


RESEARCH

Open Access



Internet data budget allocation policies for diverse smartphone applications

Saidur Rahman^{1†}, Anika Anzum Prima^{1†}, Md. Abdur Razzaque^{1*†} , Mohammad Mehedi Hassan², Abdulhameed Alelaiwi², Majed Alrubaian² and Kim-Kwang Raymond Choo^{3,4}

Abstract

In recent times, there has been a significant growth in the number of smartphone users and number and types of mobile applications (apps). Such a trend has resulted in increased Internet data consumption, particularly for users of “data hungry” apps. Thus, smartphone apps should be allocated to their required budget to minimize resource wastage without compromising on user’s quality of experience. In this paper, we develop a prioritized and dynamic budget allocation policy framework for ensuring an optimal budget allocation to each app as well as improving system performance. We formulate the optimal Internet data budget management (O-IDM) problem as a mixed-integer nonlinear programming (MINLP) problem, which maximizes the resource utilization and minimizes user penalties. We also employ runtime monitoring technique to estimate future bandwidth utilization so as to ensure budget reservation as close as to the required amount. A heuristic Internet data budget management algorithm (H-IDM) is also presented, which is designed to reduce time complexity and computational overhead of the O-IDM system. The experimental results from test-bed implementation demonstrate the effectiveness of the proposed IDM systems, in comparison to state-of-the-art approaches.

Keywords: Internet data budget, Smartphone applications, Dynamic budget allocation, Mixed-integer nonlinear programming optimization, Resource utilization

1 Introduction

In modern-day society, smartphones and mobile applications are two popular consumer technologies. This is also evidenced by the diversity of the devices (e.g., wide ranging models and makes) and applications (e.g., social, productivity, and lifestyle) [1–5]. A typical smartphone is capable of running several applications concurrently (e.g., navigation, email, browser, dating, and communication) [6–8], in addition to other built-in functionalities such as music player, camera, and sensors [9–14]. Mobile devices including smartphones could be the predominant digital device for future daily use, as it allows users to access the Internet and their data anywhere anytime [15–17]. Thus, it is important to ensure that next-generation mobile communications are cost efficient and secure and provide high quality of service and experience.

Many of these mobile applications consume significant bandwidth via Wi-Fi or mobile network, for example, using cloud applications to access or upload a large document on-the-go. If Wi-Fi signal is absent or too weak, then the mobile can autonomously switch to mobile communication [18]. Every application needs to be allocated to limited mobile Internet data budget according to its bandwidth utilization. Consequently, a significant amount of Internet data budget could be consumed by some low-priority applications running in the background (e.g., gaming), exhausting the available data allowance in a fixed Internet data budget plan. Consequently, the user would not be able to access or upload data without paying for additional bandwidth. If we impose restrictive budget allocations among applications, it could result in the user under-utilizing the Internet data budget in a particular billing cycle. Therefore, it is essential to have the capability to flexibly set/adjust an appropriate budget plan for mobile applications (e.g., using smart policies to provision different level of bandwidth resources to different applications) to maximize bandwidth resources and

*Correspondence: razzaque@du.ac.bd

†Equal contributors

¹Green Networking Research Group, Department of Computer Science and Engineering, University of Dhaka, Dhaka 1000, Bangladesh
Full list of author information is available at the end of the article

quality of user experience. If an application is allocated less budget than it requires, then this could result in *under-provisioning*, affecting its performance, and consequently quality of user experience. On the other hand, if an application is allocated too much budget, then it can result in *over-provisioning* and consequently, resource-underutilization (see Fig. 1). Thus, the aim of this work is to ensure efficient and systematic Internet data budget utilization for each application for avoiding over- or under-provisioning.

Most state-of-the-art research such as [18–21] do not consider dynamic budget allocation for mobile applications. The solution in [19], for example, seeks to satisfy the bandwidth requirement for constant and variable bit rate connections, while minimizing blocking probability. In [18], a heuristic solution for allocating budget to sensitive and non-sensitive applications was developed. However, it did not handle over- and under-provisioning issues while allocating portion of data budgets. In other words, the research did not allow the allocation of different resources to different applications. In our earlier work [22], an optimization framework for Internet data budget management was developed. However, the framework was designed for small systems, but it is not suited for larger system deployment due to the high computational complexity associated with an increased number of user applications. This is the gap we seek to contribute to in this paper.

In this work, we explore dynamic budget allocation policies that allow us to analyze the budget usage behavior of each application and decide the amount of bandwidth allocation for each application so as to maximize resource utilization. Our proposed optimal Internet data budget management (O-IDM) system leverages prioritized budget allocation to applications in order to minimize penalty for important and sensitive applications. We then develop a heuristic IDM algorithm (H-IDM) designed to reduce time complexity and computational overhead while maintaining efficient data budget utilization. Specifically, in this paper,

- we formulate the problem of optimal allocation of mobile Internet data budget (O-IDM) to different

mobile applications as a mixed-integer nonlinear programming (MINLP) problem;

- we propose O-IDM which allows us to maximize budget utilization for all applications while minimizing budget allocation errors due to under- and over-provisioning problems;
- we present a heuristic IDM algorithm (H-IDM) designed to reduce computation complexity of O-IDM;
- we present a runtime monitoring and measurement scheme to estimate budget utilization using weighted average usage prediction (WAUP) method; and
- we study the effectiveness of autoregressive integrated moving average (ARIMA) model to more accurately predict the amount of budget in future Internet data budget plan.

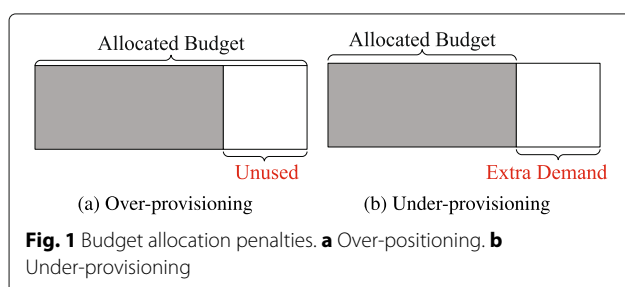
We then evaluate the performance of the proposed model with a number of existing models.

The rest of the paper is organized as follows. Section 2 reviews related work. Section 3 presents our system model. In Section 4, we formulate the optimization problem and present the proposed dynamic budget allocation scheme. Section 5 describes the effectiveness of H-IDM scheme. Section 6 presents findings from the performance evaluation, and conclusions are drawn in Section 7.

2 Related works

Allocating bandwidth among competing users or devices is a challenging problem, and this topic has been studied in the literature. In [19], the authors presented a utility-based bandwidth allocation algorithm for multiple services in heterogeneous wireless access networks consisting of WMAN, 3G cellular network, and WLAN. Bandwidth is allocated to a new arrival connection in heterogeneous wireless environment depending on utility fairness. Chen et al. [23] proposed a smart bandwidth allocation algorithm based on mobile users' personality traits and channel condition. Based on one user's data usage, the service provider could estimate this user's probability of each personality trait using diagnostic inference and then based on predictive inference to calculate this user's usage of bandwidth in the future. In [24], the authors studied the bandwidth disposition problem for heterogeneous networks. Their proposed method determines the amount of disposed bandwidth, and the decision to upgrade or downgrade the sequence of bandwidth is based on the upgrade rank or downgrade rank function.

The researchers of [25] have designed a new framework to model mobile applications and proposed efficient algorithm (MuSIC) which achieves more than 75 % of optimal solution when the number of mobile users is high. In [26–28], the authors presented traffic congestion management strategies using crowd sensing technology.



Fang et al. [29] proposed an online control algorithm for throughput-energy tradeoff in mobile device. An online incentive-compatible VM allocation mechanism is developed in [30], and resource allocation algorithms are proposed in [31, 32] to enhance network capacity. A control framework for participatory sensing system using smartphones is developed in [33].

Chaisiri et al. [34] proposed a resource provisioning mechanism in cloud computing to offer cloud users two resource provisioning plans, namely, on-demand plan and reservation plan. Their proposed optimal cloud resource provisioning algorithm minimizes under- and over-provisioning problems due to the uncertainty of user's future demand and provider's resource prices. In [35], Dong et al. proposed a randomized task assignment framework for computing tasks on participatory smartphones in a reliable fashion while minimizing workload at individual participatory devices. Martignon et al. [36] examined the potential of wireless mesh networks (WMN) in providing broadband Internet access to mobile users. A WMN operator may be reluctant to develop a bandwidth market place as some dishonest customers can put in false bids in order to pay a lower price. To overcome this problem, they proposed a mechanism to allocate the bandwidth of WMN operators to customers who are interested to pay the higher price as per their bandwidth demand. This strategy ensures that all customers reveal their real valuation of the required bandwidth. In [37], the authors studied the problem of network utilization decline due to significant traffic variability in datacenters. They then designed a virtual machine bandwidth allocation algorithm to handle highly dynamic traffic in datacenters. In other words, the proposed solution seeks to minimize performance degradation caused by frequent decreases in network throughputs.

The authors of [18] introduced an online 3G budget algorithm that decides which sensory data should be uploaded via 3G communication while others will be uploaded or downloaded later when Wi-Fi access point is available. Their optimization scheme ensures efficient 3G budget utilization but the algorithm has significant computational overheads. Therefore, the approach is both computational resource and energy hungry. Also, they proposed a heuristic algorithm which splits the 3G budget in each time cycle into two pieces, namely, reserved budget and flexible budget. Sensitive applications use reserved budget and non-sensitive applications use flexible budget. If the reserved budget runs out, then sensitive applications use the allocation from the flexible budget. But this two-state application classification (sensitive and non-sensitive) decreases the dynamicity and flexibility of bandwidth allocation. In addition, budget allocation strategies for heterogenous applications based on urgency/priority are not been explicitly discussed or analyzed.

In [22], the authors proposed an optimal budget allocation policy for each mobile application. However, their optimization scheme does not provide effective and fair utilization of 3G budget because utilization is set to zero if under-provisioning penalty occurs. Again, their budget usage ratio measurement is erroneous as a result of inaccurate budget reservation in upcoming time cycles. Another limitation of their work is that time complexity increases exponentially if the number of applications and time cycles are sufficiently large.

3 System model

In this section, we present the system model for Internet data budget utilization. We consider that a smartphone is connected with the Internet either using Wi-Fi access point or by 3/4/5G mobile Internet connection. The smartphone uses Internet data budget for urgent application usage whenever no Wi-Fi access point is available at nearby. We assume that some low-priority applications are allowed to buffer the data packets at local device until it is connected with any AP. In that case, the buffer space of the mobile phone is exhausted; either it stops data collection process for very low-priority applications or turns on mobile network Internet connection to offload collected data. When a user is in the range of a Wi-Fi access point, all the backlogged data packets in the buffer are uploaded to the destination server through Wi-Fi communication regardless of its priority. However, if Wi-Fi signal is absent or too low to upload the important data packets, then the system switches to mobile Internet communication.

We assume that a user purchases a fixed limited Internet data budget B for 3/4/5G Internet connection (e.g., 3GB monthly, 1GB weekly package). Let $|\mathcal{N}|$ represent the set of all applications running on a smartphone and their priorities are denoted by the set, $\mathcal{P} = \{p_1, p_2, p_3, \dots, p_{|\mathcal{N}|}\}$. In this case, we use higher values for high priority applications. The amount of the Internet data budget plan for each of the applications is proportional to how much important the application is. That is, real-time and interactive applications need more bandwidths and they may not tolerate significant delay; on the other hand, some low-priority applications may be delayed and reduced amount of data budget can be allocated. In this work, we dynamically prioritize all the applications running in the mobile device by estimating bandwidth usage behavior of the applications. The more bandwidth an application uses, the higher its priority is. We exploit the ARIMA formulae for estimating the runtime usage of resources by different applications and recommend a user the most appropriate amount of monthly Internet data budget plan (to be discussed in detail in Section 4.4).

The Internet data budget plan period \mathcal{T} is the validation time for the Internet package of the user (e.g.,

1 month, 1 week). Each data budget plan period \mathcal{T} is divided into multiple time cycles, as depicted in Fig. 2. We also assume that the budget allocation algorithm periodically runs every after t time cycle where t changes proportionately with \mathcal{T} and B . It tries to avoid over-provisioning as well as under-provisioning so as to maximize the resource utilization and application performances. At every scheduling interval t , a certain amount of data budget $r_{i,t}$ is reserved for each application $i \in \mathcal{N}$ and an amount $f_{i,t}$ as flexible budget; the application i uses data budget from $f_{i,t}$ in the case that the reserved amount $r_{i,t}$ is exhausted before the next time interval $t + 1$. The notations used in this paper are summarized in Table 1.

4 Internet data budget management

In this section, we present optimal Internet data budget management strategy (O-IDM) for heterogeneous applications running in a smartphone. The proposed budget allocation policy dynamically expands or shrinks the amount of bandwidth allocated to different applications over time based on the usage behavior of the applications. Our budget allocation mechanism optimizes the bandwidth resource utilization as well as reduces the penalties incurred due to over- and under-provisioning. We exploit the WAUP method to infer more accurately the bandwidth usage in future time cycles. We use the ARIMA model for recommending appropriate amount of monthly data budget plan for a user.

4.1 Optimization problem formulation

The problem of optimal allocation of bandwidth to the mobile applications is translated as maximizing the utilization of resources while minimizing the penalties incurred due to over-provisioning and under-provisioning. And, this policy needs to be maintained for all applications in all time cycles. Our target is to determine an accurate reservation of budget to an application in t_{m+1} time cycle based on usage statistics of previous M time cycles, i.e., from t_{m-M} to t_m such a way that maximization of utilization is ensured. Therefore, the optimization function is a mixed-integer non-

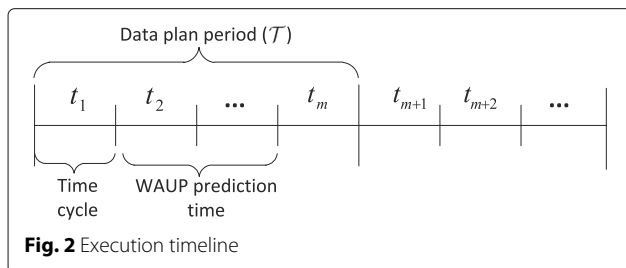


Table 1 Notations used in this paper

Notation	Description
\mathcal{N}	Set of all applications in a smartphone
\mathcal{T}	Internet data budget plan period
\mathcal{P}	Set of priorities of all applications
B	Internet data budget
B_1	Reserved budget
B_2	Flexible budget
t_m	m 'th time cycle
α	Control parameter
$U_{i,t}$	Resource utilization of application i at time cycle t
$r_{i,t}$	Amount of reserved budget for application i at time cycle t
$x_{i,t}$	Used amount of reserved budget for application i at time cycle t
$f_{i,t}$	Amount of flexible budget for application i at time cycle t
$y_{i,t}$	Used flexible budget for application i at time cycle t
ρ_i	Priority of i 'th application
Δ	Time interval for algorithm execution
$b_{i,t}$	Bandwidth usage ratio of i 'th application at time cycle t
$\bar{b}_{i,t}$	Estimated bandwidth usage ratio of i 'th application at time cycle t
M	Number of time cycles for WAUP model
w_i	Weight of i 'th time cycle
e_t	Uncorrelated Gaussian noise
θ_t	Moving average coefficient
L	Lag operator

linear programming (MINLP) problem, formulated as follows,

$$\text{Maximize: } Z = \sum_{t=1}^{|\mathcal{T}|} \sum_{i=1}^{|\mathcal{N}|} U_{i,t} \tag{1}$$

subject to:

$$x_{i,t} \leq r_{i,t}, \tag{2}$$

$$y_{i,t} \leq f_{i,t}, \tag{3}$$

$$x_{i,t} + y_{i,t} \leq r_{i,t} + f_{i,t}. \tag{4}$$

Here, $U_{i,t}$ is the resource utilization of application i at time cycle t . Given that the $r_{i,t}$, $x_{i,t}$ and $y_{i,t}$ are the amount of reserved budget, used reserved budget, and

used flexible budget for application i at time cycle t , the utilization is defined as follows:

$$U_{i,t} = \begin{cases} 1 - \frac{y_{i,t}}{r_{i,t}} & \text{if } \frac{x_{i,t} + y_{i,t}}{r_{i,t}} > 1 \\ \frac{x_{i,t}}{r_{i,t}} & \text{if } \frac{x_{i,t} + y_{i,t}}{r_{i,t}} \leq 1. \end{cases} \quad (5)$$

We observe from the first condition of Eq. 5 that the resource utilization decreases with the increasing usage of flexible budget $y_{i,t}$, which occurred due to under-provisioning of resources. Furthermore, the second condition depicts that the utilization is also decreased in the case that resource over-provisioning occurred, i.e., only a small part of reserved resource is utilized. In summary, the more appropriate amount of budget that we can allocate just to meet the actual requirement of an application, the more the resource utilization is increased and vice-versa. The constraints (2) and (3) are corresponding to bandwidth usage constraints for reserved and flexible budgets, i.e., usage must be bounded by the proportionately allocated amount for an application i at a given time cycle t . The constraint (4) states that the constraints (2) and (3) follow additive rule. However, increasing the reserve budget for an application does not always increase the utilization. The resource utilization is enhanced when the gap between the reserved amount and the usage amount is decreased. Similarly, excessive allocation of flexible data budget amount for a certain application might cause penalty to some other applications, degrading the overall resource utilization. We use a popular mathematical programming language AMPL [38] for solving the optimization problem. What follows next is that we describe in detail the budget allocation policies to different applications $i \in \mathcal{N}$ at each time cycle t .

4.2 Budget allocation policy

The total Internet budget B for a data plan is split into two parts: reserved budget (B_1) and flexible budget (B_2) in each time cycle. Initially ($t = 0$), their values are determined as follows:

$$B_1 = \alpha \times B, \quad (6)$$

$$B_2 = B - B_1, \quad (7)$$

where α is a control parameter that determines how much of the total budget is to be kept in reserved portion. When the Internet data budget plan period starts, every application is allocated a certain amount of reserved budget based on their priorities assuming all the applications have equal bandwidth usage. For the subsequent time cycles, the amount of reserved budget for an application is forecasted following its estimated bandwidth usage $\bar{b}_{i,t}$ during last M time cycles; the detail estimation process of $\bar{b}_{i,t}$ is

presented in Section 4.3. Therefore, the reserved budget of an application i at a given time cycle t is calculated as

$$r_{i,t} = \frac{p_i \times \bar{b}_{i,t}}{\sum_{i=1}^{|\mathcal{N}|} (p_i \times \bar{b}_{i,t})} \times B_1 \quad (8)$$

where p_i and $r_{i,t}$ are, respectively, the priority and reserved budget of an application $i \in \mathcal{N}$ in the time cycle t . In Eq. 8, budget reservation to each application is defined as a weighted average of priority and estimated budget usage so that judicious amount of budget is reserved to each application according to its requirement. We assume that in the first time cycle t_1 , each application i has used $x_{i,t}$ amount of data from the reserved budget. So, remaining reserved budget is $B_1 - \sum_{i=1}^{|\mathcal{N}|} x_{i,t}$.

In the case that an application has run out of its reserved budget within the current time cycle, then flexible budget is allocated to it from B_2 . If $f_{i,t}$ denotes the extra budget requirement for i 'th application in time cycle t , then

$$f_{i,t} = \frac{p_i \times T'}{\mathcal{T} \times \sum_{i=1}^{|\mathcal{N}|} p_i} \times B_2, \quad (9)$$

where \mathcal{T} is the Internet data budget plan period and T' is current time of a data plan period. Equation 9 refers to the allocated flexible budget for i 'th application in t time cycle based on both application priority and current time of data plan period. If flexible budget is allocated based on priority, only then some applications would face significant amount of budget wastage at the end of the data plan period as well as many important applications would be deprived of sufficient amount of flexible budget. If a high-priority application runs out of flexible budget, then it is assigned flexible budget amount repeatedly following its requirement; but, a low-priority application is allowed to use flexible budget one time only. This policy helps us to restrict uncontrolled usage of available data budget by low-priority applications and thus reduce penalties for high-priority ones.

If each application $i \in \mathcal{N}$ uploads or downloads $y_{i,t}$ amount of data using flexible budget, then the remaining flexible budget is $B_2 - \sum_{i=1}^{|\mathcal{N}|} y_{i,t}$. The remaining total budget after t_1 time cycle is calculated as

$$B = \left(B_1 - \sum_{i=1}^{|\mathcal{N}|} x_{i,t} \right) + \left(B_2 - \sum_{i=1}^{|\mathcal{N}|} y_{i,t} \right) \quad (10)$$

This is the budget for next time cycle; that means the assignment is additive. We calculate the total budget by adding remaining flexible and reserved budget for the next time cycle. From the second time cycle, the reserved budget is calculated according to following equation:

$$B_1 = \frac{T' + \alpha \times (\mathcal{T} - T')}{\mathcal{T}} \times B \quad (11)$$

Equation 11 helps us to dynamically update the reserved budget amount B_1 following the historical usages of data budget by the applications. It also minimizes the wastage of bandwidth later at the end of the Internet data budget plan period. The control parameter α plays an important role to start with minimum reserved amount from the first day of Internet data budget plan and to increase gradually. Therefore, it helps to minimize both the over- and under-provisioning penalties. The value of α depends on execution frequency of the budget allocation algorithm compared to the total Internet data budget plan period. For performance evaluation, we have set $\alpha = \frac{\Delta}{T}$, where Δ is the time interval between two consecutive executions of data budget allocation algorithm.

4.3 Estimation of budget usage

We calculate budget usage ratio $b_{i,t}$ after each usage interval, t , as follows,

$$b_{i,t} = \frac{x_{i,t} + y_{i,t}}{r_{i,t}} \tag{12}$$

Thus, Eq. 12 refers to how much of the allocated budget is used by an application $i \in \mathcal{N}$. We need to predict the budget usage of each application so as to infer the judicious amount of budget to be allocated in the upcoming time cycle. The possible amount of usage of the budget by an application in the next time cycle typically depends on its historical usage patterns. And the most recent usage behavior puts more impact on the future usage estimations. In this work, we exploit the WAUP method similar to the WALI model [39, 40] that works as follows. The WAUP measures the average bandwidth usage of an application $i \in \mathcal{N}$ in the current time cycle as a weighted average of last M time cycles as follows:

$$\bar{b}_{i,t} = \frac{\sum_{j=1}^M (w_j \times b_{i,j})}{\sum_{j=1}^M w_j} \tag{13}$$

where the value of weight factor w_j is determined as follows,

$$w_j = \begin{cases} 1 & \text{if } 1 \leq j \leq \frac{M}{2} \\ 1 - \frac{j - \frac{M}{2}}{\frac{M}{2} + 1} & \text{if } \frac{M}{2} < j \leq M \end{cases} \tag{14}$$

For $M = 8$, this gives weights of 1, 1, 1, 1, 0.8, 0.6, 0.4, and 0.2 for w_1 through w_8 , respectively, where the most recent four samples are equally weighted.

4.4 Budget recommendation

Based on the user's recent monthly Internet data budget usage behavior, we explore a time series analysis method to predict the future Internet data budget plan so that insufficiency or wastage of bandwidth can be avoided. A user can use a different Internet data budget for different Internet data budget plans. The usage statistics typically

show a constant variance in its fluctuation with the course of time. The ARIMA model works best for forecasting because in this case, data exhibits a consistent pattern over time with minimum amount of outliers [41]. The ARIMA(p, d, q) models are first introduced by Box and Jenkins in 1970 [42] for purposes of modeling time series data. The model is the combination of autoregression and a moving average model. The full form of ARIMA can be written as [43, 44]

$$B'_l = c + \phi_1 B'_{l-1} + \dots + \phi_p B'_{l-p} + \theta_1 e_{l-1} + \dots + \theta_q e_{l-q} + e_l \tag{15}$$

or

$$B'_l = c + \sum_{i=1}^p \phi_i B'_{l-i} + \sum_{i=1}^q \theta_i e_{l-i} \tag{16}$$

or by using lag polynomial operator,

$$\nabla^d B'_l \phi(L) = \theta(L) e_l, \tag{17}$$

where $\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \phi_3 L^3 - \dots - \phi_p L^p$ and $\theta(L) = 1 - \theta_1 L - \theta_2 L^2 - \theta_3 L^3 - \dots - \theta_q L^q$. B'_l is the correlated normally distributed random variable, e_l is an uncorrelated Gaussian noise, θ_l is moving average coefficient, and L is the lag operator.

For predicting the amount of Internet data budget for the next data plan period, ARIMA model is employed on historical usage behavior. Here, B'_l is the estimated Internet data budget for the l th Internet data budget plan period. If the usage behavior has high degree of fluctuations in the amount of data budget over time, i.e., the historical usage is not stationary, then $B'_l = B_l - B_{l-1}$ is used for forecasting purposes.

5 Heuristic Internet data budget management

For allocating reserved budget in the current time cycle, estimated budget usage needs to be calculated based on historical bandwidth usage patterns of the applications. The estimated bandwidth usage of an application is measured as a weighted average of the last M time cycles. We need to maintain a $M \times |\mathcal{N}|$ matrix to record the usage statistics of $|\mathcal{N}|$ applications for previous time cycles. The main drawback of O-IDM algorithm is that in the case $|\mathcal{N}|$ and M are both large enough, then the computational complexity of Eq. 1 increases exponentially. The computational time goes up exponentially with higher number of applications since it requires huge amount of computation and storage to update the matrix.

Considering this problem, we propose a heuristic Internet data budget management (H-IDM) algorithm to provide a feasible and real-time solution. The idea is to minimize time complexity by decreasing the value of M when $|\mathcal{N}|$ is larger. Again, when $|\mathcal{N}|$ is smaller, larger value of M is taken to increase the accuracy level of bandwidth

Algorithm 1 Heuristic IDM algorithm

INPUT: $\mathcal{N} \leftarrow$ Set of all applications, $\mathcal{T} \leftarrow$ Internet data budget plan period, and $B \leftarrow$ Total Budget

OUTPUT: Reserve budget $r_{i,t}$, $\forall i \in \mathcal{N}$ at time cycle t

```

1: while the system starts do do
2:   for all data plan periods  $T_k$  do
3:     for each cycles  $t \in \mathcal{T}$  do
4:       Calculate  $B_1$  using Eq. [6]
5:       Calculate  $B_2$  using Eq. [7]
6:       Adjust  $M$  with  $|\mathcal{N}|$ ;
7:        $M \leftarrow \left\lceil \frac{\kappa}{|\mathcal{N}|} \right\rceil$ ;
8:       for each application  $i \in \mathcal{N}$  do
9:         Predict  $b_{i,t}$  using Eq. [13]
10:        Set  $r_{i,t}$  using Eq. [8]
11:      end for
12:    end for
13:  end for
14: end while

```

estimation. This solution allows us to make a tradeoff in between the bandwidth allocation accuracy and the time complexity.

The complexity of Algorithm 1 is quite straightforward to follow. The statements 1 ~ 10 are enclosed in a loop that iterates $|\mathcal{T}|$ times, 6 ~ 9 has another loop that iterates $|\mathcal{N}|$ times, and the statement 7 has a prediction methodology that has $O(M)$ complexity. The rest of the statements have constant unit time complexities. Therefore, the overall computational complexity of the algorithm is $O(|\mathcal{T}| \times |\mathcal{N}| \times M)$.

6 Performance evaluation

In this section, we evaluate performance of the proposed optimal Internet data budget management (O-IDM) policy, heuristic Internet data budget management (H-IDM) policy, dynamic bandwidth allocation (DBA) [22], and an online heuristic algorithm [18]. We implemented both O-IDM and H-IDM algorithms on an Android device.

6.1 Experimental setup

Now, we present our testbed implementation environment and dataset used in the evaluation. To evaluate our proposed scheme, we conducted a study involving 8 participants, each having a smartphone running a number of applications on a fixed budget mobile Internet connection. We developed Android applications for each of the studied data budget allocation algorithms to be installed on the smartphones. We assume that the user applications are heterogeneous, in terms of time-sensitivity, reliability, bandwidth requirements, etc. The number of applications active on a mobile device and user's monthly

data plan vary from a wide range of 10 ~ 60 and 1 ~ 8 GB, respectively. The usage arrival and departure of applications at smartphones are exponentially distributed. As a result, the duration for which an active application varies. We also assumed that the mobile user uses both Internet data plan and Wi-Fi, which is a typical user setup. The total data plan period is assumed to be 720 h (i.e., 30 days) and the budget allocation algorithm execution time interval Δ is chosen 30 ~ 50 min. We stored the user's application usage statistics in a log file and plotted the graph points based on real-time dataset. Each graph data point represents the average value of the results from 10 or more experiments; thus, the performance graphs depict stable behavior of the studied systems.

6.2 Performance metrics

We used the following two metrics to evaluate the performance of the studied Internet data budget allocation algorithms.

- *Average resource utilization:* Equation 1 defines the resource utilization of each application in all time cycles. Our aim is to upgrade system performance and user experience by maximizing utilization and minimizing penalty. The average resource utilization is measured for individual applications and then the average is taken for all user applications running on the smartphone.
- *Average penalty:* A user application is penalized when the allocated data budget is insufficient compared to its requirement. Thus, we measure penalty as the percentage of applications that could not be run due to shortage of bandwidth during the experiments. The average is taken for all time periods and all applications.

6.3 Simulation results

6.3.1 Impacts of increasing number of applications

As shown in Fig. 3a, the average resource utilization linearly decreases with the number of applications for all studied budget allocation algorithms due to under-provisioning penalty. The proposed H-IDM policy offers as high as almost 80 % performance for higher number of applications compared to online heuristic and DBA algorithms. This happens because of its higher capability in accommodating diverse applications with different priorities and reducing computational overhead by dynamically adjusting the bandwidth allocation to the applications following their historical usage pattern. The performance of O-IDM is slightly lower than H-IDM because it requires a significant amount of time to update the matrix for a large number of applications. Thus, some applications are denied sufficient bandwidth.

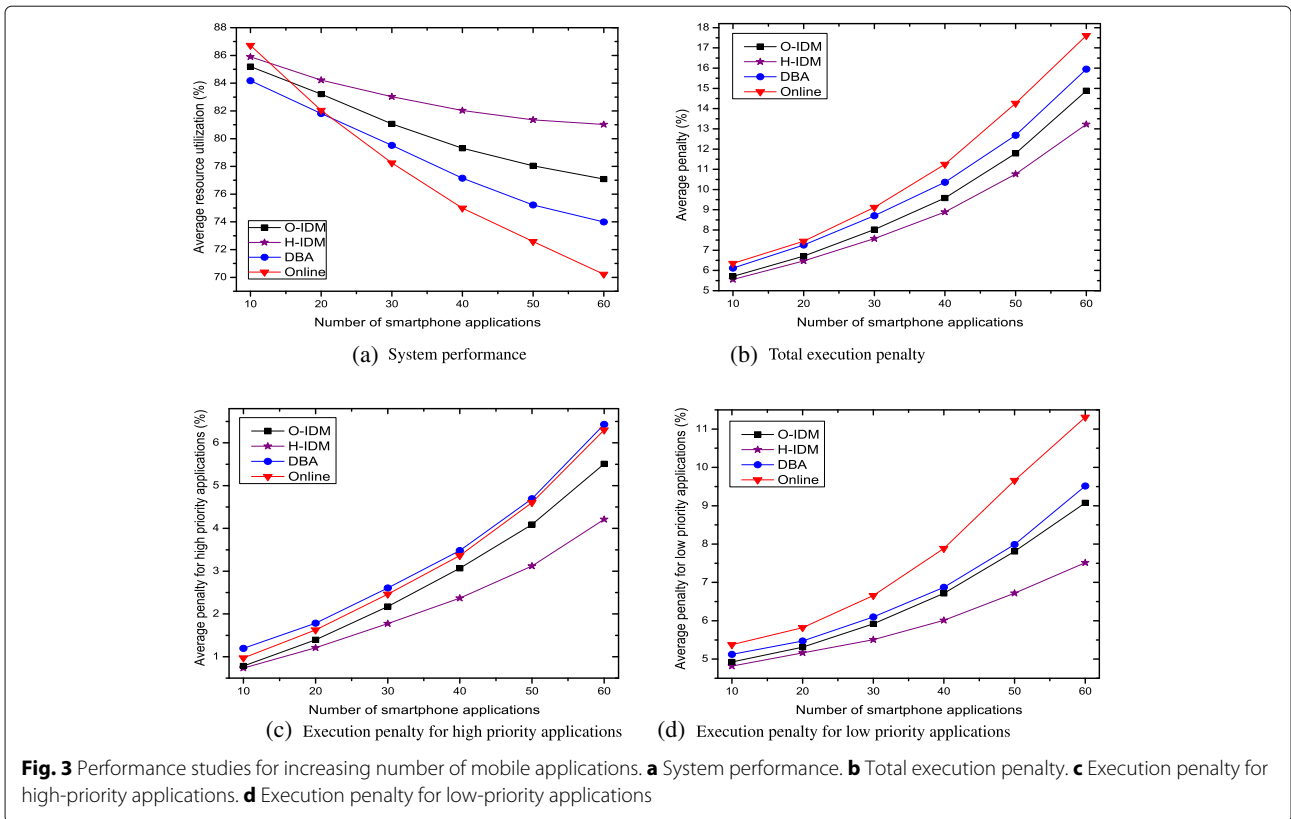


Figure 3b depicts that the percentage of penalty increases with increasing number of applications but the penalties offered by O-IDM and H-IDM algorithms remain relatively low, in comparison to online heuristic and DBA algorithms. This is because flexible budget is allocated to high-priority or sensitive applications repeatedly so that these applications can complete their tasks. On the other hand, for online heuristic and DBA algorithms, the average penalty percentage is relatively high because the applications face under-provisioning penalty due to the weakness of budget distribution policy.

Figure 3c depicts that the percentage of penalty increases with increasing number of applications but the penalty offered by H-IDM remains relatively low, in comparison to the DBA algorithm. This is due to the DBA algorithm resetting utilization to zero when under-provisioning penalty occurs. Again, high-priority applications are stuck when the flexible budget is depleted, but in the case of H-IDM, high-priority applications are allocated flexible budget repeatedly according to their requirements.

Figure 3d depicts that the percentage of penalty increases with the increasing number of applications because a fixed amount of budget is allocated among all increased number of applications. Consequently, low-priority applications do not get adequate budget. Low-priority applications also face more penalty than

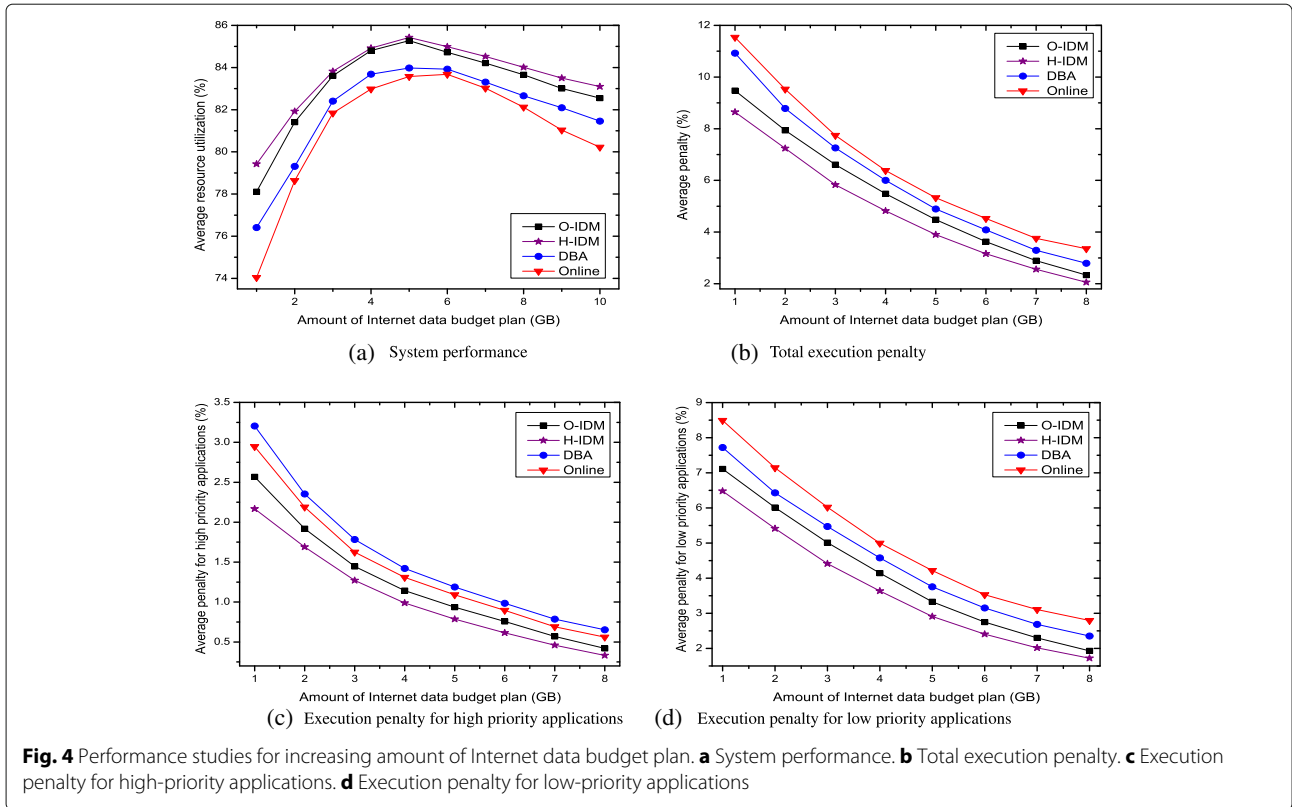
high-priority applications because they have access to flexible budget only once when they run out of reserved budget.

6.3.2 Impacts of varying amounts of data budget plan

Figure 4a shows that the average resource utilization offered by H-IDM grows with larger amount of budget, but it starts decreasing due to over-provisioning penalty when the allocated budget exceeds 7 GB. However, the average performance offered by online heuristic algorithm remains low because of their incapability of proper distribution of budget among applications and cleverly handling over- and under-provisioning penalties. The average resource utilization offered by the H-IDM is relatively higher than that of DBA algorithm since Eq. 12 increases the budget usage ratio and reduces the probability of occurring penalties in the upcoming time cycles.

Figure 4b depicts that under-provisioning penalty decreases with larger amount of allocated budget as the applications are allocated adequate amount of budget. However, the percentage of penalty offered by online heuristic algorithm is higher as it has not defined a budget allocation technique for each application.

Figure 4c depicts that average penalty decreases with larger amount of allocated budget as the high-priority applications are allocated adequate amount of budget according to their bandwidth requirement. The average



penalty decreases due to the decline in under-provisioning penalty.

Figure 4d depicts that under-provisioning penalty declines with larger amount of allocated budget because adequate amount of budget is allocated to low-priority applications. However, penalty percentage offered by online heuristic algorithm is relatively higher because bandwidth is not distributed among all applications; rather, all the sensitive applications use reserved budget and non-sensitive applications use flexible budget. When sensitive applications run out of reserved budget, they take help from flexible budget. Consequently, low-priority applications are deprived of sufficient bandwidth.

6.3.3 Impacts of varying time intervals

Figure 5a depicts that average resource utilization decreases with increasing time intervals for all the three graphs because longer time interval for algorithm execution leads to under-provisioning penalty. However, the proposed H-IDM graph has a higher average resource utilization than both online heuristic and DBA algorithms. This is because when a high-priority application runs out of its reserved budget, it is allocated flexible budget repeatedly.

Figure 5b–d depicts that the average penalty increases with the increasing time intervals because applications

face under-provisioning penalty for longer time intervals during algorithm execution as the applications get stuck due to insufficient budget. On the contrary, if time interval for algorithm execution is too short, then over-provisioning penalty occurs. Thus, we need a tradeoff between over- and under-provisioning penalties; 30 – 50-min interval gives the best result in this case.

6.3.4 Computational time analysis

Estimated budget usage needs to be calculated based on the application’s historical usage pattern for determining appropriate budget reservation of each application. The estimated budget usage of an application is measured as a weighted average of last M time cycles. For the O-IDM algorithm, the computational time increases significantly if $|\mathcal{N}|$ and M are large enough. In Fig. 6, we compare the computational time of all algorithms with increasing number of applications, ranging from 10 ~ 60. From the graph, it is clear that when the number of applications is 40 and above, the computational time of O-IDM increases exponentially because of its checking on all possible combinations required for updating the matrix for historical usage statistics. The computational overhead of heuristic algorithm is noticeably minimized for increasing number of applications as the value of M is adjusted time to time with $|\mathcal{N}|$.

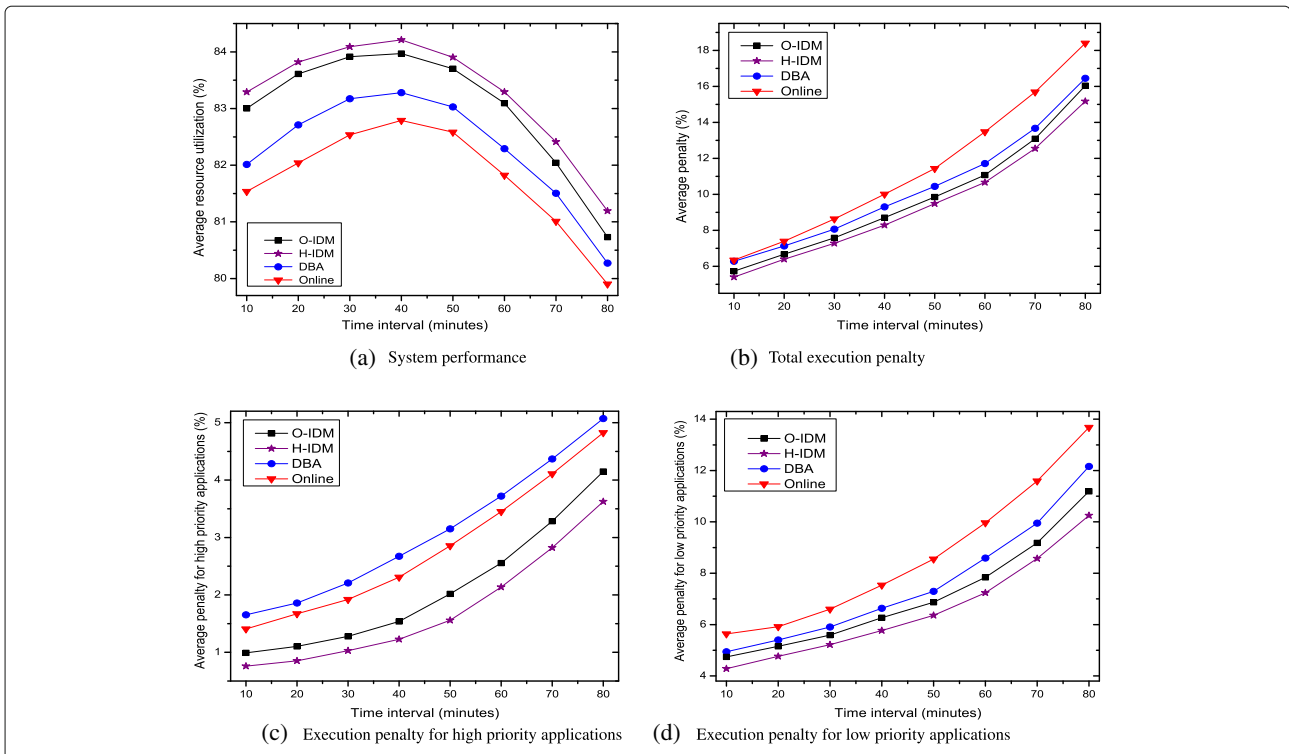


Fig. 5 Performance studies for varying time intervals. **a** System performance. **b** Total execution penalty. **c** Execution penalty for high-priority applications. **d** Execution penalty for low-priority applications

7 Conclusions

The popularity of mobile devices and applications will necessitate the development of policies for effective Internet data budget allocation in order to avoid incurring excessive charges.

In this paper, we developed policies for Internet data budget allocation to heterogeneous applications running on mobile devices, specifically smartphones. The proposed optimal Internet data budget management (O-IDM) maximizes the resource utilization while

minimizing over- and under-provisioning for all user applications. To prevent significant computational overhead and excessive delay in decision making when allocating data budget to each application in the O-IDM system, we developed an alternate heuristic solution to the IDM problem. We relied upon the historical data budget usage behavior of the user applications in our recommendation of future Internet data budget plan following ARIMA-based resource prediction analysis. The test-bed experimental results demonstrate that our proposed Internet data management policies benefit mobile device user by increasing the resource utilization and decreasing the average percentage of penalty.

Acknowledgements

The authors would like to extend their sincere appreciation to the Deanship of Scientific Research at King Saud University for its funding of this research through the research group project no. RGP-281.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Green Networking Research Group, Department of Computer Science and Engineering, University of Dhaka, Dhaka 1000, Bangladesh. ²College of Computer and Information Sciences, King Saud University, Riyadh, Kingdom of Saudi Arabia. ³Department of Information Systems and Cyber Security, University of Texas at San Antonio, San Antonio, USA. ⁴School of Information Technology and Mathematical Sciences, University of South Australia, Adelaide, Australia.

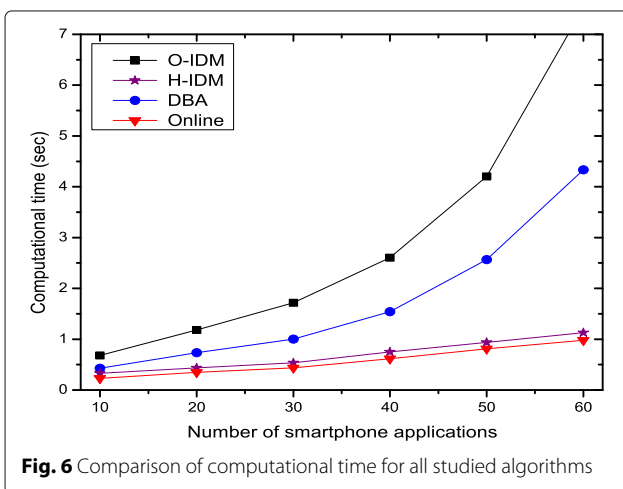


Fig. 6 Comparison of computational time for all studied algorithms

Received: 24 June 2016 Accepted: 15 September 2016

Published online: 22 September 2016

References

1. A Azfar, KKR Choo, L Liu, Android mobile VoIP apps: a survey and examination of their security and privacy. *Electron. Commer. Res.* **16**(1), 73–111 (2016). Available from <http://dx.doi.org/10.1007/s10660-015-9208-1>
2. TD Buskirk, C Andrus, in *Journal Survey Practice*. Smart surveys for smart phones: exploring various approaches for conducting online mobile surveys via smartphones, American Association for Public Opinion Research, (2012), pp. 1–11
3. A Smith, in *The Smartphone Difference*. U.S. smartphone use in 2015, Pew Research Center, USA (Pew Research Center, 2015). Available from <http://www.pewinternet.org/2015/04/01/us-smartphone-use-in-2015/>
4. MR Rahimi, J Ren, CH Liu, AV Vasilakos, N Venkatasubramanian, Mobile cloud computing: a survey, state of art and future directions. *Mobile Netw. Appl.* **19**(2), 133–143 (2014)
5. M Yang, Y Li, D Jin, L Zeng, X Wu, AV Vasilakos, Software-defined and virtualized future mobile and wireless networks: a survey. *Mobile Netw. Appl.* **20**(1), 4–18 (2015)
6. A Gutierrez, RG Dreslinski, TF Wenisch, T Mudge, A Saidi, C Emmons, et al, in *Workload Characterization (IISWC) 2011 IEEE International Symposium on*. Full-system analysis and characterization of interactive smartphone applications, USA, (2011), pp. 81–90
7. S Pokharel, KKR Choo, J Liu, Can Android VoIP voice conversations be decoded? I can eavesdrop on your Android VoIP communication. *Concurrency and Computation: Practice and Experience* (2016). Available from <http://dx.doi.org/10.1002/cpe.3845>
8. MR Rahimi, N Venkatasubramanian, S Mehrotra, AV Vasilakos, in *Utility and Cloud Computing (UCC) 2012 IEEE Fifth International Conference on*. MAPCloud: mobile applications on an elastic and scalable 2-tier cloud architecture, USA, (2012), pp. 83–90
9. J Chon, H Cha, LifeMap: a smartphone-based context provider for location-based services. *IEEE Pervasive Comput.* **10**(2), 58–67 (2011)
10. CC Chen, TC Huang, JJ Park, NY Yen, Real-time smartphone sensing and recommendations towards context-awareness shopping. *Multimedia Syst.* **21**(1), 61–72 (2015). Available from <http://dblp.uni-trier.de/db/journals/mms/mms21.html#ChenHPY15>
11. KK Rachuri, C Efstratiou, I Leontiadis, C Mascolo, PJ Rentfrow, Smartphone sensing offloading for efficiently supporting social sensing applications. *Pervasive Mobile Comput.* (2013). Available from <http://dx.doi.org/10.1016/j.pmcj.2013.10.005>
12. A Oulasvirta, T Rattenbury, L Ma, E Raita, Habits make smartphone use more pervasive. *Personal Ubiquitous Comput.* **16**(1), 105–114 (2012). Available from <http://dx.doi.org/10.1007/s00779-011-0412-2>
13. A Rahmati, L Zhong, Studying smartphone usage: lessons from a four-month field study. *IEEE Trans. Mobile Comput.* **12**(7), 1417–1427 (2013). Available from <http://dx.doi.org/10.1109/tmc.2012.127>
14. D Zhang, D Zhang, H Xiong, CH Hsu, AV Vasilakos, BASA: building mobile ad-hoc social networks on top of android. *IEEE Netw.* **28**(1), 4–9 (2014)
15. TMT Do, J Blom, D Gatica-Perez, in *Proceedings of the 13th International Conference on Multimodal Interfaces. ICMI '11*. Smartphone usage in the wild: a large-scale analysis of applications and context (ACM, New York, 2011), pp. 353–360. Available from <http://doi.acm.org/10.1145/2070481.2070550>
16. H Falaki, R Mahajan, S Kandula, D Lymberopoulos, R Govindan, D Estrin, in *Proceedings of the 8th International Conference on Mobile Systems, Applications, and Services. MobiSys '10*. Diversity in smartphone usage (ACM, New York, 2010), pp. 179–194. Available from <http://doi.acm.org/10.1145/1814433.1814453>
17. O Amft, P Lukowicz, From backpacks to smartphones: past, present, and future of wearable computers. *IEEE Pervasive Comput.* **8**(3), 8–13 (2009)
18. H Liu, S Hu, W Zheng, Z Xie, S Wang, P Hui, et al, in *INFOCOM, 2013 Proceedings IEEE*. Efficient 3G budget utilization in mobile participatory sensing applications, Turin, Italy, (2013), pp. 1411–1419
19. C Luo, H Ji, Y Li, in *2009 IEEE Wireless Communications and Networking Conference*. Utility-based multi-service bandwidth allocation in the 4G heterogeneous wireless access networks, Hungary, (2009), pp. 1–5
20. ME Khoda, MA Razzaque, A Almogren, MM Hassan, A Alamri, A Alelaiwi, Efficient computation offloading decision in mobile cloud computing over 5G network. *Mobile Netw. Appl.* 1–16 (2016). Available from <http://dx.doi.org/10.1007/s11036-016-0688-6>
21. MM Hassan, Cost-effective resource provisioning for multimedia cloud-based e-Health systems. *Multimedia Tools Appl.* **74**(14), 5225–5241 (2015). Available from <http://dx.doi.org/10.1007/s11042-014-2040-0>
22. S Rahman, AA Prima, MA Razzaque, in *2016 International Conference on 771 Networking Systems and Security (NSysS)*. Optimal allocation of 3G budget 772 for smartphones running heterogeneous applications, Dhaka, Bangladesh, vol. 773, (2016), pp. 1–6
23. J Chen, Q Liang, J Wang, in *The Proceedings of the Third International Conference on Communications, Signal Processing, and Systems Part III*. Bandwidth allocation based on personality traits on smartphone usage and channel condition, Springer, vol. 777, (2015), pp. 273–282
24. HM Huang, YH Wang, Bandwidth management method for heterogeneous wireless network. *WTOC.* **7**(4), 267–276 (2008). Available from <http://dl.acm.org/citation.cfm?id=1455946.1455953>
25. MR Rahimi, N Venkatasubramanian, AV Vasilakos, in *2013 IEEE Sixth International Conference on Cloud Computing*. MuSIC: mobility-aware optimal service allocation in mobile cloud computing, USA, (2013), pp. 75–82
26. J Wan, J Liu, Z Shao, AV Vasilakos, M Imran, K Zhou, Mobile crowd sensing for traffic prediction in internet of vehicles. *Sensors.* **16**(1), 88 (2016)
27. Z Feng, Y Zhu, Q Zhang, LM Ni, AV Vasilakos, in *IEEE INFOCOM 2014 - IEEE Conference on Computer Communications*. TRAC: truthful auction for location-aware collaborative sensing in mobile crowdsourcing, Toronto, Canada, vol. 789, (2014), pp. 1231–1239
28. H Cai, Y Gu, A Vasilakos, B Xu, J Zhou, Model-Driven Development Patterns for Mobile Services in Cloud of Things. *IEEE Trans. Cloud Comput.* **PP**(99), 1–1 (2016). doi:10.1109/TCC.2016.252600
29. W Fang, Y Li, H Zhang, N Xiong, J Lai, AV Vasilakos, On the throughput-energy tradeoff for data transmission between cloud and mobile devices. *Inf. Sci.* **283**, 79–93 (2014)
30. L Mashayekhy, MM Nejad, D Grosu, AV Vasilakos, An online mechanism for resource allocation and pricing in clouds. *IEEE Trans. Comput.* **65**(4), 1172–1184 (2016)
31. D López-Pérez, X Chu, AV Vasilakos, H Claussen, On distributed and coordinated resource allocation for interference mitigation in self-organizing LTE networks. *IEEE/ACM Trans. Netw.* **21**(4), 1145–1158 (2013)
32. D Lopez-Perez, X Chu, AV Vasilakos, H Claussen, Power minimization based resource allocation for interference mitigation in OFDMA femtocell networks. *IEEE J. Selected Areas Commun.* **32**(2), 333–344 (2014)
33. T Liu, Y Zhu, Q Zhang, A Vasilakos, Stochastic optimal control for participatory sensing systems with heterogenous requests. *IEEE Trans. Comput.* **65**(5), 1619–1631 (2016)
34. S Chaisiri, BS Lee, D Niyato, Optimization of resource provisioning cost in cloud computing. *IEEE Trans. Serv. Comput.* **5**(2), 164–177 (2012)
35. Z Dong, L Kong, P Cheng, L He, Y Gu, L Fang, et al, in *2014 Eleventh Annual IEEE International Conference on Sensing, Communication, and Networking (SECON)*. REPC: Reliable and efficient participatory computing for mobile devices, Singapore, (2014), pp. 257–265
36. F Martignon, S Paris, I Filippini, L Chen, A Capone, Efficient and truthful bandwidth allocation in wireless mesh community networks. *IEEE/ACM Trans. Netw.* **23**(1), 161–174 (2015)
37. J Guo, F Liu, X Huang, JCS Lui, M Hu, Q Gao, et al, in *IEEE INFOCOM 2014 IEEE Conference on Computer Communications*. On efficient bandwidth allocation for traffic variability in datacenters, Toronto, Canada, (2014), pp. 1572–1580
38. ALL SOLVERS FOR A Mathematical Programming Language (AMPL). <http://ampl.com/products/solvers/all-solvers-for-ampl/>. Accessed 06 March 2016
39. S Floyd, M Handley, J Padhye, J Widmer, in *Proceedings of the Conference on Applications, Technologies, Architectures, and Protocols for Computer Communication. SIGCOMM '00*. Equation-based congestion control for Unicast applications (ACM, New York, 2000), pp. 43–56. Available from <http://doi.acm.org/10.1145/347059.347397>
40. A Razzaque, CS Hong, S Lee. Autonomous traffic engineering for boosting application fidelity in wireless sensor networks, vol. E93-B (*IEICE Trans. Commun.*, 2010), pp. 2990–3003
41. Autoregressive integrated moving average models (ARIMA). <http://www.forecastingsolutions.com/arima.html>. Accessed 16 Jan 2016

42. GEP Box, G Jenkins, *Time Series Analysis, Forecasting and Control*. (Holden-Day, San Francisco, 1970)
43. X Jin, Y Dong, J Wu, J Wang, in *Information Science and Management Engineering (ISME) 2010 International Conference of*. An improved combined forecasting method for electric power load based on autoregressive integrated moving average model, China, vol. 2, (2010), pp. 476–480
44. N Citroen, M Ouassaid, M Maaroufi, in *Electrical and Information Technologies (ICEIT) 2015 International Conference on*. Long term electricity demand forecasting using autoregressive integrated moving average model: case study of Morocco, Morocco, (2015), pp. 59–64

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com
