Dispersing Instant Social Video Service Across Multiple Clouds

Zhi Wang, *Member, IEEE,* Baochun Li, *Fellow, IEEE,* Lifeng Sun, *Member, IEEE,*
Wenwu Zhu, *Fellow, IEEE,* and Shiqiang Yang, *Senior Member, IEEE*

Abstract—Instant social video sharing which combines the online social network and user-generated short video streaming services, has become popular in today's Internet. Cloud-based hosting of such instant social video contents has become a norm to serve the increasing users with user-generated contents. A fundamental problem of cloud-based social video sharing service is that users are located globally, who cannot be served with good service quality with a single cloud provider. In this paper, we investigate the feasibility of dispersing instant social video contents to multiple cloud providers. The challenge is that inter-cloud social *propagation* is indispensable with such multi-cloud social video hosting, yet such inter-cloud traffic incurs substantial operational cost. We analyze and formulate the multi-cloud hosting of an instant social video system as an optimization problem. We conduct large-scale measurement studies to show the characteristics of instant social video deployment, and demonstrate the trade-off between satisfying users with their ideal cloud providers, and reducing the inter-cloud data propagation. Our measurement insights of the social propagation allow us to propose a heuristic algorithm with acceptable complexity to solve the optimization problem, by partitioning a propagation-weighted social graph in two phases: a preference-aware initial cloud provider selection and a propagation-aware re-hosting. Our simulation experiments driven by real-world social network traces show the superiority of our design.

1 INTRODUCTION

Instant social video sharing based on the combination of online social networks and user-generated video streaming, has rapidly emerged as one of the most important social media services for users to access contents online [1]. A fundamental reason for the popularity of social video sharing is that it satisfies the users' inherent interests in sharing video contents which are generated and uploaded by users themselves [2], with their friends [3]. When viewing such user-generated videos, other users need to download the media files from servers over the Internet.

As a result, the placement of instant social video content and the network performance between the servers

Z. Wang is with the Graduate School at Shenzhen, Tsinghua University, Shenzhen, China.
B. Li is with the Department of Electrical and Computer Engineering, University of Toronto.
L. Sun (✉: sunlf@tsinghua.edu.cn), W. Zhu, and S. Yang are with the Tsinghua National Laboratory for Information Science and Technology, the Department of Computer Science and Technology, Tsinghua University, Beijing, China.
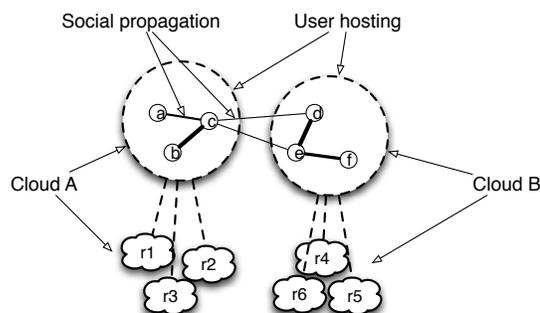

Fig. 1. Multi-cloud hosting based on preference- and propagation-aware social graph partition.

and the users can significantly affect the service quality of instant social video sharing systems. For this reason, many of them try to use the cloud-based services to deploy their systems, and take full advantage of the elastic and geo-distributed server resource availability in the cloud [4], [5], [6], [7]. Since online social network services are generally targeted at a large scale of users distributed at different geographic locations, to satisfy the needs of users, possibly with different network conditions, we may need to allocate servers across many different geographic regions and ISPs, for the sake of achieving better network performance by allocating servers in the proximity of users [8].

Today, a number of online multimedia services can be deployed over the geo-distributed cloud and network infrastructure [6]. Intuitively, multi-cloud hosting provides better geographical diversity for servers, since no single cloud provider is able to cover all the regions/ISPs across the Internet [9], to serve users with their ideal servers. The growing trend of social application and the existing geo-distributed deployment for online multimedia applications lead to the idea of *multi-cloud instant social video hosting*, or *multi-cloud hosting* in short, in that the instant video contents are dispersed to multiple cloud service providers, rather than a single cloud provider. Fig. 1 gives an example of the multi-cloud hosting (details are to be presented in Sec. 3).

A fundamental difference between an instant social video sharing system and a traditional content distri-

bution system is the presence of content propagation in the social network, in which social activities such as *sharing* are demanded by users [3]. In the context of multi-cloud hosting, besides storage and network cost, social propagation can lead to a large volume of content exchanges between different cloud providers, incurring a high cost of inter-cloud content replication.

The reason is that cloud providers tend to block content replication between the cloud providers with custom tailored pricing schemes: (1) A cloud provider typically encourages a social video sharing system to host user-generated contents, *e.g.*, the *incoming* traffic in Amazon EC2 (Elastic Cloud Computing) is not charged at all; and (2) a cloud provider charges much more for the *outgoing* traffic to a different cloud provider than inside the same cloud, *e.g.*, for the first 10 TB data transferred from an Virginia EC2 instance outside, the price is 0.12 USD/GB if the traffic goes to a server hosted by a different cloud provider, while it is only 0.02 USD/GB if the outgoing traffic remains in Amazon EC2, even though both traffics transfer between the same pair of locations [10].

Such a pricing scheme penalizes outbound transfers, and establishes a roadblock that limits replication across the boundary between different cloud providers, even though such replication is indispensable in the context of social video sharing, since users frequently share and *reshare* content from one another [11]. Taking inter-cloud propagation into account, we seek to study the design space of a multi-cloud hosting strategy that can achieve the following objectives: (1) Satisfying the *cloud-provider preference* of users, so that users are hosted with their ideal cloud providers — they can share contents to friends and view contents generated by their friends fast [12]; and (2) Reducing the *cost of inter-cloud traffic* caused by social propagation between users hosted with different cloud providers.

In this paper, we study how to efficiently host an instant social video system with multiple cloud providers based on partition of a propagation-weighted social graph. First, we conduct large-scale measurements to study the benefit of hosting social video contents with multiple cloud providers, the challenges with such multi-cloud hosting, and design guidelines from social propagation characteristics. Second, we formulate the multi-cloud hosting as an optimization problem, which is proven to be NP-hard. *Third*, based on our measurement insights, we propose to solve the problem heuristically by dividing the partition into two phases: an initial preference-aware cloud selection (so that users can upload/download the instant videos to/from their ideal servers), and a *propagation-aware* re-hosting (so as to reduce the cost of replicating the content across the boundary between multiple cloud providers caused by the social propagation). Since only a small set of social connections incurring a large amount of replication cost are re-hosted in our design, the algorithm can efficiently partition large-scale social graphs.

The remainder of this paper is organized as follows. In Sec. 2, we motivate our design by measurement studies of real-world social video sharing and cloud systems. In Sec. 3, we formulate the problem and present our multi-cloud hosting design based on the preference- and propagation-aware social graph partition. In Sec. 4, we evaluate the performance of our design using trace-driven simulations. In Sec. 5, we discuss related work. Finally, we conclude the paper in Sec. 6.

2 MOTIVATION AND DESIGN PRINCIPLES

In this section, we first present our motivation based on measurement results of an instant social video sharing system, then we present our measurement insights for the multi-cloud hosting design for instant social video contents.

2.1 Assumption: Hosting Users Instead of Contents

Before presenting the measurement results, we give the assumption made in the multi-cloud hosting.

Social connections determine how contents propagate between users in the online social network [1]. Content propagation over these social connections turns an online social network to a "user-subscribing" network, *i.e.*, each user acts as a source which generates the contents to be subscribed by others [2]. For this reason, in our study of the multi-cloud hosting of a social video sharing system, we focus on handling the hosting of users, *i.e.*, contents generated or shared by a user will be hosted by the same cloud provider assigned to host the user, and different cloud providers are assigned to host different users. Note that content instead of users is physically stored and served by the cloud servers. These contents can be either static (e.g., photos uploaded by users), or dynamical (e.g., pages generated according to different contexts). The users are actually "logical" instances, which generate and propagate contents in the online social networks. Since users instead of content are the key to social propagation, using users as logical instances to determine which content should be hosted by which servers is thus a simple design for propagation-aware multi-cloud hosting.

The benefits of hosting contents of the same user in the same cloud provider are as follows: (1) it avoids individually handling the user-generated contents, which have an extremely large number [13]; and (2) developers for the instant video sharing system can access a user's own contents *locally*, when they are hosted within the same cloud [14].

Today, user-level redirection is feasible for several content providers. For example, two different users will use different URLs (associated with their IDs) to download the same content from different servers. As social networks are popularly used by users, such user-aware redirection will be more practical in the future.

The problem is then to determine which cloud provider is assigned to host which user.

2.2 Measurement Setup

To motivate our design, we present the measurement on users' cloud-provider preference, the replication road-block across the boundary between different cloud providers, and the propagation characteristics in instant social video sharing systems, respectively. We use active and trace-based measurements as follows.

2.2.1 Instant Video Sharing and Social Propagation

We have obtained content upload and request traces from Weishi from the technical team of Tencent, an instant social video sharing system based in China. In Weishi, short videos (in 10 seconds) are generated by individuals and shared with their "followers". Each video in Weishi is transcoded into the following versions: a) 480x480, 2000 Kbps; b) 480x480, 1050 Kbps; c) 480x480, 500 Kbps; d) 480x480, 300 Kbps. The Weishi traces record two types of user activities in April 2014: (1) Video upload: each record records when a video is generated and uploaded by a particular user; (2) Video download: each item records when a video with a particular version is download by which user, and from which server.

To further study social propagation between users, we have also obtained Weibo traces, containing valuable runtime data of the system in 6 months (June 2011 — November 2011). We have collected two types of traces as follows: (1) the social relationship database, which records how users are socially connected to each other at different time points; (2) the microblogs, which are messages posted by the users — each entry includes the ID, name, IP address of the publisher, time stamp when the microblog is posted, IDs of the parent and root microbloggers if it is a re-post [15].

Weibo and Weishi are different social networks, and different types of online social network services may have different social propagation patterns and characteristics. To jointly use them for studying the service deployment for instant social video service, we carry out the following preprocessing: (1) We choose the traces of Weibo and Weishi, instead of Twitter and Vine, because Weishi and Weibo share a significant fraction of users, as Weishi was developed based on the social graph of Weibo; (2) In Weibo, we only use the propagation traces of videos, and remove other types of multimedia contents, to study the propagation patterns of videos.

2.2.2 PlanetLab-based Active Measurement

To practically study the user preference with servers located at different geographic regions over the world, we use PlanetLab-based experiments. We simulate user activities on the PlanetLab nodes, and let them download from and upload to cloud servers allocated from Amazon EC2 (Elastic Cloud Computing) [16], to study the user preference of different cloud regions.
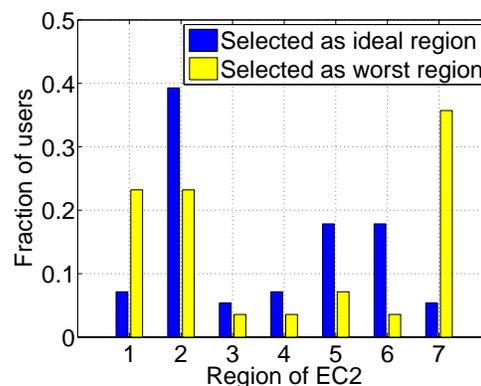

Fig. 2. User preference of regions in the Amazon cloud service.

2.3 Benefits from Multi-Cloud Hosting

2.3.1 Diverse Regions/ISPs Improve Service Quality

To show that diverse server deployment improves the service quality in instance social video sharing, we study the performance of users downloading contents from servers at different geographic locations. In particular, we measure the time users (PlanetLab nodes) spend on downloading contents from the servers allocated at different locations (7 Amazon regions are selected). The content size is 1 MB, and users download the content over HTTP, from the same type of web server. We repeated these download experiments in one week, and calculated the average download speeds of the users, to infer their preference of servers deployed at different regions.

Fig. 2 shows the preference of 55 PlanetLab nodes randomly distributed in different locations over the world. These nodes download the same content from servers deployed at 7 different regions which are randomly selected from Amazon regions. A pair of bars represent the fraction of each region being selected as the "ideal" region (i.e., a PlanetLab node downloads the contents from the region at the highest speed), against the "worst" region (i.e., a PlanetLab node downloads the contents from the region at the slowest speed), respectively. We observe that all the regions have an opportunity to be selected as the ideal region by users, indicating that different users have different region preference.

2.3.2 Multiple Cloud Providers Cover More Service Regions

Today's cloud providers are scaling their services globally, by building datacenters at different regions and with different ISPs around the world. However, it is difficult for a single cloud provider to cover all the possible regions/ISPs that an instant social video sharing system requires, to serve the users with servers deployed at their ideal regions [9].

For example, the Amazon EC2 has deployed servers at 9 regions, including 1. Virginia, 2. Oregon, 3. California,

4. Ireland, 5. Frankfurt, 6. Singapore, 7. Tokyo, 8. Sydney, and 9. Sao Paulo [16], but these regions fail to locally serve users at some locations, *e.g.*, users in China. On the other hand, some Chinese cloud providers including Tencent Cloud [17] can provide servers in a variety of these regions in China.

It is promising for a social video sharing system to allocate servers from a larger range of regions and ISPs, by utilizing more cloud providers.

Measurement insight. We observe that (1) different cloud providers have deployed servers at different regions/ISPs, and (2) user preference of different regions/ISPs is very different. As a result, multi-cloud hosting of an instant social video sharing system is appealing, since multiple cloud providers allow the system to host contents at a large range of regions/ISPs, so as to improve the possibility for users to download from and upload to their ideal servers.

2.4 Challenges of Multi-cloud Hosting for Instant Social Video Service

Next, we study the statistics of content uploads and requests in an instant video sharing service, the replication limitation caused by the pricing schemes of the cloud providers, and the dynamics of social propagation in an instant social video sharing system.

2.4.1 Instant Video Uploads and Requests

Based on the Weishi traces, we measure the content uploads and requests in an instant social video sharing service. We first study the statistics of the number of video uploads and requests in one day, as illustrated in Fig. 3. The two curves in this figure represent the number of instant videos uploaded/requested by users in each time slot (1 hour) over time. We observe that both the upload and request curves demonstrate daily patterns, with the peak hours at 8pm and 10pm, respectively. We also observe that the average number of requests is around 100x larger than that of uploads, indicating that it is likely for the popular videos to be requested by many users, located at different regions. It is thus necessary to deploy these contents into multiple clouds.

We next study the elapse between the upload of a video and the requests. We plot the CDF of the elapse between the upload time of a video and the time the first request for the video was issued in Fig. 4, over $300,000$ videos. We observe that more than 40% (resp. 55% and 85%) of the videos were requested 1 hour (resp. 8 hours and 24 hours) after they were uploaded. These observations show that it is necessary for instant videos to be deployed into multiple clouds timely.

2.4.2 Inter-Cloud Replication Cost

Another challenge is the inter-cloud *replication cost*, which is the cost from replicating contents between cloud providers due to social propagation. In this paper,
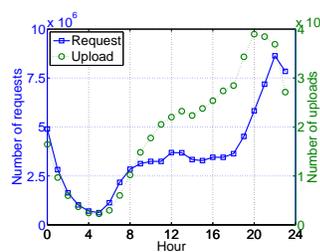

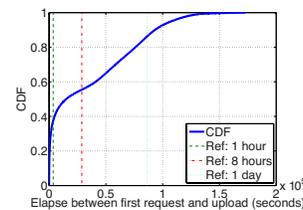Fig. 3. Number of uploads and requests of instant videos over time.

Fig. 4. CDF of elapse between upload and the first request.

propagation cost and replication cost are interchangeable. Taking the bandwidth pricing schemes used by Amazon EC2 [10] as example, we observe two important pricing schemes used in today's cloud providers as follows. (1) *Incoming content encouragement*. It is observed that cloud providers do not charge the incoming traffic from the Internet, *i.e.*, in the cloud-based social video sharing system, the contents generated by users can be uploaded to servers of any cloud providers for free. (2) *Inter-cloud replication roadblock*. However, it is observed that the outgoing traffic is generally charged. Specifically, the cloud providers charge a regular price of the outgoing traffic when contents are transferred from inside one cloud to another cloud provider; while they charge much less when the outgoing traffic is inside the same cloud. For example, the price scheme for the first 10 TB outgoing traffic in Amazon EC2 is illustrated in Table 1. When data is transferred outside an EC2 server at Virginia, the price is 0.12 USD/GB on average if the traffic goes to a server hosted by a different cloud provider; while it is only 0.02 USD/GB if the outgoing traffic remains in Amazon EC2.

This pricing scheme restricts the social video sharing systems from freely extending their service to multiple cloud providers, given that contents are indispensably replicated between servers in different cloud providers because of the social propagation between users hosted with different cloud providers. Note that the pricing in our study is an input: Our design tries to reduce the replication cost caused by the data transfer price between different cloud providers.

2.4.3 Dynamics of Social Propagation

Another challenge is related to the dynamics of social topology.

Creation and removal of social connections. One year since it was online, social connections within Tencent Weibo were still changing dramatically. Fig. 5 illustrates the creation and removal ratios of social connections related to a sample of 1 million users over time (*i.e.*, any social connection that connects at least one user in the 1 million users is included).

In Fig. 5(a), we have a baseline of the number of social connections in June 2011, and each sample in the "social connection created" curve represents the creation

TABLE 1
Data transfer price of Amazon EC2 for outgoing traffic
(USD/GB for first 1TB, November 2014) [10].

Region	To another Amazon region	To another cloud provider
Virginia	0.02	0.12
Oregon	0.02	0.12
California	0.02	0.12
Ireland	0.02	0.12
Frankfurt	0.02	0.12
Singapore	0.09	0.19
Tokyo	0.09	0.201
Sydney	0.14	0.19
Sao Paulo	0.16	0.25
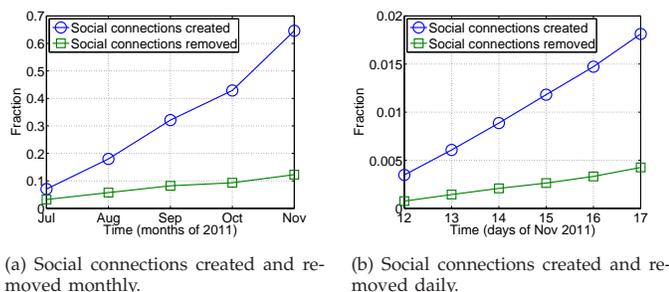
(a) Social connections created and re-moved monthly.

(b) Social connections created and re-moved daily.

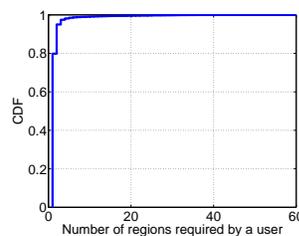Fig. 5. Dynamics of social connections in Tencent Weibo after it was online.

Fig. 6. CDF of the number of regions demanded by a user to serve all her friends with their ideal servers.

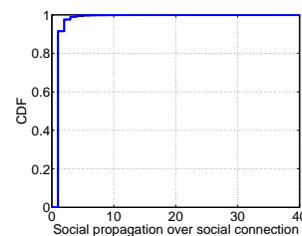Fig. 7. CDF of the propagation weights of social connections.

ratio of social connections since then, while each sample in the "social connection removed" curve represents the removal ratio of social connections since then. The creation and removal of social connections among users are relatively dynamical — in 5 months, over 65% new social connections are created. Besides, much more social connections are created (friending or following others) than removed (un-friending or un-following others). Similar results are also observed in a period of one week in November 2011 in Fig. 5(b).

For a newly deployed social video sharing system, the social connections can change dramatically over time for a long period.

Measurement insight. Multi-cloud hosting of a social video sharing system is challenging, because (1) the cloud providers are using pricing schemes that block the replication of contents between cloud providers, and (2) the social topology is changing dynamically due to frequent creation and removal of the social connections after the social video sharing system is deployed.

2.5 Principles Learnt from Measurement Studies

We study the characteristics of the social propagation in the online social network, which can guide the multi-cloud hosting design.

2.5.1 A Few Server Regions Are Enough for Most of the Users

After a content is generated or shared by a user in the online social network, her friends are the ones who are to view the content. As mentioned above, these friends have different preference of regions to download the contents from. We study how many server regions are required to host a user so that every friend of her can download the content from their ideal regions. Based on the Weibo traces, we retrieve users' geographic locations and estimate their ideal server regions based on the geographic distance, *i.e.*, an ideal server region is one closest to the user.

Fig. 6 illustrates the CDF of the number of regions demanded by a user so that all her friends can download the content from their ideal regions. We observe that this number follows a heavy-tailed distribution, *i.e.*, while some of the users need to be hosted with many regions to serve their friends with their ideal regions, most of the users only need a small number of regions, which are highly possible to be covered by a single cloud provider (which is different for different users). Since the first type of users tend to be hosted by almost all the cloud providers available to the social video sharing system, in our design, we are focused on the second type of users, to determine which cloud provider is assigned to host which user.

2.5.2 A Few Social Connections Incur a Large Amount of Propagation

In an online social network, users can reach the content generated by others through the social connections between them. We observe that the number of contents propagating over different social connections can be quite different. In Fig. 7, we plot the CDF of the *propagation weight* (*i.e.*, the number of reshares via a particular social connection in 1 day) of $10,000$ social connections, randomly chosen from all the social connections among all the 1 million users.

We observe that the distribution of the propagation weight over different social connections is also heavy-tailed. To reduce the cost of inter-cloud traffic, we need to take social connections with the dominating propagation weight into account when applying the multi-cloud hosting.

Measurement insight. We observe that, (1) for most of the users, each individual of them only needs a few number of regions to serve the contents for her

followers, which can be provided by a single cloud provider, though multiple cloud providers are needed to cover regions for all of the users as different users need different sets of cloud providers; (2) only a few number of social connections incur a large amount of social propagation, which is the cause of the dominate inter-cloud data transfer cost.

Based on the measurement insights, we will present our design of the multi-cloud hosting of an instant video sharing media system in Sec. 3.

3 MULTI-CLOUD HOSTING: DETAILED DESIGN

Fig. 8 illustrates the framework of our instant social video multi-cloud hosting proposal. We design the instant video content hosting strategies following a data-driven approach. We collect the following information: (1) the social propagation information, including the social graph between users, how they generate videos and how these contents propagate via social connections; and (2) the cloud information, including locations and ISPs of cloud servers, their upload/download speeds to users, and the resource price of these cloud servers (*e.g.*, storage, data). Based on these information, we carry out the multi-cloud hosting strategies, to partition users to different cloud providers, such that users can receive videos with good streaming quality, as well as that the overall content replication cost is minimized.

We need to strategically determine which users should be hosted by which cloud provider, so as to not only satisfy users' cloud-provider preference but also reduce the cost of inter-cloud content replication. In this section, we present our detailed design for instant social video multi-cloud hosting.

Fig. 1 illustrates the idea of hosting an instant social video sharing system with multiple cloud providers based on the preference- and propagation-aware social graph partition. In this figure, $a, b, \ldots, f$ represent users in the social network, and $r1, r2, \ldots, r6$ represent regions with servers deployed by two cloud providers A and B, where $\{r1, r2, r3\} \in A$ and $\{r4, r5, r6\} \in B$. Each user can generate and share a number of contents, which will be downloaded by her friends. The segments between users represent the social connections, which can be retrieved from the online social network. The thickness of a segment represents the propagation level between two users, *i.e.*, a thicker segment indicates that more contents propagate between two users. Recall that such propagation will cause the content replication between the servers where the two users are hosted, as presented in Sec. 2.

We assume that users a, b, and c prefer cloud A, while users d, e, and f prefer cloud B, *i.e.*, better download and upload performance can be achieved if they are hosted with their ideal cloud provider. In this example, we observe that the partition (indicated by the two large dashed circles) of the users can satisfy

the preference of all users, as well as minimize the inter-cloud propagation, since the propagation weights of social connections (c, d) and (c, e) are much smaller than that of other social connections. However, in most of the cases, satisfying users' cloud-provider preference and minimizing the inter-cloud propagation will conflict (*e.g.*, when the propagation weight between c and d is very large), and we need to strategically achieve the two objectives jointly.

Next, we present the formulation of the multi-cloud hosting problem, and our solution based on the measurement insights.

3.1 Problem Formulation

In this subsection, we will formulate the multi-cloud hosting of an instant social video sharing service into an optimization problem. In particular, the objectives we seek to achieve are as follows: (1) we need to satisfy the cloud-provider preference of users who are influenced to download the contents shared by their friends; (2) we need to satisfy the cloud-provider preference of users who generate and upload the contents; and (3) we need to reduce the inter-cloud traffic caused by social propagation between users that are hosted with different cloud providers.

Before we present the details of our design, we summarize some important notations in Table 2.

3.1.1 Social Graph and Multiple Cloud Providers

Let $\mathcal{G} = \{\mathbf{V}, \mathbf{E}\}$ denote the social graph with each node $u \in \mathbf{V}$ representing a user in the online social network, and each edge $e_{uv} \in \mathbf{E}$ denoting the propagation weight from user u to user v. Let $\mathbf{C}$ denote the set of cloud providers that the instant social video sharing system can be hosted with. Each cloud provider $c \in \mathbf{C}$ has a set of regions $\mathbf{R}_c$ where users can be hosted.

3.1.2 Cloud Preference of a User

The cloud-provider preference of a user includes the following two perspectives: (1) improving the download performance by hosting the user with a cloud so that the user's friends can download from regions that are close to them; and (2) improving the upload performance by hosting the user with a cloud with regions that are close to the user himself.

Local download index. Let $Q(u, c)$ denote the local download index of hosting user u with cloud c, to satisfy her friends to download from local regions. $Q(u, c)$ is calculated as follows:

$$Q(u, c) = \sum_{v \in \mathbf{F}_u} \max_{s \in \mathbf{R}_c} \chi(v, s),$$

where $\mathbf{F}_u$ is the set of user u's friends in the online social network, r_v is the region where user v is located, and $\chi(v, s)$ denotes the preference level for user v to download contents that are hosted at region s. $\chi(v, s)$ depends on the network condition between user v and
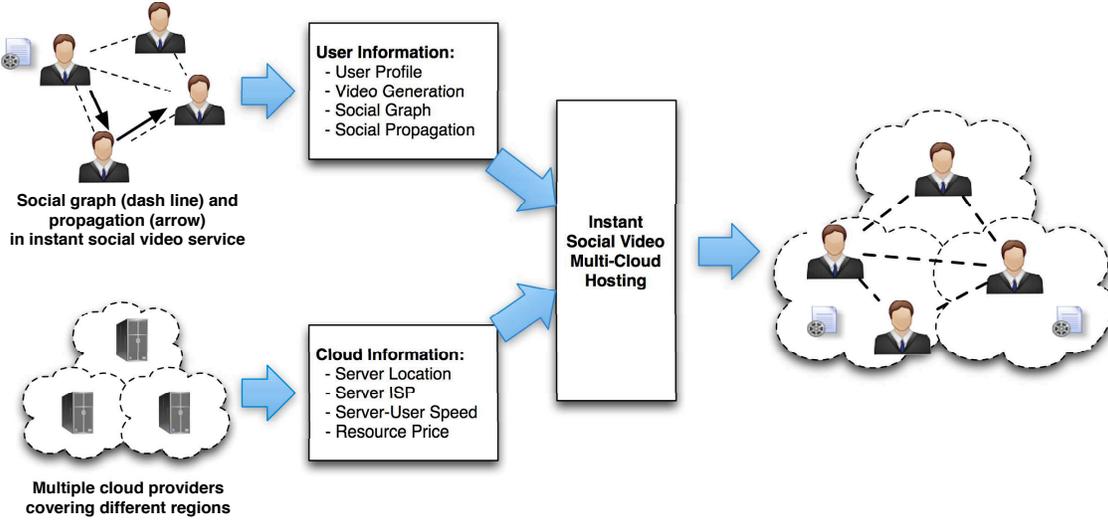
Fig. 8. Framework of instant social video multi-cloud hosting.

TABLE 2
Important notations

Symbol	Definition
u, v, w	Indices for users in the online social network
c, d, f	Indices for cloud providers
$\mathcal{G} = \{\mathbf{V}, \mathbf{E}\}$	The social graph with users in set $\mathbf{V}$ and social connections in set $\mathbf{E}$
$\mathbf{C}$	The set of cloud providers
e_{uv}	The propagation weight of social connection $u \rightarrow v$
$\mathbf{R}_c$	The set of regions in cloud provider c
$\mathbf{F}_u$	The set of friends of user u
$\chi(v, s)$	The preference level for user v to download contents from servers at region s
$\chi'(v, s)$	The preference level for user v to upload contents to servers at region s
$Q(u, c)$	The local download index of user u hosted with cloud provider c
$W(u, c)$	The local upload index of user u hosted with cloud provider c
$\psi(u, c)$	The preference of user u to the cloud provider c
$Y(u, c)$	The cost of inter-cloud replication between user u and her friends if u is hosted with cloud provider c
p_{cd}	The data-transfer price of inter-cloud traffic from cloud c to cloud d
α	The parameter used to balance satisfying users' cloud-provider preference and reducing inter-cloud propagation
γ_X	The gain of re-hosting users using the strategy X
$\phi(\cdot)$	The cost caused by inter-cloud propagation in the re-hosting
η	A threshold used to determine which social connections are considered in the re-hosting.

region s, and large $\chi(v, s)$ indicates that better network performance can be achieved in the download. The rationale of $Q(u, c)$ is that hosting a user u with a cloud c with large $Q(u, c)$ can benefit her friends, who can download the contents generated or shared by user u from their ideal cloud regions.

Local upload index. We also seek to find an ideal server for the user himself to upload the generated contents to. Let $W(u, c)$ denote the local upload index, which represents the upload performance achieved at user u when u is hosted with cloud c. $W(u, c)$ is defined as follows:

$$W(u, c) = \max_{s \in \mathbf{R}_c} \chi'(u, s) K_u,$$

where $\chi'(u, s)$ denotes the preference level for user u to upload his generated contents to a server at region s, and K_u is the average amount of content that can be generated by user u in the next time slot. A larger local upload index $W(u, c)$ indicates that better upload gain can be achieved when u is hosted with the cloud provider c.

In our experiments, $\chi(v, s)$ and $\chi'(v, s)$ can be estimated either by the geographic distance between the user and the server region, or using the historical network performance.

Cloud-provider preference of a user. In our design, the overall cloud-provider preference of a user takes both the local download performance (for the user's friends) and local upload performance (for the user himself) into consideration. The overall cloud-provider preference of a user is then the combination of the two indices. We denote $\psi(u, c)$ as user u's preference of cloud provider c, defined as follows:

$$\psi(u, c) = Q(u, c) + \beta_u W(u, c), \quad (1)$$

where β_u is an implementation parameter used to combine the two indices, depending on the characteristics of a user, *e.g.*, a large β_u for a user who frequently generates and uploads contents from a mobile device, so that a cloud provider with servers the user can upload content fast to will have a larger preference index $\psi(u, c)$ with the user.

3.1.3 Replication Cost Due to Inter-Cloud Propagation

As a unique cost in the multi-cloud hosting, the inter-cloud traffic cost is caused by the social propagation between users that are hosted with different cloud providers, due to the pricing scheme of the cloud providers we have shown in Sec. 2. In the online social network, the common social activities such as sharing contents [3] make the contents associate with different users dynamically.

Due to the high cost of replicating contents from servers in one cloud to servers in another cloud, we need to take the cost of inter-cloud content replication caused by social propagation into account. We define the replication cost for user u hosted with cloud c as follows:

$$Y(u,c) = \sum_{v \in \mathbf{F}_u} \frac{1}{2}(p_{C_u C_v} e_{uv} + p_{C_v C_u} e_{vu}),$$

where p_{cd} is the data transfer price for replicating a content from cloud c to cloud d, *i.e.*, the price that the instant social video sharing system has to pay when contents are replicated between two cloud providers rather than inside one cloud provider. C_u is the cloud with which user u is hosted. $p_{C_u C_v} e_{uv}$ is the cost of inter-cloud traffic of social connection (u,v), depending on the actual pricing scheme used by cloud providers. Next, we will discuss how these objectives are achieved in our multi-cloud hosting design.

3.1.4 Problem Formulation and Analysis

Optimization. To satisfy users' cloud-provider preference as well as reducing the inter-cloud propagation, the multi-cloud hosting can be formulated as an optimization problem by combining the two objectives, as follows:

$$\max_{\{C_u | u \in \mathbf{V}\}} \sum_{u \in \mathbf{V}} \alpha \psi(u, C_u) - (1 - \alpha)\gamma Y(u, C_u), \quad (2)$$

subject to:

$$C_u \in \mathbf{C}, \forall u \in \mathbf{V},$$

where C_u is the objective variable that determines the cloud provider with which user u is hosted, γ is the parameter used to combine the inter-cloud *propagation cost* (The cost caused by content replication across boundaries of different cloud servers due to social propagation) with the users' cloud-provider preference, and α is the parameter to adjust the weight between the two objectives. The optimization variables give the choices of cloud providers for users in the online social network.

Proof of NP-Hardness. The optimization to determine the multi-cloud hosting is NP-hard in general.

Proof: To prove this, we reduce a MCP (Multiterminal Cut Problem) [18], which is NP-hard, to it. In the MCP, we are given an edge-weighted graph G and a subset of k vertices called terminals, and asked for a minimum weight set of edges that separates each terminal from all the others. Next, we show that the MCP can be reduced to the multi-cloud hosting problem. We build a

social graph G' which has the same structure with G. The reduction is as follows. (1) We have k cloud providers $c_1, c_2, \ldots, c_k$ for the multi-cloud hosting, and the data-transfer price between any two cloud providers is 1. (2) We let the propagation weight of a social connection in G' be the same as the edge weight of the corresponding edge in G. (3) Without loss of generality, we let users $u_1, u_2, \ldots, u_k$ in G' be the ones corresponding to the k terminals in G, and we assign the cloud-provider preference of users as follows:

$$\begin{cases} \psi(u_i, c_j) = 0, & i > k, j = 1, 2, \ldots, k \\ \psi(u_i, c_j) = 0, & j \neq i, i = 1, 2, \ldots, k \ , \\ \psi(u_i, c_j) = L, & j = i, i = 1, 2, \ldots, k \end{cases}$$

where L is a const cloud-provider preference, which can be assigned with a value large enough (*e.g.*, the sum of all propagation weight) so that every user $u_i, i = 1, 2, \ldots, k$ has to be hosted with the cloud c_i to achieve the optimal multi-cloud hosting. Thus, the k users will be separated from each other in the multi-cloud hosting (they are hosted with different cloud providers respectively) while the overall social propagation is minimized. If the multi-cloud hosting problem can be solved, then the solution of the original MCP can be achieved as well, *i.e.*, the set of edges corresponding to the social connections between any two cloud providers. Thus, the multi-cloud hosting problem is NP-hard. □

3.2 Heuristic Multi-Cloud Hosting

Based on our measurement insights, we design a heuristic algorithm to solve the multi-cloud hosting problem. Algorithm 1 presents our strategy to partition the social graph, determining which users should be hosted with which cloud providers. Our algorithm includes two phases as follows: (1) In the initial preference-aware cloud selection (lines 2 – 4), a user is assigned to a cloud provider according to only the cloud-provider preference of hosting a user ($\psi(u,c)$), without considering the social propagation; (2) In the propagation-aware re-hosting (lines 5 – 14), pairs of users are assigned with different cloud providers to reduce the inter-cloud propagation. We will present the two phases as follows.

3.2.1 Preference-Aware Initial Hosting

In this phase, an initial cloud provider is assigned to host each user, only according the cloud-provider preference ($\psi(u,c)$) of users. For each user in the online social network, the ideal cloud provider is the one that can maximize its cloud-provider preference among all available cloud providers (line 3). After the initial cloud selection, users are assigned to the cloud providers that can maximize their preference; however, such assignment can result in a large cost of inter-cloud propagation ($Y(u,c)$). Next, users are adjusted to reduce the inter-cloud propagation.

Algorithm 1 Heuristic multi-cloud hosting algorithm

1: **procedure** MULTI-CLOUD HOSTING
2:    **for all** $u \in \mathbf{V}$ **do**
3:       $C_u \leftarrow \arg\max_{c \in \mathbf{C}} \psi(u, c)$
4:    **end for**
5:    Sort social connections in their propagation weight's descending order
6:    **repeat**
7:       $\{u, v\}, C_u \neq C_v \leftarrow$ head of the sorted list
8:       Calculate γ_A, γ_B, γ_C and γ_D in Eq. (3)
9:       $\gamma'_X = \max\{\gamma_A, \gamma_B, \gamma_C, \gamma_D\}$
10:      **if** $X \in \{B, C, D\}$ **then**
11:        Re-host user u and user v accordingly (Sec. 3.2.2)
12:      **end if**
13:      Remove social connection $\{u, v\}$ from the sorted list
14:    **until** $\gamma'_X < \eta$
15: **end procedure**

3.2.2 Propagation-Aware Re-Hosting

Our re-hosting strategy is to change the cloud providers for users so that the inter-cloud replication cost can be reduced. The re-hosting procedure works as follows.

First, the social connections are ranked between users that are not hosted with the same cloud in the descending order of the propagation weight (line 5). A pair of users who have a large propagation weight between them are can be hosted with the same cloud to reduce the inter-cloud replication cost.

Second, we present the re-hosting approach. Let u and v denote the pair of users whose social connection has the largest propagation weight. Fig. 9 illustrates the schemes that we can apply to improve the partition: (1) keep the original hosting strategy, as illustrated in Fig. 9(A); (2) re-host u to the cloud by which v is hosted, as illustrated in Fig. 9(B); (3) re-host v to the cloud by which u is hosted, as illustrated in Fig. 9(C); and (4) re-host both u and v to a new cloud provider f, as illustrated in Fig. 9(D). We can improve the performance by applying one among the four strategies, to re-host u and v with one cloud.

Third, to determine which strategy is used to re-host user u and user v, we use the following heuristic: we define a gain $\gamma_X, X \in \{A, B, C, D\}$ for each re-hosting scheme. γ_A is the gain of hosting user u with cloud c, and user v in cloud d, without any change to the original hosting (it is the baseline); γ_B is the gain of hosting users u and v with cloud d; γ_C is the gain of hosting users u and v with cloud c; and γ_D is the gain of hosting users u and v with a new cloud provider f. The gain is defined

as follows:

$$
\begin{aligned}
\gamma_A &= 0 \\
\gamma_B &= \alpha[\psi(u, d) - \psi(u, c)] - (1 - \alpha)\phi(u \to d) \\
\gamma_C &= \alpha[\psi(v, c) - \psi(v, d)] - (1 - \alpha)\phi(v \to c) \quad , \quad (3) \\
\gamma_D &= \max_{f \neq c, d}\{\alpha[\psi(u, f) + \psi(v, f) - \psi(u, c) \\
&\quad - \psi(v, d)] - (1 - \alpha)\phi(u, v \to f)\}
\end{aligned}
$$

where the first part ($\alpha[\cdot]$) is the gain of improving the cloud-provider preference of user u and v according to the re-hosting, and the second part ($-(1 - \alpha)\phi(\cdot)$) represents the gain by reducing the inter-cloud replication cost under different re-hosts. $\phi(\cdot)$ is defined as follows:

$$
\begin{cases}
\frac{1}{2}\sum_{w|C(w)=c}\gamma(p_{cd}e_{wu} + p_{dc}e_{uw}) \\
\quad -\frac{1}{2}\sum_{w|C(w)=d}\gamma(p_{cd}e_{uw} + p_{dc}e_{wu}) & u \to d \\[6pt]
\frac{1}{2}\sum_{w|C(w)=d}\gamma(p_{dc}e_{wv} + p_{cd}e_{vw}) \\
\quad -\frac{1}{2}\sum_{w|C(w)=c}\gamma(p_{dc}e_{vw} + p_{cd}e_{wv}) & v \to c \\[6pt]
\frac{1}{2}\sum_{w|C(w)=c}\gamma(p_{cd}e_{wu} + p_{dc}e_{uw}) \\
\quad -\frac{1}{2}\sum_{w|C(w)=f}\gamma(p_{cf}e_{uw} + p_{fc}e_{wu}) \\
\quad +\frac{1}{2}\sum_{w|C(w)=d}\gamma(p_{dc}e_{wv} + p_{cd}e_{vw}) & u, v \to f \\
\quad -\frac{1}{2}\sum_{w|C(w)=f}\gamma(p_{df}e_{vw} + p_{fd}e_{wv})
\end{cases}
$$

The re-hosting is performed according to the value of the gains. Among γ_A, γ_B, γ_C and γ_D, if γ_A is the largest, u will remain in cloud c and v in cloud d; if γ_B is the largest, u will be hosted with cloud d; if γ_C is the largest, v will be hosted with cloud c; otherwise, v will be hosted and u with a new cloud provider f, which can maximize the re-hosting gain.

Fourth, in our heuristic re-hosting, social connections with the largest propagation weight are first processed. In each pass, we only consider the social connections that can incur a high cost of inter-cloud traffic (as observed in our measurement study, the fraction of such social connections is very small). A threshold η is used to determine which social connections are considered in the re-hosting. In our experiments, η is selected to terminate the re-hosting loop when the fraction of social connections touched exceeds 20% of all the social connections. The rationale is that, according to our measurement insight in Sec. 2.5.2, in a real online social video sharing system, only the most "active" social connections will affect the replication cost.

According to our design, the complexity of the algorithm is determined by the procedure of ranking the social connections, as only a constant fraction of the most influential social connections are considered in the re-hosting. The complexity of the algorithm is thus $\mathcal{O}(|\mathbf{E}| \log |\mathbf{E}|)$, since only a constant fraction of the most influential social connections are considered in our algorithm. In a social network, as $|\mathbf{V}|$ is similar to $|\mathbf{E}|$ [19], the complexity is then similar to $\mathcal{O}(|\mathbf{V}| \log |\mathbf{V}|)$.

4 EXPERIMENTAL RESULTS

In this section, we evaluate the performance of our design using simulation experiments driven by the Weibo traces. In our experiments, we will study the satisfaction
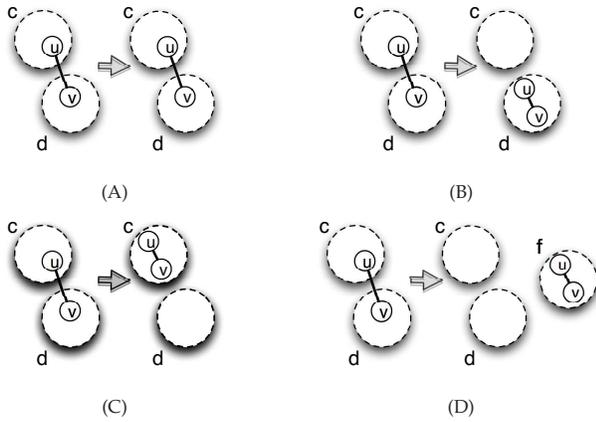
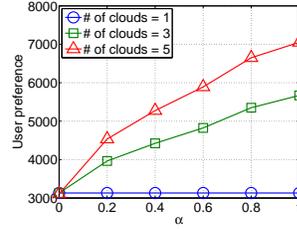Fig. 9. Strategies of re-hosting a pair of users.

Fig. 10. Satisfaction of users' cloud-provider preference.
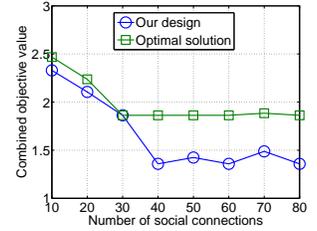

Fig. 11. Comparison between our algorithm and the brute-force algorithm with the optimal solution.

of users' cloud-provider preference, the reduction of replication cost caused by inter-cloud social propagation, and the efficiency of the heuristic algorithm.

4.1 Experiment Setup

Users and social graph. We have used a sample of $20,000$ users selected from the social graph of Tencent Weibo, by a BFS-based collection from 10 random seed users, *i.e.*, we initialize a user set with the 10 seed users, and iteratively add friends of users that are already in the set, until the size of the set reaches $20,000$ or it is self-contained.

Content generation and propagation. We also use the traces of Tencent Weibo to drive the experiments. The actions of posting microblogs in the traces are considered as generating new contents, and the actions of sharing microblogs are used to weight the social propagation between the users. Contents generated and shared need to be hosted with the cloud providers. Based on the traces, we perform the social graph partition for the multi-cloud hosting.

Cloud regions and prices. In these Weibo traces, about 80 regions are observed to have users located in. We have randomly assigned each of these regions to one of 6 cloud providers. We assume the cloud providers have the same pricing scheme: (1) the price of outgoing traffic to a different cloud provider is 1, and the price of outgoing traffic to the same cloud provider is 0, *i.e.*, the roadblock of inter-cloud replication; and (2) the price of incoming traffic is 0, *i.e.*, the encouragement of incoming contents from the Internet.

According to our design in Sec. 3, the locations of users are used to calculate their preference of cloud providers, according to Eq. 1.

Next, we present the results in our evaluation.

4.2 Performance Evaluation

4.2.1 Satisfying Users' Cloud Preference

First, we study how users' cloud-provider preference is satisfied in our design. In our experiments, a user's preference of a cloud provider is normalized, *i.e.*, the sum of the preference of all the cloud providers is 1. We denote the normalized preference as $\bar{\psi}(u, C_u)$, and use the overall satisfaction of user preference ($\sum_u \bar{\psi}(u, C_u)$) as the metric to evaluate the performance.

We study the satisfaction of user preference under different weight of parameter α. Fig. 10 illustrates the overall cloud-provider preference versus α. Different curves are generated under different numbers of available cloud providers (out of all the 6 cloud providers) for the multi-cloud hosting. We observe a general increase of users' cloud-provider preference as α grows using more than 1 cloud providers, since the users' cloud-provider preference is more considered when α is larger. We also observe that when more cloud providers are available for the social video sharing system to choose from, the overall cloud-provider preference can be improved.

4.2.2 Reducing Inter-Cloud Propagation

Next, we study the cost caused by inter-cloud propagation. To evaluate the cost, we define two metrics: (1) the propagation cost of all inter-cloud social connections calculated as $\sum_{u,v} p_{C_u C_v} e_{uv}$; and (2) the number of social connections that connect users hosted with different cloud providers.

In Fig. 12, we study the impact of α on the cost of inter-cloud propagation. The curves in Fig. 12(a) illustrate the propagation cost against α. We observe that the inter-cloud replication cost generally increases as α grows, since the weight of the inter-cloud propagation becomes smaller in the social graph partition. We also observe that the cost increases much faster when α is larger. This result indicates that strategies only considering users' cloud-provider preference can incur a high inter-cloud propagation cost for the social video sharing system.

The curves in Fig. 12(b) illustrate the numbers of the inter-cloud social connections. We observe that more social connections span different cloud providers when α grows, and the increase speed of the number of inter-cloud social connections as α increases is much more linear than the propagation cost. The reason is that our algorithm tends to divide friends with small propagation weight between them into different cloud providers, but
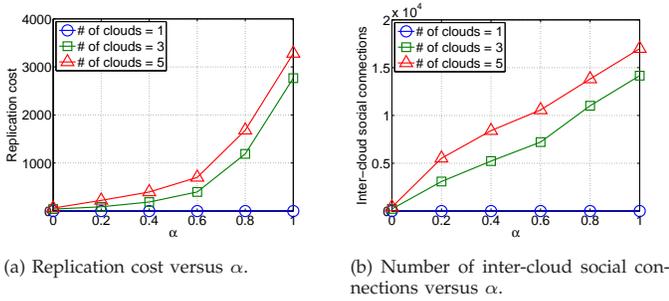
(a) Replication cost versus α.

(b) Number of inter-cloud social connections versus α.

Fig. 12. Replication cost caused by inter-cloud social propagation.

(a) User preference versus the number of cloud providers to host the instant social video sharing system.

(b) Replication cost versus the number of cloud providers to host the instant social video sharing system.

Fig. 13. Comparison of different multi-cloud hosting algorithms under different number of cloud providers.

keep them in the same cloud if the propagation weight is large.

4.2.3 Performance Comparison

In our experiments, we have also compared our algorithm (with $\alpha = 0.5$) with the following strategies: (1) Random partition, in which users are randomly assigned to the cloud providers; (2) Min-propagation, in which users are partitioned to minimize the inter-cloud propagation; and (3) Max-preference, in which users are hosted with their ideal cloud providers.

We study their performance by varying the number of cloud providers that are available for the instant social video sharing system to perform the multi-cloud hosting. In Fig. 13(a), we first study the satisfaction of users' cloud-provider preference. We observe that the user preference increases in both our design and the max-preference strategy; while the other two strategies cannot benefit from the availability of more cloud providers. We also observe that the max-preference strategy outperforms our design by about 12% when all the 6 cloud providers can be utilized in the multi-cloud hosting.

However, the max-preference algorithm incurs much larger inter-cloud replication cost due to the social propagation over the social connections between users hosted with different cloud providers. The curves in Fig. 13(b) illustrate the propagation cost against the number of available cloud providers. We observe that both our design and the min-propagation strategy remain very low inter-cloud propagation cost as the number of cloud providers increases, while the cost increases in both the max-preference and random strategies as the number of cloud providers increases. The inter-cloud propagation cost in the max-preference strategy is about 6 times larger than that in our design when all the cloud providers can be utilized. The results indicate that our design can well balance users' cloud-provider preference and the cost of inter-cloud propagation.

4.2.4 Effectiveness of the Heuristic Algorithm

Since finding the optimal solution for deploying the social video sharing system is generally NP-Hard, we present the effectiveness of our heuristic algorithm. In particular, we compare the combined objective defined
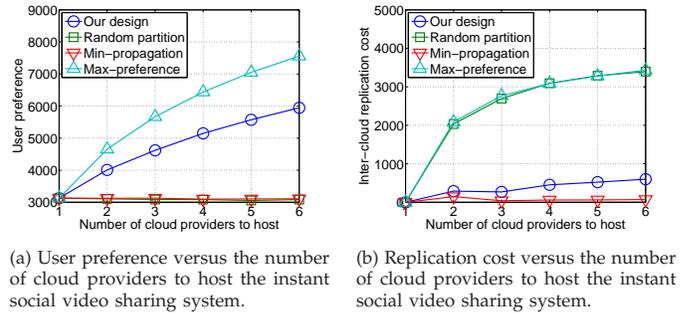
in Eq. 2 achieved by our algorithm, with the optimal value achieved by a brute-force searching. In this experiment, we generate a graph with 10 nodes (users), who have random preferences of 3 cloud providers.

By varying the number of the directed social connections between them, we compare the combined objective value achieved by both algorithms in Fig. 11. We observe that when the number of social connections (edges) is under 30 (i.e., $1/3$ of all possible social connections), our algorithm achieves similar performance as the optimal solution does — in a real social graph, the number of social connections is much smaller than that [20]. Thus, our algorithm has relatively good performance for the real online social network systems.

5 RELATED WORK

5.1 Growth of Online Social Media and Cloud-based Hosting

The philosophy of social media is to let users in an online social network not only generate the content, but also disseminate the generated content through social connections [21], including the "following" relationship in a Twitter-like system, and the "friending" relationship in a Facebook-like system.

Cloud computing has been widely used to deploy the social media systems. As both the number and the geographic distribution of the users in an online social media service are expanding, hosting such a social media system can take full advantage of the elastic cloud resource [14]. In [14], social graph is studied for locality partition, while in our study, we design partition by studying social propagation, which is determined by not only the social graph, but also user behaviors.

There are several works on hosting a social media system with different cloud computing platforms. Cheng et al. [5] have studied the migration of socialized videos in YouTube to the cloud so as to balance the load between servers. Wu et al. [6] have studied how to scale the social media service using the geo-distributed cloud resource. In our previous study, we presented that using a edge cloud framework, users can benefit from downloading from local servers [15] in social video streaming.

5.2 Instant Social Video Delivery

Today, conventional content delivery strategies are still used for delivering social contents, including the conventional CDN and P2P-based approaches [22]. Given the crowdsourced content capturing and sharing, the preferred length of online videos becomes shorter and shorter. Taking Vine as a case study, Zhang et al. [1] show that the instant social videos have short lifetime and highly skewed popularity that fast decays over time. Videos in these social trending media become more fragmented and instantaneous, which have challenged the conventional content replication and streaming strategies. Our study is to bridge the gap between the conventional content delivery strategies and instant social video delivery requirements.

5.3 Social Propagation

In a social media system, contents propagate among users due to a variety of social activities, *e.g.*, users can *reshare* contents that are originally generated by their friends, so as to make the contents available to more people in the online social network. To efficiently serve the social media contents, the propagation information has to be considered for the service deployment [23].

In an online social network, contents can be dynamically shared by social groups with very different size and geographic distribution [24]. As a result, propagation inference has become an important factor for improving the performance of social media services — a number of research efforts have been devoted to studying the content propagation in social media [2], including the traditional message propagation in Twitter-like microblogging systems [25], as well as the video propagation in YouTube-like systems [26]. Xu *et al.* [27] studied how to forecast video popularity of social video contents, and observed that social propagation is a critical factor for predicting social video contents.

5.4 Graph Partition for Distributed Social Media

A fundamental problem in hosting social video contents with a distributed system is the partition of contents and users in the social graph. Tran *et al.* [28] have studied the partition of contents in an online social network by taking users' social relationship into consideration. Newman *et al.* [29] have studied the community structure in the social network. Carrasco *et al.* [30] have proposed to partition the social graph by dividing users' activities into different time phases, since the propagation levels between two users vary over time.

These works have studied the hosting of a social media system in the context of a single cloud provider, or the cloud servers are treated equally even when they are allocated from different cloud providers — the replication roadblock across the boundary among different cloud providers does not exist. However, this assumption is no longer true under the pricing scheme of today's cloud providers [10].

In this paper, we seek to design a multi-cloud hosting strategy based on a social graph partition, which jointly takes users' cloud-provider preference, the content propagation between users, and the replication roadblock across the boundary between cloud providers caused by their pricing schemes into account.

6 CONCLUDING REMARKS

In this paper, we have studied hosting an instant social video sharing service with multiple cloud providers. Our measurement studies not only confirm the benefit of the multi-cloud hosting, but also reveal several guidelines for the multi-cloud hosting design. The multi-cloud hosting problem to optimize a combination of satisfying users' cloud-provider preference and reducing the cost caused by inter-cloud social propagation is proven to be NP-hard in general. We design a heuristic algorithm to solve the problem, by iteratively partitioning a propagation-weighted social graph — based on an initial preference-aware partition, a propagation-aware re-hosting dynamically reduces the inter-cloud propagation of the most active social connections. Trace-driven simulations further demonstrate that our heuristic can efficiently solve the multi-cloud hosting problem, and our design achieves a good balance of the two objectives under acceptable complexity — with only 12% user preference degradation, our algorithm reduces $5/6$ of the inter-cloud data transfer cost.

REFERENCES

[1] L. Zhang, F. Wang, and J. Liu, "Understand Instant Video Clip Sharing on Mobile Platforms: Twitter's Vine as a Case Study," in *ACM Network and Operating System Support on Digital Audio and Video Workshop (NOSSDAV)*, 2014.

[2] H. Kwak, C. Lee, H. Park, and S. Moon, "What Is Twitter, a Social Network or a News Media?" in *ACM International Conference on World Wide Web (WWW)*, 2010, pp. 591–600.

[3] M. Wasko and S. Faraj, "Why Should I Share? Examining Social Capital and Knowledge Contribution in Electronic Networks of Practice," *Mis Quarterly*, pp. 35–57, 2005.

[4] M. Hajjat, X. Sun, Y. Sung, D. Maltz, S. Rao, K. Sripanidkulchai, and M. Tawarmalani, "Cloudward Bound: Planning for Beneficial Migration of Enterprise Applications to the Cloud," vol. 40, no. 4, pp. 243–254, 2010.

[5] X. Cheng and J. Liu, "Load-Balanced Migration of Social Media to Content Clouds," in *ACM Network and Operating System Support for Digital Audio and Video (NOSSDAV)*, 2011.

[6] Y. Wu, C. Wu, B. Li, L. Zhang, Z. Li, and F. C. Lau, "Scaling Social Media Applications into Geo-Distributed Clouds," in *IEEE International Conference on Distributed Computing Systems (ICDCS)*, 2012.
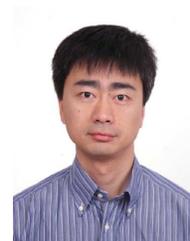
[7] H. Hu, Y. Wen, T.-S. Chua, Z. Wang, J. Huang, W. Zhu, and D. Wu, "Community Based Effective Social Video Contents Placement in Cloud Centric CDN Network," in *IEEE International Conference on Multimedia and Expo (ICME)*, 2014.

[8] X. Liu, F. Dobrian, H. Milner, J. Jiang, V. Sekar, I. Stoica, and H. Zhang, "A Case for a Coordinated Internet Video Control Plane," in *ACM SIGCOMM*, 2012.

[9] H. Liu, Y. Wang, Y. Yang, A. Tian, and H. Wang, "Optimizing Cost and Performance for Content Multihoming," in *ACM SIGCOMM*, 2012, pp. 371–382.

[10] "Amazon Data Transfer Price (Last accessed: November, 2014)," http://aws.amazon.com/ec2/pricing/#DataTransfer.

[11] S. Ye and S. F. Wu, "Measuring Message Propagation and Social Influence on twitter.com," in *Social informatics*. Springer, 2010, pp. 216–231.

[12] R. Krishnan, H. Madhyastha, S. Srinivasan, S. Jain, A. Krishnamurthy, T. Anderson, and J. Gao, "Moving Beyond End-to-End Path Information to Optimize CDN Performance," in *ACM Internet Measurement Conference (IMC)*, 2009.

[13] "Facebook photos," http://gizmodo.com/5937143/what-facebook-deals-with-everyday-27-billion-likes-300-million-photos-uploaded-and-500-terabytes-of-data.

[14] J. Pujol, V. Erramilli, G. Siganos, X. Yang, N. Laoutaris, P. Chhabra, and P. Rodriguez, "The Little Engine(s) That Could: Scaling Online Social Networks," *ACM SIGCOMM Computer Communication Review*, vol. 40, no. 4, pp. 375–386, 2010.

[15] Z. Wang, W. Zhu, X. Chen, L. Sun, J. Liu, M. Chen, P. Cui, and S. Yang, "Propagation-based social multimedia distribution," *ACM Transactions on Multimedia Computing, Communications and Applications*, vol. 9, no. 1, pp. 52–72, October 2013.

[16] [Online]. Available: http://aws.amazon.com/ec2/

[17] [Online]. Available: http://wiki.open.qq.com/wiki/

[18] E. Dahlhaus, D. Johnson, C. Papadimitriou, P. Seymour, and M. Yannakakis, "The Complexity of Multiterminal Cuts," *SIAM J. Comput.*, vol. 23, no. 4, pp. 864–894, 1994.

[19] G. Caldarelli, "Scale-free networks: complex webs in nature and technology," *OUP Catalogue*, 2007.

[20] A. Mislove, M. Marcon, K. Gummadi, P. Druschel, and B. Bhattacharjee, "Measurement and Analysis of Online Social Networks," in *ACM Internet Measurement Conference (IMC)*, 2007.

[21] A. Kaplan and M. Haenlein, "Users of the World, Unite! the Challenges and Opportunities of Social Media," *Business horizons*, vol. 53, no. 1, pp. 59–68, 2010.

[22] N. Carlsson, G. Dán, D. Eager, and A. Mahanti, "Tradeoffs in Cloud and Peer-Assisted Content Delivery Systems," in *IEEE International Conference on Peer-to-Peer Computing (P2P)*, 2012.

[23] H. Yoganarasimhan, "Impact of Social Network Structure on Content Propagation: a Study Using YouTube Data," *Quantitative Marketing and Economics*, vol. 10, no. 1, pp. 111–150, 2012.

[24] L. Backstrom, E. Sun, and C. Marlow, "Find Me if You Can: Improving Geographical Prediction With Social and Spatial Proximity," in *ACM International Conference on World Wide Web (WWW)*, 2010, pp. 61–70.

[25] S. Petrovic, M. Osborne, and V. Lavrenko, "RT to Win! Predicting Message Propagation in Twitter," in *International Conference on Weblogs and Social Media (ICWSM)*, 2011.

[26] H. Li, J. Liu, K. Xu, and S. Wen, "Understanding Video Propagation in Online Social Networks," in *IEEE International Workshop on Quality of Service (IWQoS)*, 2012.

[27] J. Xu, M. van der Schaar, J. Liu, and H. Li, "Forecasting Popularity of Videos using Social Media," *arXiv preprint arXiv:1403.5603*, 2014.

[28] D. Tran, K. Nguyen, and C. Pham, "S-CLONE: Socially-Aware Data Replication for Social Networks," *Computer Networks*, 2012.

[29] M. Newman and M. Girvan, "Finding and Evaluating Community Structure in Networks," *Physical review E*, vol. 69, no. 2, p. 026113, 2004.

[30] B. Carrasco, Y. Lu, and J. Trindade, "Partitioning Social Networks for Time-Dependent Queries," in *ACM Workshop on Social Network Systems*, 2011.

Zhi Wang received his B.E. and Ph.D. degrees from Department of Computer Science and Technology, Tsinghua University, in 2008 and 2014. He is currently an assistant professor in Graduate School at Shenzhen, Tsinghua University. His research areas include online social media, mobile cloud computing and large-scale multimedia systems. He received Outstanding Doctoral Dissertation Award from China Computer Federation, Best Paper Award at ACM Multimedia 2012, and Best Student Paper Award at MMM 2015.

Baochun Li received the B.E. degree from the Department of Computer Science and Technology, Tsinghua University, China, in 1995 and the M.S. and Ph.D. degrees from the Department of Computer Science, University of Illinois at Urbana-Champaign, Urbana, in 1997 and 2000. Since 2000, he has been with the Department of Electrical and Computer Engineering at the University of Toronto, where he is currently a Professor. His research interests include large-scale distributed systems, cloud computing, peer-to-peer networks, applications of network coding, and wireless networks. He was the recipient of the IEEE Communications Society Leonard G. Abraham Award in the Field of Communications Systems in 2000. In 2009, he was a recipient of the Multimedia Communications Best Paper Award from the IEEE Communications Society, and a recipient of the University of Toronto McLean Award. He is a member of ACM and a Fellow of IEEE.

Lifeng Sun received his B.S. and Ph.D. degrees in System Engineering in 1995 and 2000 from National University of Defense Technology, Changsha, Hunan, China. He was an postdoctoral fellow from 2001 to 2003, an assistant professor from 2003 to 2007, an associate professor from 2007 to 2013, and currently a professor all in the Department of Computer Science and Technology at Tsinghua University. His research interests lie in the areas of online social network, video streaming, interactive multi-view video, and distributed video coding. He is a member of IEEE and ACM.

Wenwu Zhu received the Ph.D. degree from New York University Polytechnic School of Engineering in 1996 in Electrical and Computer Engineering. He is with Computer Science Department of Tsinghua University as Professor of "1000 People Plan" of China. His current research interests are in the area of multimedia cloud computing, social media computing, multimedia big data, and multimedia communications and networking. He served(s) on various editorial boards, such as Guest Editor for the Proceedings of the IEEE, IEEE T-CSVT, and IEEE JSAC; Associate Editor for IEEE Transactions on Mobile Computing, IEEE Transactions on Multimedia, and IEEE Transactions on Circuits and Systems for Video Technology; Leading Editor of the Area "Computer Networks and Distributed Computing" of Journal of Computer Science and Technology. He received the Best Paper Award in ACM Multimedia 2012, the Best Paper Award in IEEE Transactions on Circuits and Systems for Video Technology in 2001, and the other 4 international Best Paper Awards.

Shiqiang Yang received the B.E. and M.E. degrees in Computer Science from Tsinghua University, Beijing, China in 1977 and 1983, respectively. From 1980 to 1992, he worked as an assistant professor at Tsinghua University. He served as the associate professor from 1994 to 1999 and then as the professor since 1999. From 1994 to 2011, he worked as the associate header of the Department of Computer Science and Technology at Tsinghua University. He is currently the President of Multimedia Committee of China Computer Federation, Beijing, China and the co-director of Microsoft-Tsinghua Multimedia Joint Lab, Tsinghua University, Beijing, China. His research interests mainly include multimedia procession, media streaming and online social network. He is a senior member of IEEE.