Optimized Resource Distribution for Interactive TV Applications

Gosala Kulupana, Dumidu S. Talagala, Hemantha Kodikara Arachchi, and Anil Fernando
Centre for Vision, Speech and Signal Processing, University of Surrey, United Kingdom
Email: {g.kulupana, d.talagala, h.kodikaraarachchi, w.fernando}@surrey.ac.uk

***Abstract*– Personalized interactive television broadcasting requires real-time audiovisual processing at levels impractical in end-user equipment. However, guaranteeing Quality of Service (QoS) also remains a challenge for cloud-based solutions. This paper proposes a group QoS optimization approach to the problem that demonstrates significant improvements in the number of users being served.**

I. Introduction

During the past decade, Interactive Television (ITV) based applications have gained a huge popularity within the consumer entertainment sector [1]. In order to facilitate the next generation of ITV, the ACTION-TV [2] project proposes a novel way of personalizing TV shows for a large number of concurrent social groups, in such a way that some individuals from each group can participate in the show while the rest enjoy the performance of their social peers. For example, in a quiz show, an interested user (i.e., active interactive-user) can become one of the contestants from the comfort of home and only his/her social group (i.e., passive interactive-users) can view his/her appearance.

Consumer devices may not be able to support this type of interactive multimedia applications due to insufficient computational resources. Therefore, cloud support is a necessity for realizing these applications. Hence, an inter-cloud [3] based media processing architecture has been proposed for personalizing broadcast media [2]. Here, the users' activity data is acquired using consumer grade sensing devices such as the Kinect, and is uploaded to cloud-based media processing servers which personalize the broadcast media by re-authoring the show with the help of pre-prepared photorealistic 3D models.

When the number of concurrent social groups proliferates, the cloud and communication infrastructure can be exhausted rapidly. Moreover, the communication delay between active interactive-users and the processing server must be maintained at an acceptable level to improve the user experience. Therefore, in the context of an application such as ACTION-TV, the design challenge is the selection of the cloud that hosts the processing server for each social group, such that overall users' QoS is maximized. Since users of a given social group may be connected to different clouds in a content delivery network (CDN) architecture as shown in Fig. 1, the optimal location of the processing cloud must be determined based on the network conditions and data center availability. In addition, the processing requirements on the cloud infrastructures scale linearly with the number of social groups. Thus, optimum cloud resource allocation becomes even more challenging with the increasing numbers of users, social groups, clouds and network conditions. This paper proposes a group based QoS maximization approach to determine the processing cloud allocation for personalizing ITV content.

This work was supported by the EU FP7 project ACTION-TV (http://www.action-tv.net), funded under the European Commission's 7th Framework Program (Grant Number: 611761).

II. Simultaneous Group QoS Maximization

In this section we introduce the ITV system, and describe the optimization problem and the associated constraints.

A. System Description

Fig. 1 illustrates an example logical network diagram of an ITV system. For simplicity, it illustrates two user groups (i.e., social groups) connected to two Internet Service Providers (ISPs) and three potential computational resources (i.e., clouds). Users are categorized into "Actively Interacting (AI)" and "Passively Interacting (PI)" users, whose content are processed by a third "processing cloud" as indicated.

The general problem for multiple user groups and cloud resources can therefore be expressed as follows. Let S denote the set of J clouds available in the network, U be the set of K ITV viewers belonging to N personalized social groups, T be the set of $J\times J$ virtual links between clouds, and E the set of K links between individual users and their ISPs. Thus,

$$\begin{aligned} S &= \{s_1, s_2, \ldots, s_J\} \\ T &= \{t_{1,1}, t_{1,2}, \ldots, t_{J,J}\} \\ E &= \{e_1, e_2, \ldots, e_K\} \\ U^{AI} &= \{u_{1,1}^{AI}, u_{1,2}^{AI}, \ldots, u_{n,L_n}^{AI}, \ldots, u_{N,L_N}^{AI}\} \\ U^{PI} &= \{u_{1,1}^{PI}, u_{1,2}^{PI}, \ldots, u_{n,M_n}^{PI}, \ldots, u_{N,M_N}^{PI}\} \\ U &= U^{AI} \cup U^{NI}, \end{aligned} \tag{1}$$

where $u_{n,y}^{x}$ denotes an individual user for $x \in \{AI, PI\}$, $y \in \{L_n, M_n\}$ (L_n, M_n correspond to the number of AI and PI users of the n^{th} social group respectively), and $n = 1, \ldots N$.

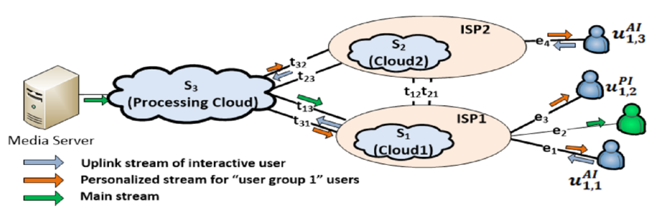

Fig. 1. Logical network diagram of two social groups in the ITV system.

B. Group Quality of Service Parameterization

The QoS of a particular user u (discarding the subscripts and superscripts for clarity) is a function of end-to-end network QoS parameters such as delay D^u, bandwidth B^u, jitter J^u and packet loss L^u. The impact (i.e., the cost) of processing the interactive streams at the j^{th} cloud for the user u of the n^{th} social group can therefore be expressed as [4]

$$c_{n,j}^{u} = \alpha_1 \times (L^u) + \alpha_2 \times (J^u) + \alpha_3 \times (D^u) + \alpha_4 \times (1/B^u), \tag{2}$$

where $\{\alpha_1, \alpha_2, \alpha_3, \alpha_4\}$ are appropriately selected constants that adequately parameterize the cost in terms of the QoS metrics. The impact on the n^{th} social group is defined by a function f_G that maps the individual user costs to a group cost, given by

$$c_{n,j}^G = f_G(\{c_{n,j}^1,\ldots,c_{n,j}^u,\ldots\}) \qquad u \mapsto u_{n,y}^x . \qquad (3)$$

Minimizing the group cost function in (3) therefore implies a maximization of the group QoS and also the user experience.

C. Optimization Criterion

In order to maximize the group QoS of the overall system, a criterion that optimizes the processing resource distribution within the system must be defined. To this effect, we first express the group costs in the matrix form

$$\boldsymbol{c}_{(N\times 1)}^G = \mathbf{C}_{(N\times NJ)}\boldsymbol{d}_{(NJ\times 1)}. \qquad (4)$$

Here, $\boldsymbol{d} = [\boldsymbol{d}_1^T, \ldots \boldsymbol{d}_n^T, \ldots \boldsymbol{d}_N^T]^T$, where $\boldsymbol{d}_n$ is a binary $J\times 1$ vector with a single nonzero element corresponding to its processing cloud index (e.g., $\boldsymbol{d}_n = [0\ 1\ 0]^T$ when the n^{th} social group is processed at the cloud s_2 in Fig. 1). $\mathbf{C}$ therefore becomes a block diagonal matrix whose the n^{th} row can be expressed as $[\,\mathbf{0}_{(1\times(n-1)J)}, [\,c_{n,1}^G, c_{n,2}^G, \ldots, c_{n,J}^G\,]_{(1\times J)}, \mathbf{0}_{(1\times(N-n)J)}\,]$.

We compute $\boldsymbol{d}$ by minimizing the l_1-norm of (4), i.e.,

$$\text{minimize} \quad \|\mathbf{C}\cdot\boldsymbol{d}\|_1 , \qquad (5)$$

subject to the conditions

$$B\cdot\mathbf{B}_{(J\times N)}[\boldsymbol{d}_1, \boldsymbol{d}_2, \ldots, \boldsymbol{d}_N]^T_{(J\times N)} \preceq \mathbf{B}_{0(J\times J)}$$
$$P\cdot\boldsymbol{1}_{(1\times N)}[\boldsymbol{d}_1, \boldsymbol{d}_2, \ldots, \boldsymbol{d}_N]^T_{(J\times N)} \preceq \boldsymbol{p}_{0(1\times J)}$$
$$\max\{\boldsymbol{J}_{(K\times NJ)}\mathrm{d}_{(NJ\times 1)}\} \leq \Delta_J$$
$$\max\{\boldsymbol{D}_{(K\times NJ)}\mathrm{d}_{(NJ\times 1)}\} \leq \Delta_D ,$$

where $\mathbf{B}_{j,n} \in \{0,1\}$ indicates the presence of a n^{th} social group user in the j^{th} cloud, B indicates the multicast bandwidth per group and $\mathbf{B}_0$ indicates the available bandwidth in virtual link set T. Similarly P indicates the processing power requirement per social group, and $\boldsymbol{p}_0$ represents the available processing power in each cloud. $\mathbf{D}$ and $\mathbf{J}$ possess a similar structure to $\mathbf{C}$ and identify the end-to-end delay and jitter of k^{th} user in the system, whereas Δ_D and Δ_J represent the maximum acceptable delay and jitter, respectively, for the interactive application.

III. Results and Discussion

We evaluate the proposed Simultaneous Group QoS Maximization (SGQM) algorithm for a variety of network and social group configurations, and compare the results with Sequential Individual QoS Maximization (SIQM) and Random Resource Allocation (RRA) (also sometimes known as the 'Best Fit' and 'First fit' methods respectively). The performance of each method is evaluated using the mean of the group cost in (4), obtained from 50 Monte Carlo simulations of different network conditions and social group configurations. The simulation environment consists of 10 potential processing clouds (J=10) and simulate random networking conditions, where the matrix elements $\mathbf{B}_{0,ij} \in (20,60)$ Mbps, $\mathbf{D}_{ij} \in (20,120)$ ms, $\mathbf{J}_{ij} \in (20,60)$ ms, $\boldsymbol{p}_{0,ij} \in (10, 85)$, L_n=2 and $M_n \in (12,20)$. Based on [2, 4], the minimum requirements for HDTV transmissions are obtained as B=8 Mbps, P=15, Δ_D=100 ms and Δ_J =50 ms. The MOSEK [5] and YALMIP [6] toolboxes for MATLAB are used to estimate the optimal solution for each scenario.

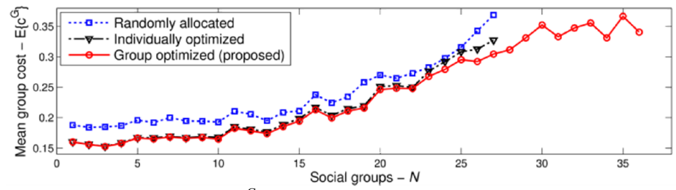

Fig. 2. Mean group cost $E\{\boldsymbol{c}^G\}$ with respect to the number of social groups N.

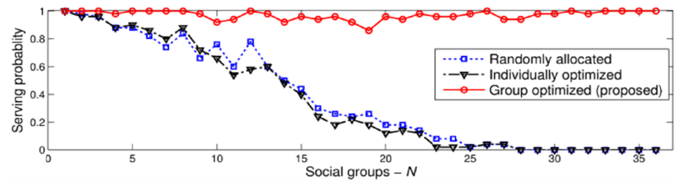

Fig. 3. Probability of serving all social groups within the ITV system.

Fig. 2 illustrates the mean group cost with respect to the number of social groups. The proposed SGQM method exhibits the lowest mean group cost and follows a rising trend with N, as is expected due to the increased complexity of the group optimization problem. In addition, Fig. 2 indicates the capability of the proposed method to optimally allocate resources for N>27, unlike the SIQM and RRA approaches. This is mirrored in Fig. 3, where the likelihood of successfully allocating resources for all user groups, i.e. the serving probability, is illustrated. While the probability of serving all user groups decreases with N for SIQM and RRA, a serving probability greater than 90% is achieved by the proposed SGQM method, illustrating the efficiency of group based resource allocation.

IV. Conclusion

In this paper, we present a simultaneous group QoS maximization technique to determine the optimal cloud resource allocation for ITV applications in delay, bandwidth and processing constrained networks. The efficient and superior utilization of resources is demonstrated over the individual and random optimization mechanisms. Future work will focus on integrating the optimal routing problem for application of these methods in real-world ITV systems.

V. References

[1] J. Chung, K. Xu, A. Colaco, C. Schmandt and V.O.K. Li, "My second bike: A TV-enabled social and interactive riding experience" *IEEE Consumer Commun. and Networking Conf.,* Las Vegas, 2010, pp. 1 – 5.

[2] ACTION-TV, [Online]. Available: *http://www.action-tv.net*

[3] R. Buyya, R. Ranjan and R.N. Calheiros, "Inter-cloud:Utility-Oriented Federation of Cloud Computing Environments for Scaling of Application Services", *Proceedings of the 10th Int. Conf. on Algorithms and Architectures for Parallel Processing,* Busan,Korea,2010,pp. 13-31.

[4] H.J. Kim and K.S. Choi, "A study on a QoS/QoE correlation model for QoE evaluation on IPTV service," *12th International Conf. on Advanced Commun. Technology,* Phoenix Park, Ireland, 2010, pp. 1377 – 1382.

[5] The MOSEK optimization toolbox for MATLAB [Online]. Available: http://www.mosek.com

[6] YALMIP: A Toolbox for Modeling and Optimization in MATLAB [Online]. Available: http://users.isy.liu.se/johanl/yalmip