Contract-Theoretic Modeling for Content Delivery in Relay-Based Publish-Subscribe Networks

Yifan Li, Ping Wang, Dusit Niyato, and Yang Zhang
Nanyang Technological University, Singapore 639798

Abstract—Mobile social network (MSN) has been widely studied as a novel and effective way for communication in the current era of fast developing mobile devices, which enables the contents and services to be delivered opportunistically when mobile users contact. In this paper, we aim to optimize the content delivery from the content provider (CP) to the subscribers via relays in between. Tandem queuing model is applied to model the content delivery process, based on which absorbing Markov chain is used to derive the quality of the content delivery service received by the subscribers. The validity of the proposed queuing model has been verified by our simulation results. Observing that the content provider always has dominant control on the content delivery process and is pursuing the maximum profit by strategically designing the “rights” and “obligation” items for the subscribers, contract theory is adopted to reach an economically optimal solution. The numerical results verify the effectiveness of the contract-theoretic approach in maximizing the content provider’s profit, and the capability to ensure the satisfaction of the heterogeneous subscribers with different quality of service (QoS) requirements.

I. INTRODUCTION

Recently, a mobile social network (MSN) has been widely studied in the community, and applied to enable various communication applications amongst people with mobile devices. An important issue in MSNs is the content distribution problem. Generally, there exists one or more content sources, i.e., the content providers (CPs), with the purpose of gaining profits by delivering the contents to the subscribers using communication facilities such as base stations or access points. Due to the transmission reachability of a content provider, the content data can be relayed if the subscribers are not within the communication range.

Note that such a relay-based publish-subscribe (pub-sub) infrastructure is intrinsically an asymmetric information system. The content provider plays a dominant role in the content delivery process. Being in such a situation, the content provider is capable of designing “rights” and “obligation” items for the subscribers with egoistic motivation. Here the “rights” can be considered as being provided with the content delivery services, while the “obligation” can be the payment corresponding to the received service quality. As a promising technique for such an asymmetric scenario, contract theory is a suitable tool to tackle the content delivery problem, which motivates us to model the content provider to make a “take-it-or-leave-it” offer to the subscribers. The main contributions of this paper can be summarized as follows: (1) To analyze the content delivery performance of a relay-based publish-subscribe network, we construct a tandem queuing model to theoretically analyze the performance of the content delivery process; (2) Taking the heterogeneity of the subscribers into account, we propose an optimal contract design scheme for the content provider, to maximize the content provider’s profit while at the same time the subscribers are satisfied with the contract items designed for their particular types. To the best of our knowledge, this is the first work to apply contract theory to the scenario of the relay-based pub-sub networks.

The rest of the paper is organized as follows. Section II reviews the related work. Section III describes the system model of the relay-based publish-subscribe mobile social network, as well as the abstract queueing model of the network. The design and optimization of contract for content distribution in the system is provided in Section IV, where the self-interested and rational contracting strategies of the content provider and subscribers are discussed. In Section V, we present the performance evaluation and numerical results. Section VI concludes the paper.

II. RELATED WORK

A. Publish-Subscribe System

Publish-subscribe (pub-sub) is a typical paradigm for content dissemination in a distributed system with autonomous users. As described in [1], a pub-sub system is driven by the events that the subscribers notify the subscription information to the publishers, and the publishers efficiently manage the subscriptions and make deliveries. The pub-sub system adopted in a mobile environment has been studied in [2]. In [3], an efficient multicast algorithm for content-based pub-sub systems has been developed. A tree structure is established to describe the connections among publishers, content brokers and subscribers. However, in [3], the issues of mobile systems, e.g., contact rate and bandwidth, have not been discussed.

B. Contract Theory

Contract theory [4] studies how the two rational and self-interested entities in a system reach agreements based on an economic incentive. Recently, contract-theoretic approaches are employed to analyze resource distribution in wireless systems. For example, in [5], the contract theory model is applied to optimally solve the spectrum resource allocation in cognitive radio systems. In [6], the authors employ contract to optimize the cooperative spectrum transmission problem. In [7], contract theory is used for the client (Google, as the example in [7]) to offload portions of computing tasks to the smartphone users (as contract buyers). In [8], the authors discuss the application of contract theory to allocating bandwidth in a geo-location information brokerage system.

III. SYSTEM MODEL

Consider the scenario that a content provider (CP) needs to deliver contents to users scattered around it, including users nearby and users far away. To achieve this goal, a relay-based publish-subscribe model is adopted.

As shown in Fig. 1, we consider a relay-based publish-subscribe network, which consists of the content provider,
mobile users (subscribers) and relays. The distance between the content provider and the subscribers could be large that multiple relays are required to relay contents. In this case, the relays can be marked by different tiers according to the sequence they receive the content. For instance, the relays directly contacting with the content provider are defined as Tier-1 relays, and the relays which relay for the Tier-1 relays are Tier-2 relays, so on and so forth. The users directly connected with the content provider are defined as the CP users, while the users connected with Tier-n relays are defined as Tier-n users. The content source, the intermediate nodes, and the destination comprise a content delivery path. Since each content has a life cycle, indicating that if the content delay exceeds some deadline, it will be useless. We use the delivery ratio within deadline (DRWD) to measure the quality of the content delivery service, which is defined as:

\[ \eta = P(t \leq T). \]  

(1)

where \( T \) is the deadline of the content, and \( P(t \leq T) \) denotes the probability that the content can be delivered to the user within deadline \( T \). In this system model, the bandwidth between a transmitter and receiver is allocated in such a way that with a unit bandwidth, one content can be transmitted through one contact between the transmitter and the receiver. Note that given the contact rates, in order to provide service quality \( \eta \) to the subscriber, the content provider has to guarantee a certain bandwidth provision to the subscriber. Due to the cost introduced by bandwidth provision, the content provider has to charge the subscribers for requiring a certain service quality \( \eta \).

IV. QUEUING MODEL ANALYSIS

Considering the content arrival and departure at each node, the content delivery along a certain path can be modeled as a tandem of queues. For the simplicity of presentation, we investigate the tandem queueing model for the content delivery from the content provider to the Tier-1 mobile users. The model can be extended to content delivery paths from the content provider to Tier-n mobile users straightforwardly.

Let \( Q_1 \) denote the queue of contents at the content provider, while \( Q_2 \) denotes the queue of contents at the Tier-1 relay \( R_1 \). We take a typical content delivery path \( Q_1 - Q_2 \) as an instance. The contents are injected into the queue at the content provider. It is assumed that the content generation is a Poisson process with average rate \( \lambda \) [9]. Note that the departure of contents from \( Q_1 \) occurs only when the Tier-1 relay \( R_1 \) contacts with the content provider. Generally, the time period between two consecutive meetings of the Tier-1 relay and the content provider is assumed to be exponentially distributed [9]. In the example, let \( \mu_1 \) denote the contact rate between the Tier-1 relay \( R_1 \) and the content provider, then the time period between two consecutive contacts is an exponentially distributed variable with mean \( 1/\mu_1 \). Therefore, \( R_1 \) can be considered as the server of \( Q_1 \) with service rate \( \mu_1 \) if given a unit bandwidth. Similarly, subscriber \( S_1 \) can be considered as the server of \( Q_2 \) with service rate \( \mu_2 \) if given a unit bandwidth, where \( \mu_2 \) denotes the contact rate between subscriber \( S_1 \) and Tier-1 relay \( R_1 \).

In the example scenario, there are two cases of the content delivery service: (1) content delivery to subscribers directly connected to the content provider (referred to as Case I), and (2) content delivery to the Tier-1 subscribers (referred to as Case II). Due to the page limit, we only discuss Case I in later part of this paper. Note that Case II can be analyzed in the similar way.

**Case I:** For the content delivery service from the content provider to the subscribers directly connected to it, the state transition of the corresponding Markovian queue with bulk services is shown in Fig. 2. In this figure, \( r_1 \) is the bulk size, indicating that \( r_1 \) units of bandwidth is allocated by the content provider. In other words, the subscriber can obtain up to \( r_1 \) contents through one contact with the content provider.

Let \( \pi(k) \) denote the stationary probability that the queue contains \( k \) contents in its buffer space. According to the state-transition-rate diagram shown in Fig. 2, the balancing equations can be obtained as follows:

\[
\begin{align*}
\pi(0) \cdot \lambda &= \sum_{k=1}^{r_1} \pi(k) \cdot \mu_1, \\
\pi(k) \cdot (\lambda + \mu_1) &= \pi(k-1) \cdot \lambda + \pi(k+r_1) \cdot \mu_1, \\
(1 \leq k \leq K_1 - r_1) \\
\pi(k) \cdot (\lambda + \mu_1) &= \pi(k-1) \cdot \lambda, \\
(K_1 - r_1 < k < K_1) \\
\pi(K_1) \cdot \mu_1 &= \pi(K_1 - 1) \cdot \lambda.
\end{align*}
\]

By introducing the normalizing equation \( \sum_{k=0}^{K_1} \pi(k) = 1 \), the above balancing equations can be solved. Consequently, the stationary probabilities \( \pi(k) \) \((0 \leq k \leq K_1)\) can be obtained.

Next we derive the time required for a particular content to be transmitted to a subscriber, i.e., the content delay. A content has to wait until all contents ahead of it have been transmitted, i.e., the number of contents ahead of it becomes zero. We model an absorbing Markov chain with 0 (denoting the situation that there is no content ahead of the concerned one) being the absorbing state and other states (state \( k(k > 0) \) denotes that there are \( k \) contents ahead of the tagged one) being the transient states. The content delay is the absorption time of the considered absorbing Markov chain. Take \( r_1 = 2 \) as an example, then based on the proposed queueing model, the transition rate matrix describing the transition rates among all states (including the absorbing state and the transient
the transition rates among all the transient states (absorbing state). Thus, the transition rate matrix describing probabilities obtained by removing the first row and the first column of states) is shown in the following:

\[
\mathbf{\tilde{S}} = \begin{bmatrix}
1 & 0 & 0 & \cdots & 0 \\
\mu_1 & -\mu_1 & 0 & \cdots & 0 \\
\mu_1 & 0 & -\mu_1 & \cdots & 0 \\
0 & \mu_1 & 0 & \cdots & 0 \\
\cdots & \cdots & \cdots & \cdots & \cdots \\
0 & \cdots & \cdots & \mu_1 & 0 & -\mu_1
\end{bmatrix}
\]  

(2)

where the element in the \(i^{th}\) row, \(j^{th}\) column in matrix \(\mathbf{\tilde{S}}\) denotes the transition rate from state \(i\) (\(i = 1\) corresponds to the absorbing state) to state \(j\) (\(j = 1\) corresponds to the absorbing state). Thus, the transition rate matrix describing the transition rates among all the transient states \(\mathbf{S}\) can be obtained by removing the first row and the first column of \(\mathbf{\tilde{S}}\).

Let \(\alpha\) denote the initial probabilities of starting in any state (i.e., phase), which can be obtained from the stationary probabilities \(\pi(k)\) as mentioned above. Let \(t\) denote the time required to reach the absorbing state (i.e., the content delay), then \(t\) follows a phase-type distribution denoted by \(PH(\alpha, \mathbf{S})\). The cumulative distribution function (CDF) of \(t\) (\(t > 0\)) is given by [10]

\[
F(t) = 1 - \alpha \exp(\mathbf{S}t)1,
\]

(3)

from which the DRWD can be obtained. Note that the derived DRWD in this section will be used in the contract design in the following section, as a metric to quantify the quality of the content delivery service.

V. OPTIMAL CONTRACT DESIGN FOR CONTENT DISTRIBUTION

A. CONTENT PROVIDER SIDE

Considering the content delivery service as a kind of “commodity”, then the relay-based publish-subscribe network can be modeled as a trading market with the content provider being the seller and the subscribers being the consumers (i.e., buyers). The content provider sets the service qualities it can provide with, and the corresponding prices for that service. Intuitively, higher price has to be paid for higher service quality offered. Notified by the quality and price information, the subscribers can optimally decide which qualities to purchase.

As in Section IV, only Case I is considered in the proposed model. In Case I, note that different subscribers can have different contact rates with the content provider. Let \(\mathcal{G}_{CP}\) denote the set of subscribers connected to the content provider directly, which can be divided into different groups according to different contact rates with the content provider. Let \(\mathcal{G}_{CP} = \{\mathcal{G}_{i,CP}\} (i = 1, \ldots, I)\) be the set of subscriber groups in Case I, where \(I\) is the number of such groups. The subscribers in the same group \(\mathcal{G}_{i,CP}\) are considered to have the same contact rate (denoted by \(\rho_i\)) with the content provider.

Considering one single subscriber group \(\mathcal{G}\) (\(\mathcal{G} = \mathcal{G}_{i,CP}\)), let \(\Gamma\) and \(\Pi\) denote the set of all service qualities and the set of all prices, respectively. Within group \(\mathcal{G}\), each quality of service \(\eta \in \Gamma\) corresponds to a price \(p \in \Pi\). As the content delivery along different paths is independent, the content provider can solve the utility maximization problem separately, for each content delivery path. The utility of the content provider (obtained by selling a service quality \(\eta\) to a subscriber in group \(\mathcal{G}\) (\(\mathcal{G} = \mathcal{G}_{i,CP}\)) is defined as the difference between the monetary gain collected from the subscriber in \(\mathcal{G}\) and the cost of providing service quality \(\eta\) (e.g., due to bandwidth allocation), which is denoted by:

\[
U^\mathcal{G}_{CP}(\eta) = p^\mathcal{G}(\eta) - C^\mathcal{G}(\eta)
\]

(4)

Note that from the queueing analysis in Section IV, there is a one-to-one mapping between the bandwidth allocation \(W\) (which determines the bulk size in the queueing model) and the service quality \(\eta\). Thus, \(\eta\) can be considered as a function of \(W\), and \(C^\mathcal{G}(\eta(W))\) can be written as \(C^\mathcal{G}(W)\) for simplicity. Similar as [5], in this work, the cost function is defined as:

\[
C^\mathcal{G}(W) = \delta \cdot W^\beta
\]

(5)

where \(\delta\) is a weighting factor, \(0 < \beta < 1\) is a parameter controlling the behavior of the cost function. Thus, the cost function for Case I can be obtained from:

\[
C^\mathcal{G} = \delta_1 \cdot (W^\mathcal{G}_{CP})^\beta_1, \quad \mathcal{G} = \mathcal{G}_{i,CP},
\]

(6)

where \(W^\mathcal{G}_{CP}\) is the bandwidth required to provide service quality \(\eta\) to the subscriber in group \(\mathcal{G}_{i,CP}\) by the content provider.

The aim of the content provider is to maximize its utility by strategically designing the service quality it can provide, as well as the corresponding cost. As a rational content provider, negative utility from providing a service can never be accepted, and thus it will always set \(p^\mathcal{G}(\eta) \geq C^\mathcal{G}(\eta)\), \(\forall \eta \in \Gamma\) for the subscriber in group \(\mathcal{G}\).

B. SUBSCRIBER SIDE

In practice, different subscribers may have different requirements for the content delivery quality, thus the valuation of the same service quality \(\eta\) can be different for different subscribers. The valuation of service quality \(\eta\) to subscriber \(S_i\) is defined as:

\[
V_i(\eta) = \kappa \cdot \log(1 + \vartheta_i \cdot \eta)
\]

(7)

where \(\kappa > 0\) is a weighting factor, while \(\vartheta_i > 0\) can be described as the time-sensitivity of subscriber \(S_i\). A higher \(\vartheta_i\) refers to a higher requirement for the service quality, i.e., less tolerance on content delivery delay. According to different preference for the quality of the content delivery service, the
subscribers in group \( G \) are classified into different types. We refer to a subscriber as a type-\( \theta \) subscriber if \( \theta_i = \theta \).

Considering subscriber group \( G_{CP} \), we denote the set of all subscriber types as \( \Theta_{CP} \). According to Eq. (7), for a service quality \( \eta \), the valuation of a subscriber with type \( \theta \) is calculated by

\[
V_S(\theta, \eta) = \kappa \cdot \log(1 + \theta \cdot \eta), \quad \theta \in \Theta_{CP}.
\] (8)

We define the utility of a subscriber with type \( \theta \) in group \( G \) by purchasing a service quality \( \eta \) as follows:

\[
U_{\theta}^{G}(\theta, \eta) = V_S(\theta, \eta) - p^G(\eta), \quad \forall \theta \in G_{CP}.
\] (9)

### C. Optimal Contract Design

According to Eqs. (4) and (9), it can be observed that the content provider wants to sell the service at a higher price to maximize its utility, while the subscriber prefers to purchase the required service quality with lower price, resulting in conflict objectives between them. To resolve such a problem, a contract-theoretic approach is applied in the following. Specifically, the problem for the content provider is to choose a set of service qualities in terms of DRWD and a price for each service quality for subscribers in each group. Considering a specific group \( G \) (\( G = G_{CP} \)), since only one service quality will be chosen by each type of subscribers, the content provider will assign a service quality \( \eta^G(\theta) \) and a corresponding price \( p^G(\eta^G(\theta)) \) to each subscriber type \( \theta \in \Theta_G \). For simplicity, we write \( p^G(\eta^G(\theta)) \) as \( p^G(\eta^G) \), since \( \eta^G(\theta) \) is a single value function. Therefore, it is enough to design a contract consisting of \( M \) contract items, where \( M \) is the number of the subscriber types. The contract designed for the considered problem is defined as follows:

**Definition 1. CP-G Contract:** The CP-G contract \( C_{CP-G} \) proposed by the content provider to group \( G \) (\( G = G_{CP} \)) of subscribers has the form of \( CP-G = (\eta^G(\theta), p^G(\theta)) \), \( \forall \theta \in \Theta_G \), where \( \eta^G(\theta) \) is the content delivery service quality (in terms of DRWD) provided by the content provider to the subscribers with type \( \theta \) in group \( G \), and \( p^G(\theta) \) is the price required to be paid by the content provider by the subscribers with type \( \theta \) in group \( G \), in order to get service quality \( \eta^G(\theta) \).

However, for a feasible CP-G contract, two properties have to be satisfied, namely the incentive compatibility (IC) and the individual rationality (IR) [4]. In the considered problem, IC property indicates that each subscriber prefers to buy the content delivery service aligned to his type, i.e.,

\[
V_S(\theta, \eta^G) - p^G(\eta^G) \geq V_S(\theta, \eta^G(\bar{\theta})) - p^G(\bar{\theta}), \quad \forall \theta \neq \bar{\theta},
\] (10)

for all type \( \theta \in \Theta_G \). IR means that a subscriber will choose the contract item designed for his type only when this choice can provide him with a non-negative utility, i.e.,

\[
V_S(\theta, \eta^G(\theta)) - p^G(\theta) \geq 0, \quad \forall \theta \in \Theta_G.
\] (11)

Note that for a feasible contract \( C_{CP-G} \), the overall utility of the content provider obtained from selling content delivery services to subscribers in group \( G \) is calculated by:

\[
U_{CP}^G = \sum_{\theta \in \Theta_G} N^G_\theta \left[ p^G(\theta) - C^G(\eta^G(\theta)) \right], \quad \forall \theta \in \Theta_G,
\] (12)

where \( N^G_\theta \) is the number of type-\( \theta \) subscribers in group \( G \). The optimal contract for the content provider, denoted by \( C_{CP-G}^{G,*} \), is a feasible contract maximizing its overall utility.

In the proposed model, we consider a finite number of subscriber types, denoted by \( \theta_1, \theta_2, \ldots, \theta_M \). Without loss of generality, it is assumed that \( \theta_1 < \theta_2 \cdots < \theta_M \). Then, Eq. (12) can be written as:

\[
U_{CP}^G = \sum_{m=1}^{M} N^G_{\theta_m} \left[ p^G(\theta_m) - C^G(\eta^G(\theta_m)) \right].
\] (13)

Note that with the purpose of maximizing \( U_{CP}^G \), both variables \( \eta^G(\theta_m) \) and \( p^G(\theta_m) \) have to be decided simultaneously, for \( m = 1, \ldots, M \). We define a quality assignment \( \eta^G = \{\eta^G(\theta_1), \eta^G(\theta_2), \ldots, \eta^G(\theta_M)\} \) as a feasible quality assignment, if the constraints \( \eta^G(\theta_1) \leq \eta^G(\theta_2) \leq \cdots \leq \eta^G(\theta_M) \) hold. To design an optimal contract for the considered problem, similar method in [5] (i.e., sequential optimization technique) can be applied. Specifically, in the first step, we derive the best prices under a fixed feasible service quality assignment, while for the second step, the best service quality assignment is derived.

Let \( U_{CP}^{G,*}(\eta^G) \) denote the maximum utility the content provider can achieve from selling services to subscriber group \( G \) under a given feasible quality assignment \( \{\eta^G\} \), which can be obtained from:

\[
U_{CP}^{G,*}(\eta^G) = \max_{p^G(\theta)} \sum_{m=1}^{M} N^G_{\theta_m} \left[ p^G(\theta_m) - C^G(\eta^G(\theta_m)) \right]
\] (14)

subject to the price constraints in the following:

- \( 0 \leq p^G(\theta_1) \leq V_S(\theta_1, \eta^G(\theta_1)) \)
- \( p^G(\theta_m) \geq p^G(\theta_{m-1}) + V_S(\theta_{m-1}, \eta^G(\theta_{m-1})) - V_S(\theta_{m-1}, \eta^G(\theta_m)), \quad 2 \leq m \leq M \)
- \( p^G(\theta_m) \leq p^G(\theta_{m-1}) + V_S(\theta_{m-1}, \eta^G(\theta_m)) - V_S(\theta_m, \eta^G(\theta_{m-1})), \quad 2 \leq m \leq M \)

The price constraints are sufficient and necessary conditions for the contract to be feasible, which can be proved by mathematical induction as in [5]. For the problem defined in Eq. (14), the optimal price assignment with given service quality assignment, which is denoted by \( p^{G,*} = \{p^{G,*}(\theta_1), p^{G,*}(\theta_2), \ldots, p^{G,*}(\theta_M)\} \), can be obtained from:

\[
p^{G,*}(\theta_1) = V_S(\theta_1, \eta^G(\theta_1))
\] (15)
\[
p^{G,*}(\theta_m) = \max \left( p^{G,*}(\theta_{m-1}) + V_S(\theta_{m-1}, \eta^G(\theta_m)) \right.
- \left. V_S(\theta_m, \eta^G(\theta_{m-1})) \right), \quad 2 \leq m \leq M
\] (16)

Note that the optimality of the price assignment and the uniqueness can be proved similarly as in [5]. For simplicity of denotations, let \( \Delta_1 = 0 \), and \( \Delta_m = V_S(\theta_m, \eta^G(\theta_m)) - V_S(\theta_m, \eta^G(\theta_{m-1})) \), \( \forall m = 2, \ldots, M \). Note that \( \Delta_m \) represents the gap between the valuation that can be achieved by a \( \theta_m \) subscriber from choosing service quality designed for type-\( \theta_m \) subscribers, and that can be achieved from choosing service quality designed for the subscribers with the nearest lower type, i.e., type \( \theta_{m-1} \). Then, the optimal prices denoted by Eq. (16) can be written as:

\[
p^{G,*}(\theta_m) = V_S(\theta_1, \eta^G(\theta_1)) + \sum_{i=1}^{m} \Delta_i, \quad m = 1, \ldots, M
\] (17)

By substituting the optimal price assignment \( p^{G,*}(\theta_m) \) under fixed service quality assignment into Eq. (14), we have
\[
U_{CP}^{G}(\eta_{G}) = \sum_{m=1}^{M} N_{\eta_{m}}^{G} \left[ V_{S}(\theta_{m}, \eta_{m}^{G}(\theta_{m})) + \sum_{i=1}^{m-1} (V_{S}(\theta_{i}, \eta_{G}^{\ast}(\theta_{i})) - V_{S}(\theta_{i+1}, \eta_{G}^{\ast}(\theta_{i}))) - C_{G}^{G}(\eta_{m}(\theta_{m})) \right]
\]

Compared with Eq. (13) containing two variables, i.e., the service quality assignment and the price assignment, there is only one variable, i.e., the service quality assignment, in Eq. (18). The optimal service quality assignment can be denoted as
\[
\eta_{G}^{\ast} = \arg \max_{\eta_{G}} U_{CP}^{G}(\eta_{G}), \quad (18)
\]
which can be derived as follows. Considering that the subscribers enter group \(G\) in sequence, from the lowest type \(\theta_{1}\) to the highest type \(\theta_{M}\). Thus the first element of the optimal service quality assignment, i.e., \(\eta_{G}^{\ast}(\theta_{1})\), can be obtained from \(\eta_{G}^{\ast}(\theta_{1}) = \arg \max_{\eta_{G}} N_{\eta_{1}}^{G} (V_{S}(\theta_{1}, \eta_{G}^{\ast}(\theta_{1})) - C_{G}^{G}(\eta_{G}^{\ast}(\theta_{1})))\). Based on \(\eta_{G}^{\ast}(\theta_{1})\), when the subscribers of type \(\theta_{2}\) enter group \(G\), the second element of the optimal service quality assignment can be obtained from
\[
\eta_{G}^{\ast}(\theta_{2}) = \arg \max_{\eta_{G}^{\ast}(\theta_{2})} \left[ N_{\eta_{1}}^{G} (V_{S}(\theta_{1}, \eta_{G}^{\ast}(\theta_{1})) - C_{G}^{G}(\eta_{G}^{\ast}(\theta_{1}))) + N_{\eta_{2}}^{G} (V_{S}(\theta_{2}, \eta_{G}^{\ast}(\theta_{2})) - C_{G}^{G}(\eta_{G}^{\ast}(\theta_{2}))) \right]
\]
where \(\Phi(\theta_{2}, \eta_{G}^{\ast}(\theta_{2})) = V_{S}(\theta_{2}, \eta_{G}^{\ast}(\theta_{2})) + V_{S}(\theta_{1}, \eta_{G}^{\ast}(\theta_{1})) - V_{S}(\theta_{2}, \eta_{G}^{\ast}(\theta_{1})) - C_{G}^{G}(\eta_{G}^{\ast}(\theta_{2}))\). By applying this scheme, the \(m^{th}\) element (i.e., \(\eta_{G}^{\ast}(\theta_{m})\)) of the service quality assignment can be obtained based on all elements determined before, i.e., from the first element to the \((m-1)^{th}\) element. Finally, the optimal service quality assignment \(\eta_{G}^{\ast}\) can be obtained. Combined with the corresponding price assignment given by Eq. (17), all contract items of the CP-G contract for the considered problem are optimally designed.

**VI. PERFORMANCE EVALUATION**

**A. Performance Metrics Obtained from Queueing Model**

In this section, we present the analytical results as well as simulation results of the performance metrics obtained from the proposed queueing model, as shown in Fig. 3. TheCDF of content delay with different queue service rate (i.e., \(\mu\)) is shown in Fig. 3(a). It can be observed that as \(\mu\) increases, the CDF curve of content delay converges with a higher speed. The CDF of content delay is shown for different bandwidth allocation at the content provider (i.e., \(r_{1}\)) is shown in Fig. 3(b). In order to verify that validity of the analytical results, we have done simulation of the queueing model using MATLAB. Some simulation results are shown as examples in Figs. 3(c)-(f), it can be found that the simulation results match the analytical results very well (the number of content arrivals is set to be 10000 in our simulation). Note that the queue service rate is dependent on the contact rate between the subscriber and the content provider, it can be observed from Fig. 3(e) that as \(\mu\) increases, the DRWD keeps increasing, which accords with the expectation that when the subscriber has a frequent contact with the content provider, it can always get the latest content more easily, resulting in a higher DRWD.

The variation of the DRWD with \(r_{1}\) is shown in Fig. 3(f). It can be observed that the DRWD increases as \(r_{1}\) increases, however, the increasing speed becomes obviously slower when \(r_{1} \geq 3\), indicating that increasing bandwidth allocation will not always bring apparent improvement of the content delivery service. This observation can also be verified by the more and more indistinguishable difference in the CDF of content delay as \(r_{1}\) increases, as shown in Fig. 3(b). Therefore, when the improvement of the service quality cannot compensate the cost has to be paid, the content provider has no motivation to increase the bandwidth allocation.

**B. Optimal Contract Design for the Content Provider**

In this section, we implement the proposed contract \(C_{CP-G}\) for the content provider in Case I, illustrating that the contract design is optimal from the perspective of both the content provider and the subscribers. The content arrival rate \(\lambda\) is 0.5 contents per hour, and the life cycle time of the content is set to be 1.5 hour. The parameters of the cost function shown in Eq. (6) is set to be: \(\delta_{1} = 0.02, \delta_{2} = 0.02\).

In Case I, we consider four subscriber groups with contract rate with the content provider to be 0.6, 0.8, 1 and 1.2 contact per hour, respectively, which are denoted by group \(G_{0.6}\), group \(G_{0.8}\), group \(G_{1}\) and group \(G_{1.2}\), as shown in Table I, in each of which there are 20 subscribers. In each subscriber group, we consider discrete types \(\theta_{1} = 1\) and \(\theta_{2} = 10\), indicating that some subscribers have lower requirements for service quality (type \(\theta_{1}\)), while others have higher requirements for service quality (type \(\theta_{2}\)). The number of \(m_{1}\) \((m = 1, 2)\) subscribers is 10. From Table I, we have two observations as shown in the following:

- Within one subscriber group, as type \(\theta\) increases, the service quality designed for type-\(\theta\) subscribers will increase, as well as the corresponding price they have to pay.
- Comparing between different subscriber groups, the subscribers with higher contact rate with the content provider will receive higher service qualities than those in the other groups.

Note that the content provider will always select the service quality maximizing its utility, and the IC constraints are implicitly satisfied in the contract design process, therefore, subscribers with type \(\theta\) will benefit most by selecting the contract item designed for type \(\theta\). As a result, we can verify that the proposed contract design is optimal, with the benefits of both the content provider and the subscribers taken into account.

**VII. CONCLUSION**

In this paper, the content distribution problem as a critical issue in relay-based publish-subscribe networks are discussed. As the subscribers are heterogeneous, they may have different requirements on QoS of content delivery. We have developed the absorbing Markov chain model to derive the delivery ratio within deadline (DRWD), based on which we introduced a contract-theoretic approach to optimize the content distribution strategy to maximize the benefit of content provider while ensuring the satisfaction of mobile users. The simulation results verify our analytical results and have shown that the proposed contract-theoretic approach is capable of maximizing the profit of the content.


