is clear that the DVT round-trip delay of WeChat is shorter than that of Skype, Facetime, and QQ — decreasing from 252 msec to 135 msec.

Further, we follow the Equation (3) in § III to analyze the delay jitter of the four video telephony services. As presented in Fig. 9, we see that the average delay jitter of WeChat is 50 msec. On the other hand, the average jitter of Skype, Facetime and QQ is as large as 110 msec, 103 msec, and 128 msec, respectively. It is worth noting that there is a minor portion of out-of-order UDP packets according to our measurements. To make Fig. 9 tidy, we ignore these out-of-order UDP packets. The results show that WeChat reduces the round-trip delay slightly but improves the delay jitter.

### C. ISP Awareness

Internet in China is quite different from Internet in the US. In particular, China has a very simple AS topology: it has two major ISPs (*i.e.*, Unicom and Telecom), each has a giant AS that is built on top of a nationwide backbone network [39]. In § III, we have mentioned the “ISP barrier” between Unicom and Telecom. And the previous work has confirmed that cross-ISP data delivery usually bears poor performance [37]. This observation highlights the need for service providers to optimize data delivery across different ISPs.

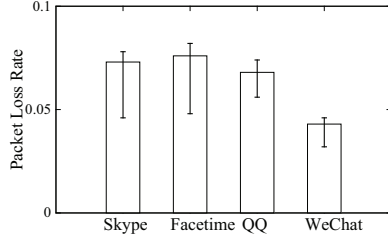


Fig. 10. Packet loss rate.

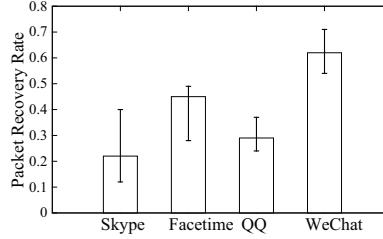


Fig. 11. Packet recovery rate in time.

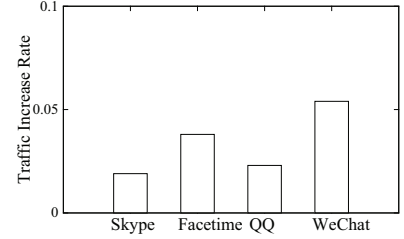


Fig. 12. Retransmission traffic increase rate.

The most effective method to address the “ISP barrier” problem is to restrict the communication data flow within a single ISP. To examine whether this method is applied in DVT services, in the benchmark experiments we alternatively use Internet (ADSL) accesses provided by Unicom and Telecom. We observe that WeChat utilizes an ISP-aware cloud server selection mechanism for both Beijing and Shanghai users. In most cases (more than 98%), a Unicom user *automatically* connects to a Unicom cloud server, and a Telecom user automatically connects to a Telecom cloud server. Here we say “in most cases” because we do find in some cases, a Telecom user connects to a Unicom cloud server, but the chances are very small — this phenomenon can be reasonably attributed to the temporary overload of Telecom cloud servers in Tianjin. Apparently, such ISP awareness is enabled by the WeChat system in itself. On the contrary, we did not observe obvious ISP awareness in Skype, Facetime, or QQ systems.

#### D. Packet Loss and Recovery

To achieve high video telephony smoothness, fast data delivery (with fewer hops and shorter delay) is one side, while fewer delivery failures (with lower packet loss rate) and efficient recovery (with higher packet recovery rate) is the other side. Fig. 10 shows the average packet loss rate of Skype, Facetime, QQ, and WeChat. We discover that the packet loss rate of WeChat is 4.3%, while that value of Skype, Facetime, and QQ is 7.3%, 7.6%, and 6.8%, respectively. The reason is related to the WeChat’s twin clouds based data delivery infrastructure — the reduction in the number of relay hops (refer to Fig. 7) in the transmission path can decrease the packet loss rate.

The UDP protocol (used by almost all the mainstream Internet video telephony services) does not have a self-checking mechanism (like the TCP protocol) to ensure the eventually successful arrival of each data packet. Thus, its upper-layer applications must implement their own packet checking and recovering mechanisms. Through the benchmark experiments, we find that all the four video telephony services employ packet retransmission to recover from packet loss. In particular, we find that WeChat adopts an aggressive retransmission strategy to recover from packet loss when we make a domestic video call between Beijing and Shanghai. Whenever a UDP packet is found lost, multiple (1 to 4) duplicate UDP packets are instantly transmitted to make up for the loss.

The results for packet recovery rate are depicted in Fig. 11, which indicates that the average packet recovery rate of WeChat is 62%. On the other hand, the average packet recovery rate of Skype, Facetime, and QQ is 22%, 45%, and 29%, respectively. Facetime has better packet recovery rate, which can be attributed to the fact that Facetime also retransmits some lost packets. The eventual DVT packet loss rates of the four services are calculated as follows:

- Skype:  $7.3\% \times (1 - 22\%) = 5.7\%$ .
- Facetime:  $7.6\% \times (1 - 45\%) = 4.2\%$ .
- QQ:  $6.8\% \times (1 - 29\%) = 4.8\%$ .
- WeChat:  $4.3\% \times (1 - 62\%) = 1.6\%$ .

Both the aggressive packet retransmission strategy and the conservative strategy (adopted by Skype, Facetime, and QQ) have their specific pros and cons. The aggressive retransmission strategy has the appeal of improving video telephony smoothness. However, this approach results in high network overhead. To quantify the impact of packet retransmission, we measure the retransmission traffic increase rate (*i.e.*, the ratio of the number of redundant retransmission packets over the total number of retransmission packets). Fig. 12 compares the average retransmission traffic increase rate of WeChat with those of the other three video telephony services. The retransmission traffic increase rate of WeChat is about 7.4% while that value of Skype, Facetime, and QQ is 1.9%, 3.8%, and 2.3%, respectively.

#### E. Video Startup Time

We investigate DVT startup time under a variety of network environments with high dynamics, and find that WeChat generally achieves similar DVT startup time to those of the other three video telephony services. Usually, the DVT startup delay of Skype, Facetime, QQ, and WeChat is all around 3 seconds.

### V. EVALUATION OF TRANSOCEANIC VIDEO TELEPHONY

The previous section confirms that video telephony smoothness is sensitive to several key metrics, such as the round-trip delay, the number of relay hops, the packet loss rate, the packet recovery rate, and the ISP awareness. In this section, we further analyze how these metrics influence the video

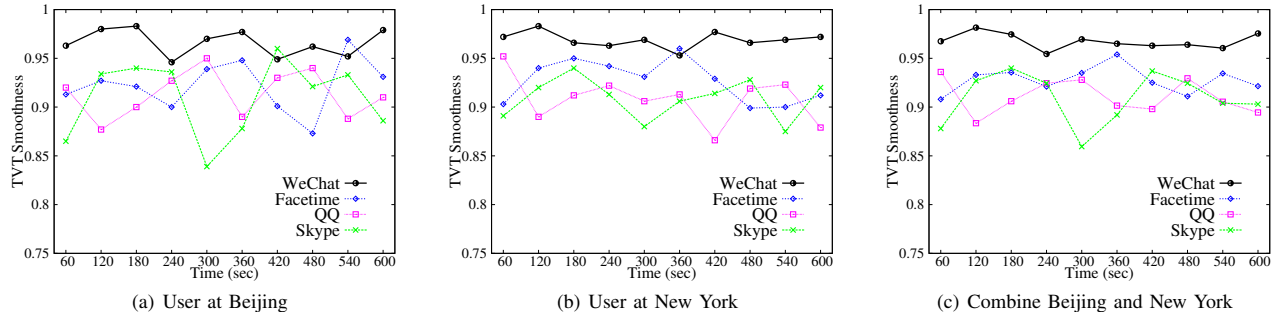


Fig. 13. Transoceanic video telephony (TVT) smoothness.

telephony smoothness under the scenario of transoceanic video telephony (TVT).

### A. Video Telephony Smoothness

In Fig. 13(a) and 13(b), we depict the typical TVT smoothness of the four video telephony services on both sides, *i.e.*, the Beijing user and the New York user. Also, each point on a curve represents the average value of TVT smoothness in one minute. Clearly, WeChat generally achieves higher TVT smoothness than Skype, Facetime, and QQ. Combining the two sub-figures together (see Fig. 13(c)), we observe that the TVT smoothness of WeChat is 97% on average (very close to 98%, the satisfactory value for video telephony users), ranging between 94.9% and 98.6%. Although the TVT smoothness of WeChat is already quite high, it is slightly lower than the DVT smoothness (98.4% on average, refer to Figure 5(c) in § IV), which is within our expectation.

On the other hand, the average TVT smoothness of Skype, Facetime, and QQ is merely 90.8%, 92.6%, and 91.2%, respectively. As mentioned in § I, the 6% increase (90.8%, 92.6%, 91.2% → 97%) in TVT smoothness will lead to obvious improvement in user-perceived quality of service. These results reveal that WeChat achieves remarkably better video telephony smoothness under more challenging (*i.e.*, transoceanic) scenarios.

Fig. 20 compares the stability (in terms of variance) of TVT smoothness of the four video telephony services. We find that WeChat also achieves higher stability (smaller variance) than Skype, Facetime, and QQ. Therefore, we conclude that WeChat outperforms the other services in terms of both absolute value and stability of TVT smoothness.

### B. Relay Hops and Round-trip Delay

To understand how the data delivery solution impacts the TVT smoothness, we measure the number of relay hops and the round-trip delay under TVT scenarios. As shown in Fig. 14, our experiment results indicate that WeChat greatly decreases the number of relay hops from 21 to 11, compared with Skype, Facetime, and QQ. Compared with the DVT scenario, WeChat only increases the number of relay hops by 2: from 9 to 11 (refer to Figure 7 in § IV). This is attributable

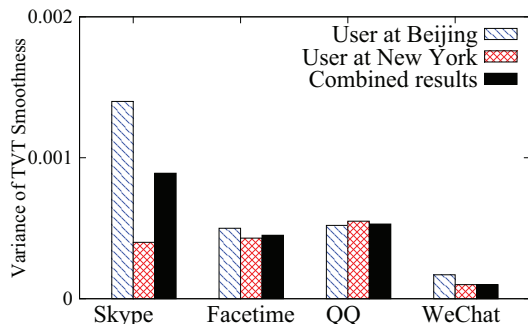


Fig. 20. Variance of TVT smoothness.

to its special twin clouds based infrastructure which constructs a “highway” across the ocean.

On the other hand, for all the other three services, the number of relay hops is greatly increased by 9: from 12 to 21 (also refer to Figure 7). This is an intuitive phenomenon, because they all completely rely on the best-effort Internet as its UDP packet delivery infrastructure.

Ideally, the reduction in the number of relay hops leads to the reduction the round-trip delay. Fig. 15 presents the average TVT round-trip delay of Skype, Facetime, QQ, and WeChat. Compared with the other services, WeChat significantly decreases the TVT round-trip delay from 1200 msec to 360 msec. Furthermore, Fig. 16 reflects that the average delay jitter of WeChat is merely 90 msec, but the average delay jitter of Skype, Facetime, and QQ is 240, 156, and 281 msec, respectively. In conclusion, the TVT round-trip delay of WeChat is not only much shorter but also more stable than Skype, Facetime, and QQ.

### C. ISP Awareness

As mentioned in § IV-C, cross-ISP data delivery in China usually suffers from serious performance degradation, and thus WeChat develops the ISP-aware cloud server selection mechanism. In this section, we explore whether this mechanism is also applied in TVT scenarios. Similarly, we alternatively use Internet (ADSL) accesses provided by Unicom and Telecom. Again, we find that a Unicom user automatically connects

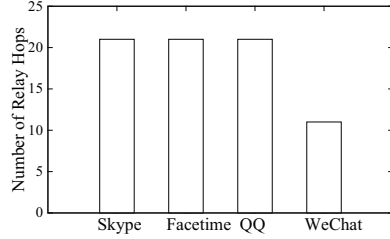


Fig. 14. Number of relay hops.

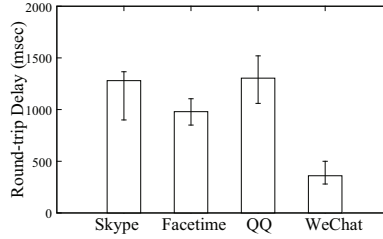


Fig. 15. Round-trip delay.

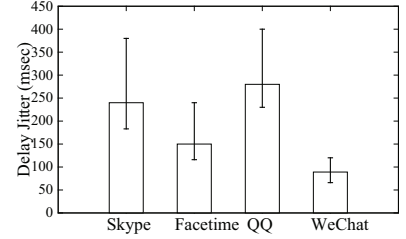


Fig. 16. Delay jitter.

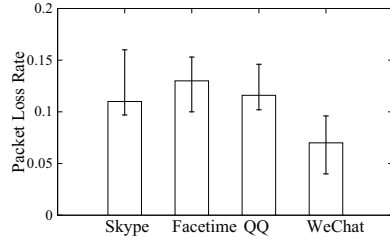


Fig. 17. Packet loss rate.

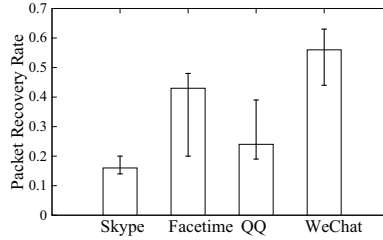


Fig. 18. Packet recovery rate in time.

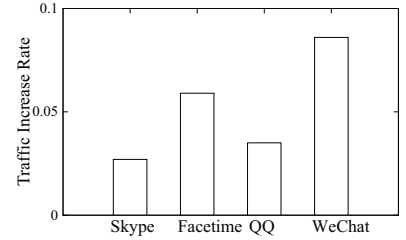
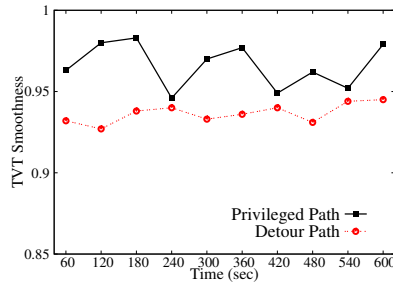
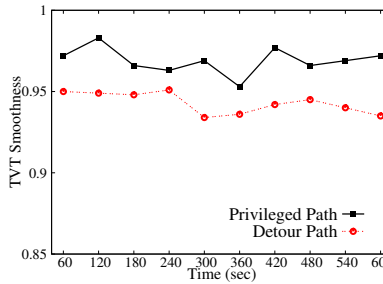


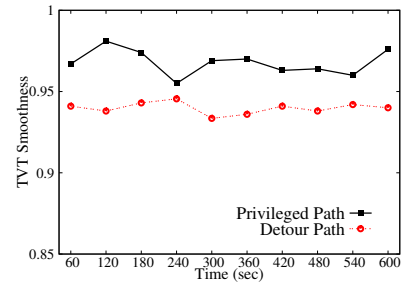
Fig. 19. Retransmission traffic increase rate.



(a) User at Beijing



(b) User at New York



(c) Combine Beijing and New York

Fig. 21. Transoceanic video telephony (TVT) smoothness.

TABLE I  
COMPARISON OF WECHAT PERFORMANCES USING THE MAJOR AND DETOUR TRANSMISSION PATHS.

Data transmission path	TVT smoothness	Number of relay hops	Round-trip latency	Packet loss rate	Packet recovery rate
BJ ↔ Tianjin ↔ <b>Fremont</b> ↔ NY	97%	11	360 msec	7%	56%
BJ ↔ Tianjin ↔ <b>Shenzhen</b> ↔ NY	94%	15	516 msec	8.2%	27%

to a Unicom cloud server, and a Telecom user connects to a Telecom cloud server. Apparently, such ISP awareness is enabled by the WeChat system in itself. On the contrary, we did not observe obvious ISP awareness in Skype, Facetime, or QQ systems.

#### D. Packet Loss and Recovery

As we have discussed in § IV-D, decreasing the packet loss rate together with the (*in time*) packet recovery rate can yield considerable improvement on video telephony smoothness. Fig. 17 shows the average packet loss rate of Skype, Facetime, QQ, and WeChat. As expected, with the help of its special and stable twin clouds based infrastructure, WeChat effectively

reduces the packet loss rate to 7%, while that value of Skype, Facetime, and QQ is 11%, 13%, and 11.6%, respectively. This result indicates that the twin clouds based data delivery solution possesses better robustness under various network environments.

Although all the four video telephony services retransmit packets to recover from packet loss, the aggressive retransmission strategy of WeChat provides more benefits, of course, at the cost of much more network overhead. The specific results are listed in Fig. 18, indicating that the average packet recovery rate of WeChat is as high as 56%, while the average packet recover rate of Skype, Facetime, and QQ is 16%, 43%, and 24%, respectively. More importantly, the *eventual* TVT packet



loss rates of the four services are calculated as follows:

- Skype:  $11\% \times (1 - 16\%) = 9.2\%$ .
- Facetime:  $13\% \times (1 - 43\%) = 7.4\%$ .
- QQ:  $11.6\% \times (1 - 24\%) = 8.8\%$ .
- WeChat:  $7\% \times (1 - 56\%) = 3\%$  — the lowest.

In addition, although the aggressive retransmission strategy improves video telephony smoothness, it inevitably increases the network load. Specifically, as shown in Fig. 19, the retransmission traffic increase rate (refer to § IV-D) of Skype, Facetime, QQ, and WeChat is 2.7%, 5.9%, 3.5%, and 8.5%, respectively. Obviously, the packet retransmission overhead brought by WeChat is much more than that of the other services.

#### E. Shortage of Twin Clouds' Capacity

Our comprehensive measurement results have revealed multi-fold advantages of the twin clouds based data delivery solution of WeChat. Nevertheless, all these advantages depend on a single performance bottleneck — the twin clouds based infrastructure must have sufficient bandwidth capacity to accommodate all the TVT users' bandwidth requirements.

As the transoceanic bandwidth is extremely expensive, deploying and maintaining twin clouds incurs continuously high cost. Even for Tencent (the back-end supporter of the WeChat service), a major Internet service provider in China, the capacity of the privileged backbone is limited. In our experiments, we actually observe the cases when the capacity of the twin clouds is in shortage. When the capacity of the privileged backbone is temporarily exhausted, we discover that WeChat uses an alternative cloud located at Shenzhen (China) to take over the data delivery. In this detour path (see Fig. 1), the Shenzhen cloud takes the place of the Fremont cloud.

In Fig. 21(a) and 21(b), we depict the TVT smoothness of the two delivery paths on both sides, *i.e.*, the Beijing user and the New York user. We use privileged path to denote the path that starts in Beijing and passes through Tianjin, Fremont, and New York. In addition, the detour path denotes the path that starts in Beijing and passes through Tianjin, Shenzhen, and New York. The combined result of Fig. 21(a) and 21(b) is shown in Fig. 21(c). It is clear that the detour path leads to performance degradation compared with the twin clouds based path. The detailed comparisons between the privileged path and the detour path are summarized in Table I. Fortunately, even in this case, the TVT smoothness of WeChat (94%) is still better than that of Skype, Facetime, and QQ.

#### F. Video Startup Time

Another shortcoming of WeChat we have observed in our study is its longer video startup time. The TVT video startup delay of Skype, Facetime, and QQ is usually less than 5 sec, but for WeChat it is around 10 sec. On the other hand, our analysis in § IV-E shows that the DVT startup delay of all the

four services is about 3 sec. The major reason for the longer video startup time can be reasonably traced to the sophisticated resource provisioning mechanism (*e.g.*, cloud server selecting, bandwidth scheduling, and load balancing) utilized in the twin clouds. This is not a trivial problem, as for some impatient users 10-second video startup time may well let them down.

## VI. CONCLUSION

As the Internet becomes more powerful and pervasive, user expectations of high-quality Internet video telephony services (in particular smoothness) are continuously rising, even in the context of transoceanic video telephony (TVT). It is therefore critical for services providers to understand how to wisely invest their resources to maintain sound QoS of TVT.

In this paper, we conduct the first comprehensive study of real-world TVT performance, compared with the performance of domestic video telephony (DVT). Our key takeaway lies in that the WeChat's twin clouds based data delivery infrastructure significantly optimizes several important performance metrics (*e.g.*, number of relay hops, round-trip delay, and packet loss rate) and thus effectively improves the TVT smoothness. On the other hand, we find that the twin clouds based infrastructure has its accompanied shortcomings, *e.g.*, longer video startup time and shortage of privileged bandwidth.

Our study provides useful implications from both commercial and technical perspectives. From a commercial perspective, our research results can help video telephony service providers better invest their resources. From a technical perspective, our study not only discusses the technical design of TVT systems but also motivates the needs of better solutions (*e.g.*, better pro-active routing to reduce startup time and adaptive resource provisioning technique).

Several interesting future research directions remain. For example, we wonder how video codec and rate adaption techniques affect the TVT performance. Besides, this paper concentrates on two-party TVT service, and we are interested in whether the "twin clouds" are also applicable to multi-party transoceanic video conferencing. Finally, although this paper has explored the key differences of WeChat from three other video telephony services (*i.e.*, Skype, Facetime, and QQ), still an important "proxy server"-based video telephony service (*i.e.*, Google+) is missing in our research. The reason is unfortunately political — Google+ is generally not available in Mainland China. As a result, next step we plan to investigate the similarities and differences between WeChat and Google+ by making TVT calls between other Asia Pacific areas (*e.g.*, Hong Kong and Singapore) and US.

## VII. ACKNOWLEDGEMENTS

This work is supported by the High-Tech Research and Development Program of China ("863 China Cloud" Major Program) under grant 2015AA01A201, China NSF under grant 61471217, and China Postdoctoral Science Fund under grant 2014M550735. Zhi-Li Zhang is supported in part by US NSF

under grants CNS-1117536 and CNS-1411636, and US DoD ARO MURI Award W911NF-12-1-0385.

## REFERENCES

- [1] "WeChat," <http://www.wechat.com>.
- [2] Z. Li, J. Cao, and G. Chen, "ContinuStreaming: Achieving High Playback Continuity of Gossip-based Peer-to-Peer Streaming," in *Proceedings of International Parallel and Distributed Processing Symposium (IPDPS)*. IEEE, 2008, pp. 1–12.
- [3] Z. Li, J. Cao, G. Chen, and Y. Liu, "On the Source Switching Problem of Peer-to-Peer Streaming," *Journal of Parallel and Distributed Computing (JPDC)*, vol. 70, no. 5, pp. 537–546, 2010.
- [4] J. Xu and B. W. Wah, "Exploiting Just-Noticeable Difference of Delays for Improving Quality of Experience in Video Conferencing," in *Proceedings of Multimedia Systems*. ACM, 2013, pp. 238–248.
- [5] X. Zhang, Y. Xu, H. Hu, Y. Liu, Z. Guo, and Y. Wang, "Profiling Skype Video Calls: Rate Control and Video Quality," in *Proceedings of INFOCOM*. IEEE, 2012, pp. 621–629.
- [6] T.-Y. Huang, K.-T. Chen, and P. Huang, "Tuning Skype's Redundancy Control Algorithm for User Satisfaction," in *Proceedings of INFOCOM*. IEEE, 2009, pp. 1179–1187.
- [7] Y. Lu, Y. Zhao, F. Kuipers, and P. Van Mieghem, "Measurement Study of Multi-party Video Conferencing," in *NETWORKING*. IEEE, 2010, pp. 96–108.
- [8] B. Grozev, L. Marinov, V. Singh, and E. Ivov, "Last N: Relevance-based Selectivity for Forwarding Video in Multimedia Conferences," in *Proceedings of the 25th ACM Workshop on Network and Operating Systems Support for Digital Audio and Video*. ACM, 2015, pp. 19–24.
- [9] D. Niu and B. Li, "Congestion-aware Internet Pricing for Media Streaming," in *Proceedings of the 3rd INFOCOM Workshop on Smart Data Pricing*. IEEE, 2014, pp. 571–576.
- [10] D. Bonfiglio, M. Mellia, M. Meo, D. Rossi, and P. Tofanelli, "Revealing Skype Traffic: When Randomness Plays with You," *SIGCOMM Computer Communication Review*, vol. 37, no. 4, pp. 37–48, 2007.
- [11] D. Bonfiglio, M. Mellia, M. Meo, and D. Rossi, "Detailed Analysis of Skype Traffic," *IEEE Transactions on Multimedia*, vol. 11, no. 1, pp. 117–127, 2009.
- [12] N. M. Markovich and Krieger, "Statistical Analysis and Modeling of Skype VoIP Flows," *Computer Communications*, vol. 33, pp. 11–21, 2010.
- [13] D. Rossi, S. Valenti, P. Veglia, D. Bonfiglio, M. Mellia, and M. Meo, "Pictures from the Skype," *SIGMETRICS Performance Evaluation Review*, vol. 36, no. 2, pp. 83–86, 2008.
- [14] H. Sengar, Z. Ren, H. Wang, D. Wijesekera, and S. Jajodia, "Tracking Skype VoIP Calls over the Internet," in *Proceedings of INFOCOM*. IEEE, 2010, pp. 1–5.
- [15] K. Suh, D. R. Figueiredo, J. Kurose, and D. Towsley, "Characterizing and Detecting Relayed Traffic: A Case Study using Skype," in *Proceedings of INFOCOM*. IEEE, 2006, pp. 6–11.
- [16] D. Bonfiglio, M. Mellia, M. Meo, N. Ritacca, and D. Rossi, "Tracking down Skype Traffic," in *Proceedings of INFOCOM*. IEEE, 2008, pp. 261–265.
- [17] L. De Cicco, S. Mascolo, and V. Palmisano, "Skype Video Responsiveness to Bandwidth Variations," in *Proceedings of Network and Operating System Support for Digital Audio and Video*. ACM, 2008, pp. 81–86.
- [18] N. Hermanns, L. Hamm, and Z. Sarker, "A Framework and Evaluation of Rate Adaptive Video Telephony in 4G LTE," in *Proceedings of World Telecommunications Congress*. VDE, 2014, pp. 1–6.
- [19] Y. Feng, B. Li, and B. Li, "Airlift: Video Conferencing as a Cloud Service using Inter-datacenter Networks," in *International Conference on Network Protocols*. IEEE, 2012, pp. 1–11.
- [20] X. Chen, M. Chen, B. Li, Y. Zhao, Y. Wu, and J. Li, "Celerity: A Low-delay Multi-party Conferencing Solution," in *Proceedings of the International Conference on Multimedia*. ACM, 2011, pp. 493–502.
- [21] Y. Xu, C. Yu, J. Li, and Y. Liu, "Video Telephony for End-consumers: Measurement Study of Google+, iChat, and Skype," in *Proceedings of Internet Measurement Conference*. ACM, 2012, pp. 371–384.
- [22] Y. Zhao, Y. Liu, C. Chen, and J. Zhang, "Enabling P2P One-view Multi-party Video Conferencing," *IEEE Transactions on Parallel and Distributed Systems*, vol. 25, no. 1, pp. 73–82, 2014.
- [23] S. Zhang, D. Niu, Y. Hu, and F. Liu, "Server Selection and Topology Control for Multi-party Video Conferences," in *Proceedings of Network and Operating System Support for Digital Audio and Video*. ACM, 2014, pp. 43–48.
- [24] S. A. Baset and H. Schulzrinne, "An Analysis of the Skype Peer-to-Peer Internet Telephony Protocol," in *Proceedings of INFOCOM*. IEEE, 2006, pp. 1–11.
- [25] Y. Chu, S. Rao, S. Seshan, and H. Zhang, "Enabling Conferencing Applications on the Internet Using an Overlay Multicast Architecture," *ACM SIGCOMM Computer Communication Review*, vol. 31, no. 4, pp. 55–67, 2001.
- [26] F. Dabek, R. Cox, F. Kaashoek, and R. Morris, "Vivaldi: A Decentralized Network Coordinate System," vol. 34, no. 4, pp. 15–26, 2004.
- [27] T.-Y. Huang, P. Huang, K.-T. Chen, and P.-J. Wang, "Could Skype be More Satisfying? A QoE-Centric Study of the FEC Mechanism in an Internet-scale VoIP System," *IEEE Network*, vol. 24, no. 2, pp. 42–48, 2010.
- [28] C. Liang, M. Zhao, and Y. Liu, "Optimal Bandwidth Sharing in Multiswarm Multiparty P2P Video-conferencing Systems," *IEEE/ACM Transactions on Networking*, vol. 19, no. 6, pp. 1704–1716, 2011.
- [29] C. Yu, Y. Xu, B. Liu, and Y. Liu, "Can You See Me Now?: Rate Control and A Study of Mobile Video Calls," in *Proceedings of INFOCOM*. IEEE, 2014.
- [30] W. Kho, S. A. Baset, and H. Schulzrinne, "Skype Relay Calls: Measurements and Experiments," in *Proceedings of INFOCOM*. IEEE, 2008, pp. 1–6.
- [31] J. Jansen, P. Cesar, D. C. Bulterman, T. Stevens, I. Kegel, and J. Issing, "Enabling Composition-based Video-Conferencing for the Home," *IEEE Transactions on Multimedia*, vol. 13, no. 5, pp. 869–881, 2011.
- [32] Y. Liu and L. Guo, "An Empirical Study of Video Messaging Services on Smartphones," in *Proceedings of Network and Operating System Support for Digital Audio and Video*. ACM, 2014, pp. 79–84.
- [33] "WhatsApp Aims for Video Calls, 1B Users," <http://www.veooz.com/news/GuX96Z.html>.
- [34] "You May Not Use WhatsApp, But the Rest of the World Sure Does," <http://www.wired.com/2014/02/whatsapp-rules-rest-world/>.
- [35] "Wireshark," <http://www.wireshark.org>.
- [36] M. Claypool, D. Finkel, A. Grant, and M. Solano, "Thin to Win? Network Performance Analysis of the OnLive Thin Client Game System," in *Proceedings of Network and Systems Support for Games*. IEEE, 2012, pp. 1–6.
- [37] Y. Huang, Z. Li, G. Liu, and Y. Dai, "Cloud Download: Using Cloud Utilities to Achieve High-quality Content Distribution for Unpopular Videos," in *Proceedings of ACM Multimedia*. ACM, 2011, pp. 213–222.
- [38] Z. Li, Y. Huang, G. Liu, F. Wang, Z.-L. Zhang, and Y. Dai, "Cloud Transcoder: Bridging the Format and Resolution Gap between Internet Videos and Mobile Devices," in *Proceedings of ACM NOSSDAV*. ACM, 2012, pp. 33–38.
- [39] Y. Tian, R. Dey, Y. Liu, and K. Ross, "Topology Mapping and Geolocating for China's Internet," *IEEE Transactions on Parallel and Distributed Systems*, vol. 24, no. 9, pp. 1908–1917, 2013.