A Dynamic Reservation Scheme in Online Charging System for Family Shared Plan

Huai-Sheng Huang, Tzu-Chung Su, Jia-Ming Liang and Yu-Chee Tseng
Department of Computer Science, National Chiao Tung University, Hsinchu, Taiwan
E-mail: {huanghs, tcsu0224, jmliang, yctseng}@cs.nctu.edu.tw

***Abstract*—For simplifying billing and limiting data usage, *Family Shared Plan* (*FSP*) is now a popular plan provided by telecommunications operators. Based on the *3rd Generation Partnership Project (3GPP)* specifications, *Online Charging System* (*OCS*) has to determine the *Granted Unit* (*GU*) and create reservations per session before a service is delivered by the serving network to a *User Equipment* (*UE*). However, fixed GUs cannot adapt to dynamic consumer behaviors, which usually require dynamic GUs. Clearly, giving small GUs will incur too many signaling messages for making reservations, but giving excessive GUs will cause unbalanced distribution of resources, especially in FSP cases. How to determine GUs is an open issue that can be self-defined by telecommunications operators. In this paper, we propose a new scheme for dynamically assigning GUs to UEs belonging to the same FSP, based on their historical data usage and total monthly data allowance. Simulation results show that our scheme can substantially save signalings by at least 22% compared to fixed scheme under unpredictable behaviors.**

***Index Terms*—4G, Long Term Evolution (LTE), Online Charging System (OCS), 3GPP, family shared plan.**

I. Introduction

The *Diameter* signaling protocol [1] has played an important role in the mobility world in the past few years. According to [2], the growth in signaling traffic correlates directly to the pricing plans in *Long Term Evolution* (*LTE*) networks. In Fig. 1, the global LTE Diameter signaling traffic is categorized according to policy, *Online Charging System* (*OCS*), mobility and *Offline Charging System* (*OFCS*). As can be seen, the OCS signalings are doubled every year and are estimated to account for 25% of signaling volumes in 2017. Obviously, OCS will become the fastest growing Diameter use case, with a global *Compound Annual Growth Rate* (*CAGR*) of 180%.

Since Verizon launched its first *Family Shared Plan* (*FSP*) in June 2012 [3], FSP has become a popular telecommunications tariff plan. Subscribers belonging to the same FSP can use their own devices to surf the Internet and share the same quota with different consumer behaviors. For example, a family of three persons may use the same FSP, where the father may use the Internet mainly for work, the mother may use the Internet mainly for audio/video communications, and the teenager may use the Internet mainly for online games and social media. As more devices are attached to a FSP, it means that the monthly data allowance may be consumed by different patterns. In fact, operators usually set different policies to deal with the monthly data allowance over usage issue [4]. China Mobile (Hong Kong) will disconnect the subscribers from the network to prevent bill shock, Singtel (Singapore) will charge on a per gigabytes (GB) basis, while NTT docomo (Japan) will limit the subscribers to access some services such as video streaming after over quota. While these rules only apply to the single-subscriber case, this work looks at the quota reservation issue under the FSP case.

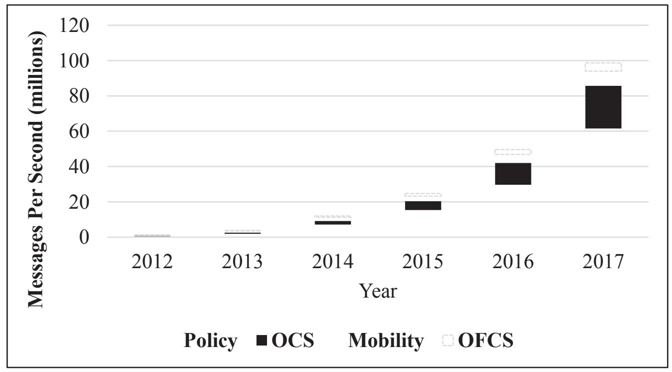

Fig. 1. Global LTE Diameter signaling volumes [2].

Based on the *3rd Generation Partnership Project* (*3GPP*) specification, Fig. 2 shows the OCS architecture for *IP Multimedia Subsystem* (*IMS*) services [5]. When a subscribers enables an application (such as Youtube or Line), the *IMS Application Server* communicates with the *Session Based Charging Function* (*SBCF*) using the Diameter protocol. The SBCF performs session-based charging and credit control by communicating with the *Rating Function* (*RF*) in order to determine the value of the requested resources of the session. On the other hand, the SBCF communicates with the *Account Balance Management Function* (*ABMF*) to query and update the subscriber's account and counter status. When the prepaid user's credit depletes, the ABMF connects the Recharge Server to trigger the recharge account function. The *Charging Gateway Function* (*CGF*) is integrated with the SBCF and transfers OCS *charging data records* (*CDR*) files to the operator's *Billing System*. In *Online Charging Functions* (*OCF*), there is another function called *Event Based Charging Function* (*EBCF*). However, we will not use it in this work.

In a message-exchange manner, Fig. 3 describes a session-based account reservation scenario to show how the serving

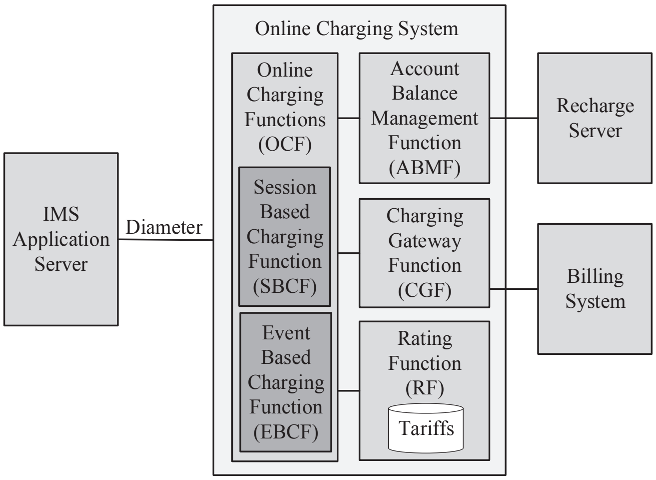

Fig. 2. Online Charging System architecture for IMS services [5].

network interacts with the SBCF and the ABMF over Diameter (e.g., Ro and Rc reference points) [6]. When the SBCF receives an online charging request for a certain service, it requests the account and counter information of the subscriber from the ABMF. It then manages the subscriber's context information and determines the tariff for the requested service by involving the RF. More importantly, SBCF has to determine the *Granted Unit* (*GU*) and create reservations per session before the services are delivered by the serving network. Here, we can imagine the ABMF as a "cash box" and the GU as a "wallet", where the wallet is the actual place where the serving network consumes granted units to deliver services to the user. Note that as the session parameters change (e.g. QoS), the GU may also be changed during a session. Also note that the reservation process may be repeated multiple times during a session. Finally, after the session ends, the serving network will notify the SBCF the used GU and the SBCF will request the ABMF to substract the actual consumed GUs from the cash box.

Since how to determine the proper amount of GU is an open issue, it deserves careful investigation for the FSP case. Fixed or multiplicative GU is typically adopted by telecommunications operators. However, both schemes can not adapt to dynamic consumer behaviors. In particular, with the widespread use of mobile devices, users may change their Internet behaviors in very short time. Obviously, for the FSP case, small GUs will incur too many signalings for reservation, but excessive ones will cause unbalance of resource distribution. This motivates us to consider the assignment of dynamic GUs for UEs belonging to the same FSP based on their historical data usage and total monthly data allowance.

The rest of this paper is organized as follows. Section II surveys related works. We propose our reservation scheme in Section III and compare its performance against the fixed and multiplicative schemes in Section IV. Section V concludes this paper.

II. Related Work

The 3GPP defines several specifications for OCS and Diameter. Reference [6] defines several Diameter-based offline and online charging applications in 3GPP networks including all charging parameters, scenarios, and message flows. Reference [7] specifies the charging functionality and charging management in 3GPP networks; it addresses the internal aspects of the OCS and contains the architecture and functions of the OCS logical components and the functionality of those OCS interfaces. Reference [8] specifies the policy and charging control functionality for evolved 3GPP packet switched domains, including both 3GPP accesses and non-3GPP accesses. Reference [9] is a study report of the Rc reference point supported by ABMF. The study covers the motivation, requirement analysis, architectures, key issues analysis, and recommendations. Reference [10] specifies a Diameter application that can be used to implement real-time credit-control for a variety of end-user services such as network access, *Session Initiation Protocol* (*SIP*) services, messaging services, and download services.

A lot of charging-related works have been developed based on the 3GPP specifications. Reference [11] propose three key aspects of online charging with respect to information utilization: a) signaling aspect, b) inter-domain aspect, and c) service- and component-based aspect. It also compares works in the literature based on the proposed criteria. Reference [12] defines a model and some usage scenarios in which the *User Charging Context* (*UCC*), such as the available budget in a subscriber's account or the available list of networks, may be utilized in online charging to increase "value for money". In [13], some comparable measures for future online charging solutions are proposed by identifying the relevant context-related charging events. Reference [5] develops an analytic model to investigate the performance of the *Recharge Threshold-based Credit Reservation* (*RTCR*) mechanism for *Universal Mobile Telecommunications System* (*UMTS*). Through this, network operators can select the appropriate recharge thresholds to reduce the probability of in-progress service sessions being forced-termination. To avoid session suspension during credit reservation in the OCS prepaid mechanism, [14] proposes a *Credit Pre-reservation Mechanism* (*CPM*) before the credits at the *gateway GPRS support node* (*GGSN*) are actually depleted.

III. The Proposed Scheme

To demonstrate how OCS works, we first give some examples below.

Example 1. This is a non-FSP case. Suppose that a user has a monthly data allowance of 5 gigabytes (GB). Every time when she surfs the Internet, SBCF will first check her balance (signaling 2 and 3 in Fig. 3). Then SBCF, for example, will reserve 10 megabytes (MB) from ABMF (signaling 7 and 8) and notify the serving network to start the service (signaling 9). After the GU is consumed, a request will be sent to SBCF (signaling 11) and then another reservation of 10 MB will be made via contacting ABMF (signaling 13 and 14). The process repeats until the session ends. Suppose that there is a 7 MB of

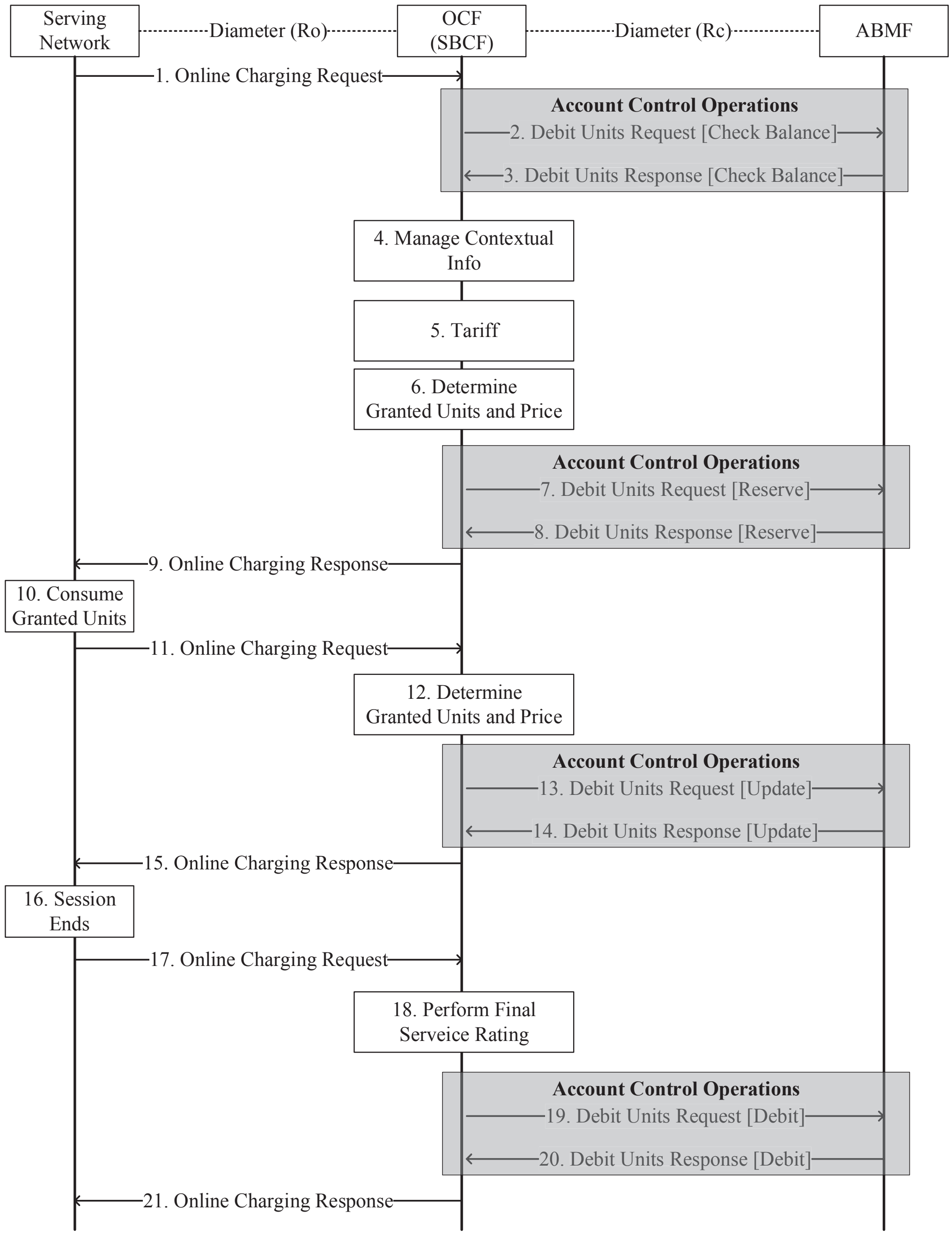

Fig. 3. Session-based reservation for an IMS service [7].

GU used after the session ends. This number will be notified to the SBCF (signaling 17) and then ABMF (signaling 19 and 20). Finally, the balance will be debited at the ABMF.

Example 2. Following the above example, the subsequent reservations do not need to be 10 MB each time. It can be dynamically adjusted based on how fast the GU is consumed.

Example 3. Consider the FSP case. Suppose that the plan includes 3 users with a total monthly allowance of 5 GB. Every time when a member uses the Internet, SBCF will first check the balance (signaling 2 and 3) and then make a proper reservation through ABMF (signaling 7 and 8) based on the users past usage pattern. The similar reservations can be made repeatedly when the GU is consumed. Clearly, the FSP scenario is more complicated because each user may keep some GU at hand when there is a session going on. Excessive reservations may make a wrong impression that the resource at the ABMF is depleted.

We now formally define the GU reservation problem in this work. We consider a FSP with M UEs, who share a total monthly data allowance of α MB over a charging period consisting of λ days. Our framework allows the subscribers to recharge after data overage, in which case we simply increase the value of α.

When UE_i enables an application (such as Youtube or Line), UE_i will be allocated a *Default GU*, which is a fixed amount of GU set by the operator, at ABMF. When the *Default GU* has been exhausted, SBCF will request a new GU. This is repeated RR_i runs (denoted as $GU_{i,j}$; j=1..RR_i) until the session ends. Here, RR_i is the *Reservation Runs* of UE_i, which is counted once whenever SBCF sends a request to ABMF for reserving a GU. The new GU may be different from the *Default GU* depending on the scheme being applied. For example, assuming *Default GU* is 10 MB, ABMF will reserve 10 MB for UE_i when she opens the web browser for work. UE_i can surf the Internet until 10 MB has been exhausted and then SBCF will get a new GU, say 20 MB if the total monthly data allowance is not exhausted yet. The problem is how to predict an appropriate GU for minimizing the total *Reservation Runs* in the FSP case, i.e.,

$$Min\left\{\sum_{i=1}^{M} RR_i\right\}. \tag{1}$$

A. Fixed Scheme (FS)

This is the most basic scheme, where the SBCF always reserves a fixed amount of GU when there is a request received by the SBCF. We use this as a baseline for comparison purpose.

B. Individual Scheme (IS)

In this scheme, we consider each UE's usage pattern in the past charging period per session. Let the total data usage by UE_i in the past month be $PreD_i$ and the total number of sessions initiated by UE_i be S_i. Whenever a request is made by UE_i, we let

$$GU_i = \frac{PreD_i}{S_i} \tag{2}$$

be the GU to be reserved for UE_i. As there are more requests, the same value can still be used. Despite being easy to implement, IS is not suitable for FSP because SBCF may reserve too large GU and cause the resource occupied problem when some UE_i has heavy data usage but few sessions.

C. Multiplicative Scheme (MS)

The *Multiplicative Scheme (MS)* uses a *Default GU* first and counts the number of reservation runs made by a UE during a session. It increases GU by the default value after every m reservation runs. Specifically, let Δ_{def} be the *Default GU* whenever a new session is started. The SBCF should count the number of reservation runs made by a UE during a session. Let $GU_{i,j}$ be the amount to be reserved for the j-th request made by UE_i during a session. We let

$$GU_{i,j} = \lceil j/m \rceil \times \Delta_{def}. \tag{3}$$

Here, Δ_{def} can be set by the past experience. For example, it can be the value determined by Eq. (2) in the Individual Scheme.

D. Dynamic Reservation Scheme (DRS)

Here we propose a scheme, called Dynamic Reservation Scheme (DRS), for dynamically assigning GUs to UEs belonging to the same FSP based on their recent data usage patterns. We assume that the operator has logged data usage of each UE_i in the past. There are three kinds of logged data:

- Long-term statistics: This logs the average session duration A_i^d and the corresponding data rate A_i^r of a particular period (say 7:00 PM $\sim$ 8:00 PM) in the past 30 days.
- Short-term statistics: This logs the session's average data rate B_i^r in the past *P* hours (say 0.5, 1, or 2 hours) before the current time.
- Current-session statistics: This logs the (currently on-going) session's average data rate C_i^r and its remaining GU at hand C_i^q so far. To achieve this goal, we suggest to change OCS by requesting the serving network to report C_i^r and C_i^q to SBCF periodically with an interval of Δ_q. For this reason, we replace RR_i by the sum of signalings including *Reservation Runs* and reporting signalings. Whenever UE_i is allocated some GU from SBCF, a timer Δ_q is set. When this timer expires, a report to SBCF is sent and Δ_q is reset.

The main workflow of the scheme is illustrated in Fig. 4. How DRS works is sketched as follows.

1) When a request is received due to a newly started session, we set t_0 to be the current time retrieve the A_i^d and the A_i^r corresponding to t_0. Then let

$$GU_i = A_i^d \times A_i^r \tag{4}$$

and $t_{end}^i = t_0 + A_i^d$ to be the expected time for this GU_i to be exhausted. On the other hand, when a request is received because the previous GU is exhausted, let t_1 be the current time. We use the short-term statistics to set

$$GU_i = \left\{\begin{matrix} (t_{end}^i - t_1) \times B_i^r & ; & if\ t_{end}^i > t_1 \\ A_i^d \times B_i^r & ; & otherwise \end{matrix}\right\}. \tag{5}$$

Note that GU_i is a tentative value, which may be changed by step 2.

2) Since we are concerned about a FSP, we have to take into account the on-going sessions of other family members. Let UE_k be another family member who has an on-going session, $k \in M$. Recall that we have logged for UE_k two statistics: C_k^r and C_k^q. We will calculate a value EGU_k, which stands for UE_k's future need of GU after it finishes using its GU_k. Let its reporting time be t_2. Then the expected GU for UE_k to complete its current session after t_2 is

$$X = (t_{end}^k - t_2)^+ \times C_k^r. \tag{6}$$

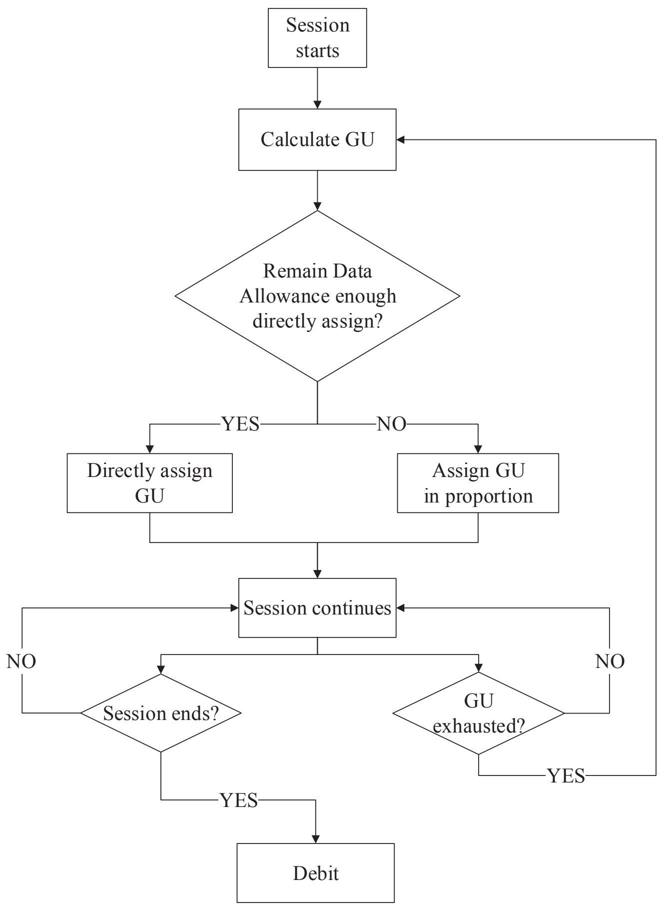

Fig. 4. Workflow of DRS.

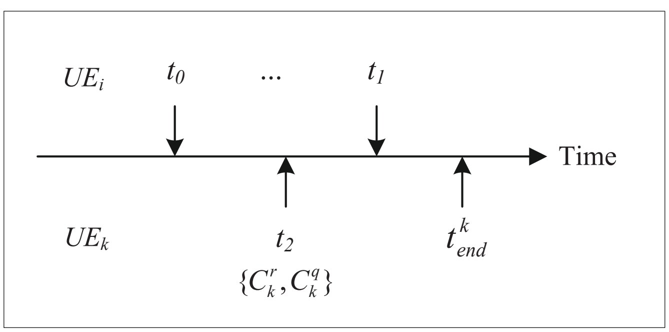

Fig. 5. Example of calculating EGU_k.

Here, $(\cdot)^+$ returns 0 if the value inside the parentheses is negative. Since at t_2, the remaining GU of UE_k is C_k^q,

$$EGU_k = (X - C_k^q)^+. \quad (7)$$

The above concept is shown in Fig. 5.

3) Let *RD* be the remaining data allowance of this FSP in this charging period. We check the following condition:

$$GU_i + \sum_{k \neq i} EGU_k \geq RD. \quad (8)$$

If the condition is true, the remaining data allowance is sufficient. We will keep the GU_i in step 1 unchanged. Otherwise, we should normalize the GU_i as follows. Let *Sum* = $GU_i + \sum_{k \neq i} EGU_k$. Then we let

$$GU_i^{'} = \frac{GU_i}{Sum} \times RD. \quad (9)$$

The new $GU_i^{'}$ will is the actual value to be returned to the serving network.

IV. Performance Evaluation

In this section, we develop a simulator in C++ language to verify the effectiveness of the proposed scheme. Table I and Table II show the assumed behaviors of three occupations, including geek, worker and student. For an FSP, we add geeks, workers and students into it and observe the impact of the short-term statistics of our proposed scheme, Dynamic Reservation Scheme (DRS), on the number of *Reservation Runs*. In the simulation, we compare DRS with Fixed Scheme (FS), Individual Scheme (IS) and Multiplicative Scheme (MS). Note that each simulated result is averaged by at least 10000 experiments.

TABLE I
Notation of the assumed behaviors

Notation	Definition
H	High usage ; 700 - 1100 KB per minute
M	Middle usage ; 400 - 600 KB per minute
L	Low usage ; 100 - 300 KB per minute
X	None usage

TABLE II
Assumed behavior of the occupations

Time	Geek	Worker	Student
00:00 - 01:00	H	X	X
01:00 - 02:00	H	X	X
02:00 - 03:00	H	X	X
03:00 - 04:00	H	X	X
04:00 - 05:00	X	X	X
05:00 - 06:00	X	X	X
06:00 - 07:00	X	X	X
07:00 - 08:00	X	H	M
08:00 - 09:00	X	M	L
09:00 - 10:00	X	M	L
10:00 - 11:00	X	M	L
11:00 - 12:00	X	M	L
12:00 - 13:00	L	L	L
13:00 - 14:00	L	M	L
14:00 - 15:00	L	M	L
15:00 - 16:00	L	M	L
16:00 - 17:00	L	M	L
17:00 - 18:00	L	H	M
18:00 - 19:00	L	L	M
19:00 - 20:00	H	L	M
20:00 - 21:00	H	L	H
21:00 - 22:00	M	H	H
22:00 - 23:00	H	H	L
23:00 - 24:00	H	L	X

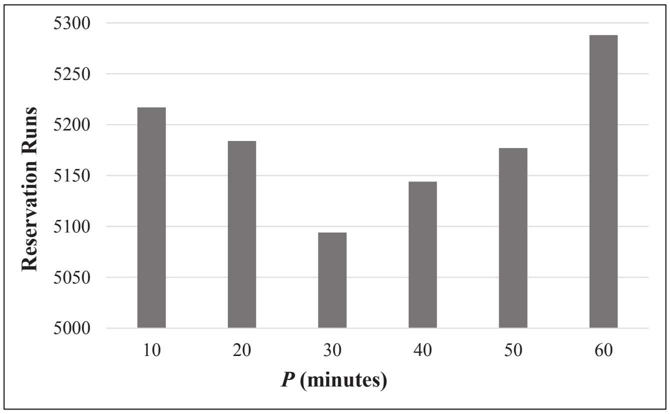

Fig. 6. The impact of logging period on reservation runs.

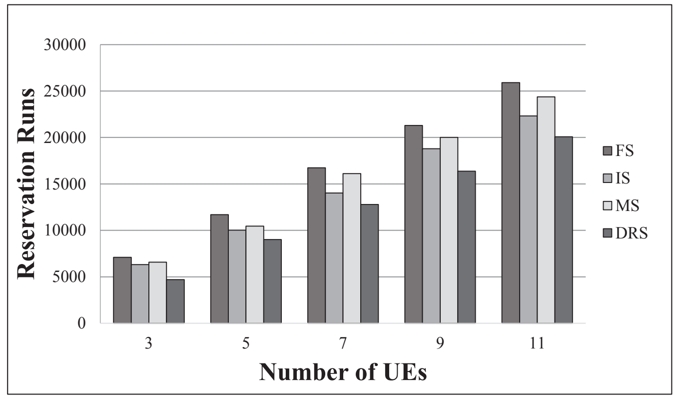

Fig. 7. Comparison of different schemes on the reservation runs.

A. Inference of Short-term Statistics

First, we observe the influence of the logging period (*P*) on *Reservation Runs* of the proposed DRS. We assume that there are three members belonging to the same FSP and their assumed behaviors of each GU_i is set randomly based on the Table II. In Fig. 6, the new GU_i will be calculated while the previous one is exhausted. We can observe that the number of *Reservation Runs* decreases dramatically because the sufficient logging period can help our scheme to predict GU_i precisely. However, *Reservation Runs* will increase when $P \geq 40$. Therefore, we choose 30 as the most appropriate value of the logging period for our scheme.

B. Comparisons on Reservation Runs

Now, we investigate the impact of different numbers of UEs in the same FSP on *Reservation Runs* of all schemes. Note that the assumed behavior of each GU_i is set randomly based on the Table II. Every time a reservation request comes, we count on the signaling counter. In the MS scheme, we set *m* as 2, i.e., the $GU_{i,j+1}$ for each GU_i will be multiplied by 2 once $GU_{i,j}$ exhausted every two times in the same on-going session. In the DRS scheme, once an on-going session reports its current statistics, a signaling should also be considered. We can observe that DRS can reduce *Reservation Runs* at least 22% compared to fixed scheme because the proposed DRS can assign dynamical size of GUs for UEs based on their historical data usages and total monthly data allowance.

V. Conclusions

In this paper, we have proposed a new reservation scheme which assigned the more appropriate size of GUs for UEs in the same FSP based on their historical data usages and total monthly data allowance. Simulation results have shown that the proposed DRS can reduce the communication signaling at least 22%, as compared to fixed scheme under unpredictable behaviors. In the future work, we will consider how to predict the appropriate session duration and thus we can calculate the most appropriate size of GUs for all UEs in the same FSP to further save signalings.

Acknowledgments

Y.-C. Tseng's research is co-sponsored by CHT, Taiwan.

References

[1] Network Working Group RFC 6733. Diameter Base Protocol; Request for Comments, October 2012.

[2] Oracle White Paper. LTE Diameter Signaling Index. 2nd Edition, December 2013.

[3] First family share plan. http://www.cnet.com/news/at-t-mobility-ceo-family-data-plan-coming-soon-scoop/.

[4] H.-S. Huang, T.-C. Su, J.-C. Wu, W.-D. Tsai, Y.-C. Tseng, F.-C. J. Lin, B.-T. Liny, and W.-H. Hu. A Comparison of 4G Telecommunications Tariff Plans in Asia Countries. *Asia-Pacific Network Operations and Management Symposium (APNOMS)*, pages 1–6, 2014.

[5] S.-I. Sou, H.-N. Hung, Y.-B. Lin, N.-F. Peng, and J.-Y. Jeng. Modeling Credit Reservation Procedure for UMTS Online Charging System. *IEEE Trans. on Wireless Communications*, 6(11):4129–4135, 2007.

[6] 3GPP TS 32.299. 3rd Generation Partnership Project; Technical Specification Group Service and System Aspects; Telecommunication management; Charging management; Diameter charging applications; Technical Specification. v12.3.0 (Release 12), December 2013.

[7] 3GPP TS 32.296. 3rd Generation Partnership Project; Technical Specification Group Service and System Aspects; Telecommunication management; Charging management; Online Charging System (OCS): Applications and interfaces; Technical Specification. v12.1.0 (Release 12), December 2013.

[8] 3GPP TS 23.203. 3rd Generation Partnership Project; Technical Specification Group Service and System Aspects; Policy and charging control architecture; Technical Specification. v12.3.0 (Release 12), December 2013.

[9] 3GPP TR 32.825. 3rd Generation Partnership Project; Technical Specification Group Service and System Aspects; Telecommunication management; Charging management; Rc reference point study; Technical Report. v10.0.0 (Release 10), March 2010.

[10] Network Working Group RFC 4006. Diameter Credit-Control Application; Request for Comments, August 2005.

[11] T. Grgic and M. Matijasevic. An overview of online charging in 3GPP networks: new ways of utilizing user, network, and service-related information. *International Journal of Network Management (IJNM)*, 23(2):81–100, 2012.

[12] T. Grgic and M. Matijasevic. "Smarter" online charging for over-the-top services by introducing user context. *Joint IFIP Wireless and Mobile Networking Conference (WMNC)*, pages 81–87, 2012.

[13] T. Grgic and M. Matijasevic. Performance metrics for context-based charging in 3GPP Online Charging System. *International Conference on Telecommunications (ConTEL)*, pages 171–178, 2013.

[14] H.-Y. Lee and Y.-B. Lin. Credit pre-reservation mechanism for UMTS prepaid service. *IEEE Trans. on Wireless Communications*, 9(6):1867–1873, 2010.