Prediction-based Personalized Offloading of Cellular Traffic Through WiFi Networks

Suyeon Kim, Yohan Chon, Seokjun Lee, and Hojung Cha
Department of Computer Science, Yonsei University, Seoul, Korea
{sykim, yohan, sjlee }@cs.yonsei.ac.kr, hjcha@yonsei.ac.kr

Abstract—**Mobile data offloading through WiFi is an essential requirement to reduce cellular network traffic. While extensive attempts have been made at mobile data offloading, previous studies have rarely addressed practical issues, such as dealing with diverse user contexts. In this paper, we propose a personalized data offloading scheme to provide maximum throughput within the cellular budget in daily life. We propose an adaptive policy that considers a user's mobility patterns, cellular budget, and network usage for applications. The proposed system employs an adaptive model to predict the throughput of WiFi APs and the network usage of smartphones. Among the three types of predictor model (i.e., spatial, temporal, and spatio-temporal), the system automatically chooses the optimal model for each mobile user without user intervention. The experimental results from 10 mobile users show that the proposed system provides 29% higher throughput than previous systems and minimizes extra data charges.**

Keywords-cellular data offloading; LTE networks; WiFi networks

I. INTRODUCTION

The proliferation of smartphones has caused an explosion of cellular network traffic, and traffic increases by 78% every year [1]. This trend continues to accelerate due to diverse smartphone applications, such as social networking, instant messaging, and video streaming. To reduce cellular network traffic, network operators charge an extra fee on data packets beyond the data plan. The research community has proposed various data offloading policies through WiFi networks [2, 4, 5, 6, 7, 8, 9, 10, 13]. However, prior work on offloading policies has limited in handling diverse contexts in real deployments. Commercial smartphones simply prefer WiFi networks to cellular networks although its network quality is poor, and mobile users should manually switch from WiFi networks to cellular networks in practice.

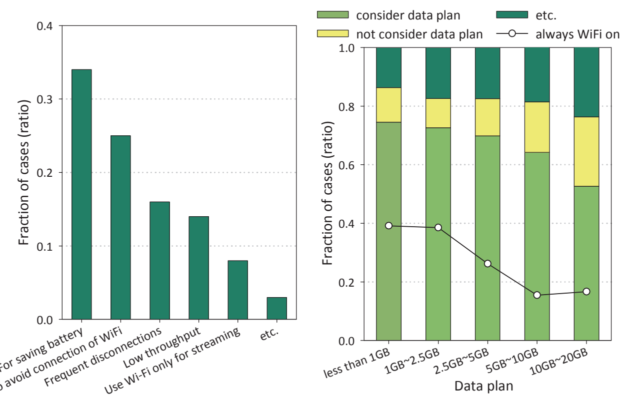
The main question for offloading is which network should be used for data traffic—the cellular network or WiFi network? Traditional approaches always prefer WiFi networks to cellular networks if WiFi is available [2, 4, 5]. However, WiFi networks have relatively low coverage and unstable throughput compared to cellular networks [12]. Thus, works in [6, 7] considered the quality of WiFi networks to deliver maximum throughput to mobile users. They monitored the number of surrounding users or channel status to estimate the quality of WiFi networks. Recently, several studies have considered user-centric information, such as the cellular budget, battery consumption, or the network usage of applications [6, 7, 8, 9, 11]. However, we argue that previous work has not fully considered the diverse contexts of mobile users in real deployment. For example, the cellular budget of mobile users is diverse (e.g., from 100 MB to 10 GB), and they use the cellular network differently according to the time of day. In addition, the throughput of WiFi APs significantly varies according to location and time. Thus, the offloading policy should be personalized for each user to optimize performance of network usage.

In this paper, we propose a personalized WiFi offloading policy that considers diverse user contexts, such as mobility patterns, cellular budgets, and the network usage of applications. Our research goal is to maximize the network throughput of smartphones within monthly cellular budgets. The proposed policy learns the mobility pattern and network quality according to location and time, then automatically switches the smartphone network between cellular network and WiFi network without user instruction. The key technical challenges are (1) predicting the throughput and the quality of WiFi networks in daily life and (2) predicting the network usage of diverse applications in smartphones.

The proposed system comprises three components: a throughput predictor, a network usage predictor, and an offloading scheduler. The throughput predictor predicts the throughput of WiFi APs based on a user's mobility pattern. The network usage predictor predicts the amount of network traffic used in smartphone applications. Both components adaptively use spatial, temporal, or spatio-temporal predictors based on a user's characteristics. The offloading scheduler determines the offloading decision of which network (i.e., Long-term evolution (LTE or 4G) or WiFi) should be used for network traffic. In other words, the proposed system generates a personalized offloading schedule to maximize the network throughput.

The primary contributions of our study are as follows: (1) we consider the case where cellular's throughput is higher than that of WiFi since users could use the networks such as LTE. This is different from previous studies that consider only the 3G networks; (2) we propose an adaptive policy that manages diverse contexts in practice. The proposed policy is transparently operated in smartphones, and mobile users could obtain maximum throughput through WiFi and LTE within their monthly data plans. We have validated the proposed scheme in real deployments for two months.

978-1-4673-8779-8/16/$31.00 ©2016 IEEE



(a) (b)

Fig. 1. (a) The reason of turning off the WiFi; (b) WiFi usage pattern according to the monthly data plan.

II. Preliminary and Motivation

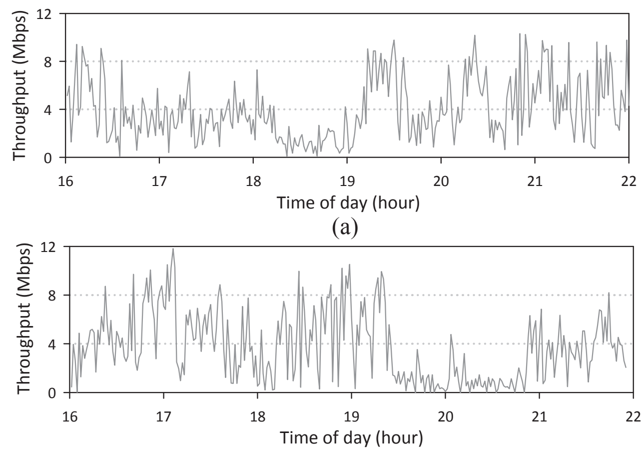
A. *WiFi Usage of Smartphone Users*

To explore problems in WiFi usage, we surveyed 1,110 users in Seoul, Korea about their WiFi usage in smartphones. In Seoul, the WiFi networks are available almost everywhere, even in metro and subway. The ages of the responders were 3% teenagers, 66% twenties, 17% thirties, 7% forties, and 7% fifties. The monthly data plan of the responders were less than 1GB (19%), 1GB-2.5GB (24%), 2.5GB-5GB (23%), 5GB-10GB (20%), and unlimited (10%). The survey question was how do you normally set WiFi on your smartphone—always on or always off—and why? The responses show that only 30% of responders normally set WiFi as always on. The rest of the responders said that they turn off WiFi to save battery (34%) or to prevent the autonomous connection of WiFi (55%). The latter problem is caused by the low quality of WiFi networks such as low throughput, frequent disconnections, and connection but unavailable Internet use, as shown in Fig. 1(a). The result suggests that more than half of the mobile users felt inconvenient with current WiFi usage, and they set WiFi off and only turn it on when they are within the coverage of high-quality WiFi APs. Fig. 1(b) shows WiFi usage patterns according to monthly data plans. Intuitively, users with low-data plans tend to set WiFi on and consider their data plans to avoid extra data charges. However, despite their low-data plans (i.e., less than 1GB), only 40% of users set WiFi on in daily life due to the low quality of WiFi networks. This result suggests that the quality of current WiFi networks falls far short of users' demands. Thus, the offloading policy should consider the low quality of WiFi networks to preserve quality-of-service in Internet use.

The survey results show that mobile users carefully consider their data plan, but the quality of WiFi networks is problematic in practice. This finding indicates that an offloading policy should filter out the use of low quality WiFi networks while preventing extra data charges.

B. *Characteristics of Throughput in WiFi Networks*

Based on the survey, we noticed that the mobile users generally felt inconvenient when using WiFi. We then considered the actual quality of WiFi networks in practice. To investigate the quality of WiFi networks in real deployments, we implemented an application to measure the throughput of WiFi APs every minute and deployed experimental smartphones in a subway and a cafeteria.



(a)

(b)

Fig. 2. The trace of throughput of WiFi APs (a) in the subway and (b) in cafeteria.

Fig. 2 shows the trace of throughput measured during a day. In the subway, the throughput is about 3.9Mbps on average, but the throughput significantly varies over short durations (i.e., 1 minute) as shown in Fig. 2(a). In particular, the throughput decreased to 1.58Mbps at commute time (6:00–7:00 p.m.) when users are crowded in the subway. This pattern was similar in the cafeteria as illustrated in Fig. 2(b). The average throughput was approximately 4.1Mbps, but the throughput was very low around dinner time (7:30–9:00 p.m.). The results indicate that the throughput of WiFi APs fluctuates widely according to the time of day, and the quality of WiFi networks is closely linked to the number of surrounding users. Based on this observation, we predict the throughput of WiFi APs in the spatio-temporal aspect in the offloading policy to handle the variation of throughput.

III. Related Work

Offloading policies for mobile data has been extensively studied in the research community. Table I summaries the characteristics of previous work. Traditional approaches focused on maximizing the volume of offloading data since the throughput of WiFi is significantly higher than 3G networks [2, 4, 5]. Balasubramanian et al. [2] predicts the WiFi availability in order to delay data traffic when a user is moving. Han et al. [4] offloaded data by relaying data between mobile users, and Yetim et al. [5] divided the data into several buckets to consider the low coverage of WiFi networks. However, these studies did not consider temporal variations in the throughput of APs, and they assume the use of a 3G network, which has a lower throughput than WiFi networks. An offloading policy using LTE networks should be designed differently to previous studies since the LTE networks have higher throughput than WiFi networks.

Studies in [6, 7, 8, 9] considered various contexts, such as cellular budget, energy consumption, and application types

TABLE I. Summary of related systems for WiFi offloading

System	Cellular Network Type	WiFi Offloading Type	Policy Goal	Considered Factor			Prediction	
				Through-put	*Cellular Budget*	*Energy Consumption*	*Access Point*	*Network Usage*
Balasubramanian et al. [2]	3G	Delayed	Maximize offloading	AP-dependent	-	-	Spatial	-
Lee et al. [3]	3G	Non-delayed, delayed	Estimate offloading amount	-	-	-	-	-
Han et al. [4]	3G	Delayed	Maximize offloading through opportunistic communications	-	-	-	Spatial	-
Yetim et al. [5]	-	Delayed	Maximize offloading	AP-dependent	-	-	Spatial	-
Multi-Nets [6]	3G	-	1. Minimize energy 2. Maximize offload 3. Maximize throughput	Signal-dependent	-	-	-	-
AMUSE [7]	3G	Semi-delayed	Maximize throughput within certain budget	AP-dependent	Constant budget	5 types	Spatial	Type-dependent
Liu et al. [8]	3G	-	Maximize data transmission within certain budget	-	Constant budget	-	Temporal	-
Lee et al. [9]	LTE	-	Maximize file transmission within certain budget	AP-dependent	Constant budget	-	-	-
Proposed method	**LTE**	**Non-delayed**	**Maximize throughput within certain budget**	**AP & time-dependent**	**Weighted budget on day of month**	**-**	**Spatio-Temporal**	**AP & time-dependent**

as well as the throughput of networks. Multi-Nets [6] proposed a seamless switching policy between a 3G network and a WiFi network to maximize the throughput with minimum energy consumption. AMUSE [7] considered the throughput, cellular budget, and application type in order to delay traffic to maximize the throughput gain. Liu et al. [8] used WiFi offloading to maximize the amount of collected data in participatory sensing within the cellular budget. Lee et al. [9] considered the network throughput, cellular budget, and energy consumption in order to maximize the amount of data in file transfers. However, these studies have room for improvement since they did not consider diverse contexts in real deployments. For example, previous studies used a constant value for cellular budgets [7, 8, 9] or energy consumption that is not realistic in practices. Thus, our scheme handles diverse usage patterns of cellular budgets to create a personalized policy by using a weighted cellular budget on the LTE network.

Traditional approaches [2, 4, 5, 14] assumed the delayed offloading. Recent studies [3, 7] did not delay the offloading, or delayed the network usage of only a few applications (e.g., mail sync) since the delay in network traffic significantly decreases the user experience. For example, when a user launches web browser, it is not practical to delay webpage loading until a WiFi AP is available. In contrast, the proposed policy would not assume the delay of network traffic to preserve user experience on phone usage.

The key factor of the offloading policy is the performance of predictions of throughput. Previous work shows a low accuracy of the predictions [2, 4, 5, 7] since they did not consider the variation of throughput in the temporal aspect, as illustrated in Table I. Although a few prediction method used temporal features [8, 15], they did not reflect spatial factors. In this context, we designed an offloading policy that uses the spatio-temporal model for throughput predictions. We also consider the spatio-temporal pattern of network usage for personalized offloading decisions.

IV. WIFI OFFLOADING SYSTEM

The goal of the proposed system is to adaptively offload mobile data through a WiFi network to provide the maximum throughput within a monthly data plan. We designed a personalized offloading policy that automatically switches the network between an LTE network and a WiFi network. Our policy would not delay the traffic to preserve user experience on phone usage. The system uses the historical pattern of the throughput of WiFi APs, network usage, and the monthly cellular budget. Based on the historical pattern, the policy predicts the expected throughput with application usage, and then determines which network (i.e., LTE or WiFi) should be used for globally maximum throughput in a month.

Fig. 3 shows the architecture of the proposed system. The proposed system runs as a background service on a smartphone. The system first learns the throughput and application usage of the user. When the user associates a WiFi AP, the system estimates the SSID, BSSID, RSSI, downlink throughput, and the latency of the WiFi AP with a timestamp. In addition, the system traces the network usage of smartphone applications. The collected data is then uploaded to the server. Then, the throughput predictor predicts the expected daily throughput (Section IV.A), and the network usage predictors predict the network usage of applications (Section IV.B). Finally, based on the predicted throughput and network usage, the offloading scheduler performs offloading decisions for optimal network usage (Section IV.C).

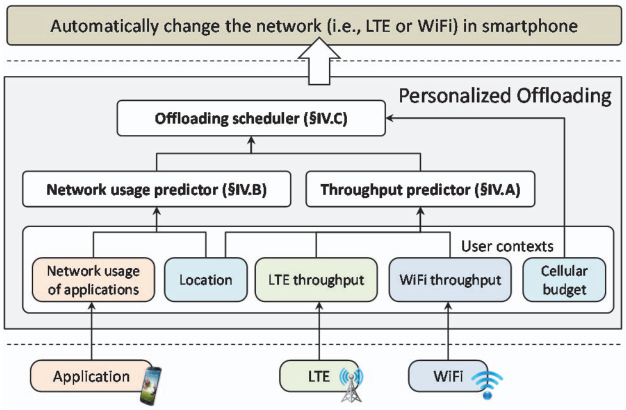


Fig. 3. System architecture of the proposed offloading system

A. Throughput Predictor

The throughput predictor predicts the throughput of WiFi APs that a user encounters daily. WiFi throughput is related to location (i.e., APs) and the number of concurrent users in the frequency band. However, the estimation of concurrent users is challenging in the device side. Thus, to infer the quality of WiFi network indirectly, we use the time-of-day in the predictor rather than the direct estimation of concurrent users. The basic concept of the throughput predictor is that the throughput of a WiFi AP varies according to the location and time of day. For example, the throughput of an AP in an office is different from that of an AP in a subway, and the throughput of the AP in a subway varies according to the time of day. The system should correctly infer the expected throughput of WiFi APs to choose the optimal network, either LTE or WiFi, for data offloading.

We used the spatio-temporal model for the throughput predictor since the set of used WiFi APs is tightly related to the mobility pattern of mobile users and the throughput of an AP varies according to the time of day. We made use of a simple Markov predictor since previous work [16] observed that the complex predictors about mobility pattern were at best only negligibly better than the simple Markov predictor in practice. The reason is that a user visits a certain routine pattern, and the throughput is strongly related to places a user visit. Let $AP_t = \{ap^1, ap^2, \ldots, ap^n\}$ be the WiFi AP a user may encounter and $p_t(ap^i)$ be the probability a user may associate ap^i at time t, then the expected throughput th_t a user may obtain at time t is defined as:

$$th_t = \sum_{i=1}^{n}(p_t(ap^i) \times th^i),$$

where th^i is the average throughput of WiFi ap^i at time t. In other words, the expected throughput is the average throughput of WiFi APs that would be encountered at specific time. We used a simple Markov predictor [17] to calculate p_t. Consider the association history of APs $H = (ap_1, t_1, d_1), \cdots, (ap_m, t_m, d_m)$, in which t_i is the association time and d_i is the connection duration at WiFi AP ap_i. From H, we extract the association history $A = ap_1, ap_2, \cdots, ap_n$ in which ap_1, ap_2 indicate that a user associated ap_2 after using ap_1. From A, the recent k-associated APs is $A(n-k+1, n) = ap_{n-k+1}, \cdots, ap_n$. Then, the order-k (or O(k)) Markov predictor generates the probability p_t at the current AP ap_i, defined as $p_{ij} = \Pr\left(ap_{x+1} = ap_j \middle| A(x-k+1, x) = ap_{i-k+1}, \cdots, ap_{i-1}, ap_i\right)$. To predict the overall throughput in a day, we generate the discrete histogram distribution R from th_t. Based on th_t, we discretize a day into time buckets (i.e., 15 minutes), and th_t at each time bucket is estimated by historical throughput data. R is used as the reward function in the offloading scheduler. Based on the generated function, the system predicts the throughput of each 15-minute bucket for the next 24 hours.

In case of throughput of cellular network, we set the LTE's throughput higher than that of WiFi since the estimation of LTE's throughput in daily life is practically limited due to the consumption of LTE data budget. In other words, users should consume their data budget for estimating the throughput of LTE. The use of higher throughput of LTE network is feasible since LTE's throughput is practically higher than that of WiFi in most places as we presented in our preliminary experiments.

B. Network Usage Predictor

The network usage predictor predicts the network usage of applications on a smartphone. The prediction is challenging since mobile users show diverse patterns of application usage. For example, the same application may be used by different ways due to different paths of interactions or duration of interactions. Considering such unknown variables is indeed challenging. Thus, we focus on the prediction of data usage according to location and time-of-day based on the finding in the previous work [18]. They showed that typical users used only 9.8 apps for 90% of their usage time. In other words, the usage pattern of application in each user showed a trivial variation. Thus, we use an adaptive model to handle the variation of usage patterns in each user according to location and time-of-day.

The basic assumption of the network usage predictor is that the network usage pattern is dependent on location and/or time of day. For example, user A may use a high volume of network traffic at home, but user B may use more network traffic during his or her commute. In these cases, the network usage of user A is related to location, but the usage of user B is dependent on the time of day. In addition, a mobile user may exhibit different patterns of network usage (e.g., high traffic during lunchtime and low traffic in the afternoon) although the user is stationary (e.g., in the office). Thus, we designed a network usage predictor to adaptively choose an optimal predictor according to the characteristics of individual users.

We used three types of model: a location-dependent model, a time-dependent model, and a spatio-temporal model. The location-dependent model considers that the network usage of applications is related to location. For example, a user tends to use a large amount of network usage at home or in the office. Let $p(l^i)$ be the probability that a user may stay at location l^i, $p(app^i)$ the probability that a user may use application i, and $u(app^i)$ the expected network usage of application i, then the expected network usage at time t is defined as:

$$U_t^a = \sum_{i=1}^{n}\{p(l^i) \times \sum_{j=1}^{m}\{p(app^j) \times u(app^j)\}\}.$$

The time-dependent model assumes that the network usage of an application is related to the time of day. For example, supposed that a user tends to use a high volume of network traffic during commute time or lunch/dinner time. Therefore, let $p_t(app^i)$ be the probability that a user may use application i at time t and $u_t(app^i)$ the network usage of application i at time t, then the expected network usage at time t is defined as:

$$U_t^b = \sum_{i=1}^{n}\{p_t(app^i) \times u_t(app^i)\}.$$

Lastly, the spatio-temporal model considers that the network usage of an application is related to both location and the time of day. For example, when a user is stationary at home, he or she may use a weather application in the morning time and a streaming application in the evening. Let $p_t^i(app^j)$ be the probability that a user may use application j in location i at time t and $u_t^i(app^j)$ the network usage of application j in location i at time t, then the expected network usage at time t is defined as:

$$U_t^c = \sum_{i=1}^{n}\{p(l^i) \times \sum_{j=1}^{m}\{p_t^i(app^j) \times u_t^i(app^j)\}\}.$$

The reason we used three separate models is that the patterns are diverse across all users. The location is an important factor for certain user, whereas both location and time are important for another user. For example, user A mainly visits private places such as home and office every different time-of-day, while user B tends to visit public places such as subway station, cafeteria, and so on. Then, in the case of user A, the location constraint is strongly related to throughput since the throughput of private WiFi networks is relatively constant than that of public ones. The use of time-of-day decreases the performance because the factor is not relevant to throughput. However, in user B's case, both location and time are important since the throughput of public WiFi networks would highly fluctuate according to time of day as we discussed in Section 3.

The system adaptively chooses the optimal model from the three models based on a user's characteristics. To determine the optimal model at day, we calculate the error between the predicted usage of network traffic and the actual usage from day 1 to day $i-1$. Let e^a, e^b, and e^c be the errors of each model U^a, U^b, and U^c, respectively, the optimal model is the model U that has the minimum error from the three models. Similar to the throughput predictor, we generate the discrete histogram distribution C within specific time buckets (i.e., 15 minutes) based on U. C is used as the cost function in the offloading scheduler.

C. Offloading Scheduler

The goal of the offloading scheduler is to maximize the network throughput within a given cellular budget. The key problem is choosing an optimal moment to offload data through a WiFi network. We formulated the offloading scheduler as a Markov decision process (MDP). The MDP is a stochastic process that contains a 4-tuple$(\mathbb{S}, \mathbb{A}, \mathbb{P}, \mathbb{R})$. The finite set of states in which $\mathbb{S} = \{l_t, l_{t+1}, \ldots\}$ is a set of visited locations and action set $\mathbb{A} = \{a_t, a_{t+1}, \ldots\}$ is a set of actions taken on states, i.e., the activation of WiFi on each bucket. $\mathbb{P}_a(l_t, l_{t+1})$ is the transition probability from state l_t to l_{t+1}, when action a is taken. Here, we set $\mathbb{P}$ to a constant probability since the activation of WiFi is independent to a location change. $\mathbb{R} = \{R_{a_t}(l_t, l_{t+1}), R_{a_{t+1}}(l_{t+1}, l_{t+2}), \ldots\}$ is an expected reward at the transition between the states for the action taken. In our case, the reward is the expected throughput of the WiFi AP or the LTE network at time t, as shown in line 10 at Algorithm 1 (i.e., $R = th_{LTE}\ or\ th_t$). Our goal is to maximize the cumulative function of rewards, defined as follows:

Algorithm 1. Main algorithm for an offloading decision

Input: Offloading moment *i*, cellular budget *b/d*
Output: Set of offloading moments derived from *V()*

1: **if** *b/d* < 0 **then** // Boundary condition 1
2: return *invalid*
3: **end if**
4: **if** *t* > *maxTime* **then** // Boundary condition 2
5: return *0*
6: **end if**
7: **if** *V(t, b/d)* is *valid* **then** // Memorization
8: return *V(t, b/d)*
9: **end if**
10: *V(t, b/d) = max(*th_{LTE} *+ V(t+1, b/d- U*$_t$*),* th_t *+ V(t+1, b/d))*
// Use LTE or WiFi
// Cellular budget decreases if LTE is used
11: return *V(t, b/d)*

$$\text{maximize} \sum_{i=0}^{|\mathbb{A}|} R_{a_i}(l_t) \ \text{subject to}\ C \le E\ ,$$

where C is the total network traffic through an LTE network and E is a given cellular budget. We calculated C by using the cost function generated by the network usage predictor. For example, if a user used cellular network at given time t, a user consumed cellular budget by U_t, as shown in line 10 at Algorithm 1. The solution for this problem involves designing an optimal policy π and computing the value function V(·), which is expressed as:

$$\pi = \mathrm{argmax}_a\{\sum_{t=0}^{maxTime} R_{a_i}(l_t) + V_{a_{i+1}}(t, e - U_t)\},$$

where e is the remaining cellular budget, $V(t, e)$ is the optimal throughputs at given state (t, e), U_t is the used network traffic if the action a does not activate WiFi and zero if the action a activates WiFi, and π is a set of planned network usage (e.g., LTE at time 0, WiFi at time 1, WiFi at time 2, and so on). Here, the time bucket is 15 minutes for a unit. $V(t, e)$ produces the offloading schedules, and the state of the function is a set of input parameters (i.e., time and remaining cellular budget). In other words, an optimal offloading schedule is allocated to maximize the throughput of the network within a given cellular budget. Algorithm 1 presents the pseudo-code of the offloading scheduler using the dynamic programming technique. The output of the dynamic programming is the set of offloading moments, and inputs are the remaining cellular budget, predicted throughput as the reward function, and predicted network usage as the cost function. Here, the daily cellular budget is calculated as b/d, where b is the remaining cellular budget and d is the remaining days in the month. The offloading scheduler generates the pre-planned schedule for a day without feedback. The input 'i' in the algorithm is the i-th time bucket in terms of 15 minutes unit. In the usage scenario, the scheduler runs as the background service in smartphone, and it generates the offloading schedule without user intervention. The system could re-schedule the decision based on the user's current location, but the location sensing is beyond the scope of our research.

V. Evaluation

We divided the evaluation into two parts: macro benchmark and micro benchmark. In the macro benchmark, we validated the performance of the three components: the throughput predictor, the network usage predictor, and the offloading scheduler. The micro benchmark was conducted to evaluate the performance of the offloading scheduler depending on the cellular budget and learning period.

A. *Data Collection*

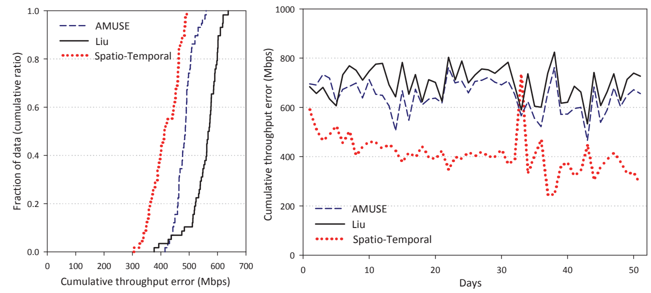
For the evaluation, we collected network and application usage data from 30 smartphone users for two months. We chose the participants by considering various factors such as age, data plan, occupation, and application usage pattern. We implemented a usage monitoring application and provided it to the participants. The application runs in the background and collects various information such as the WiFi fingerprints for place recognition, network usage of each application, and information of connected AP including SSID, BSSID, RSSI, downlink throughput, and latency at an interval of three minutes. To estimate the throughput, our mobile application accessed 5 public web-pages, and we estimated the amount of bytes the device had received. We measured the average of RX rate at a specific location. The collected data was uploaded to a Microsoft Azure server once per day only when the smartphone was using a WiFi network to avoid wasting the cellular budget. During the experiment, the participants turned WiFi on an average of 14 hours per day, and the average volume of network usage per day was 235MB. The average volume of WiFi network usage was 228MB.

B. *Macro Benchmarks*

We validated the overall performance of the proposed system compared to related approaches. Note that we denote the proposed system as the adaptive method.

AMUSE [7]. The method is the representative one that predicts the throughput of WiFi APs using the location-dependent predictor [2, 4, 5, 7]. The network usage predictor also used the spatial model without the adaptive selection scheme.

Liu et al. [8]. This method uses the location-independent predictor to predict the throughput of WiFi APs. The temporal model is used for the network usage predictor without the adaptive selection scheme.



(a) (b)

Fig. 4. Cumulative error of the throughput predictor (a) of all users and (b) of one user.

Spatio-Temporal Method. This method uses the spatio-temporal model to predict both the throughput of WiFi APs and the network usage. We did not use the adaptive selection of network usage predictor in this method, but we applied the fallback mechanism as proposed in [16]. When the number of samples was zero, the scheme used the spatial method. Similarly, the fallback of the spatial method is the temporal method. This is a variation of the proposed scheme.

We implemented the proposed algorithm and the related schemes in Java and validated the performance based on the real traces collected by 30 users for two months. The trace-based experiment is commonly used in the network-related systems due to the practical difficulty of real deployment.

1) *Prediction of the Throughput of WiFi APs*

To evaluate the performance of the throughput predictor, we measured the cumulative error of three prediction methods: spatial, temporal, and spatio-temporal. Each time a participant uses a WiFi network, the throughput is predicted based on the previous usage pattern. The cumulative error is calculated by accumulating the difference between actual and predicted throughput for an individual day. Fig. 4(a) shows the CDF graph of cumulative throughput error of all participants according to the type of method used. Fifty percent of the error of the spatial, temporal, and spatio-temporal methods was smaller than 486Mbps, 567Mbps, and 412Mbps, respectively. This result indicates that our spatio-temporal prediction method achieves higher accuracy than other methods. The temporal method has the largest error, which indicates that the throughput of a network is more dependent on WiFi APs than on time of day. The spatio-temporal method has the smallest error because of the consideration of throughput variation according to location and time of day. We also traced the daily cumulative throughput error of a single participant as shown in Fig. 4(b). On most days, the spatio-temporal method is more accurate than the other methods, and the error decreased as the learning period increased.

To investigate why the spatio-temporal method is superior to the other methods, we further analyzed the actual throughput of WiFi APs used in the experiment. Fig. 5(a) shows the throughput distribution of the three most-used APs as a histogram. The throughput is different depending on the AP; the most frequent throughputs of AP1, AP2, and AP3

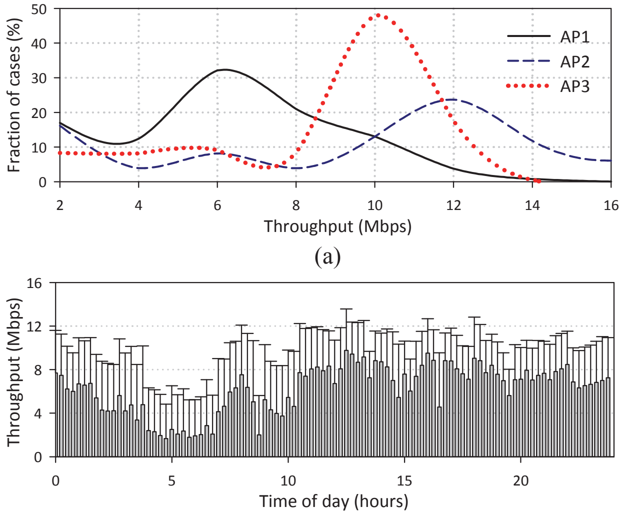


Fig. 5. (a) Distribution of the throughput of WiFi APs; (b) Distribution of the throughput of a single WiFi AP according to time of day.

are 6Mbps, 12Mbps, and 10Mbps, respectively. This clearly shows that WiFi throughput is location-dependent since association with a certain AP is highly related to the location. Fig. 5(b) shows the average WiFi throughput of a single AP for each hour during the experiment. The throughput in the nighttime is relatively lower than in the daytime. Moreover, the throughput fluctuated from 2 to 15 Mbps from time to time. The result of this analysis indicates that the spatio-temporal method is most suitable for the prediction of WiFi throughput.

2) *Prediction of Network Usage*

To evaluate the network usage predictor, we measured the cumulative error of the three prediction methods in the same manner as the throughput predictor. The volume of network usage is predicted based on the previous usage pattern. The cumulative error is calculated by accumulating the difference between the actual and predicted volume of network usage for a single day. Fig. 6 shows the cumulative prediction error of all participants according to the type of method. The temporal method shows the largest cumulative error, whereas the spatio-temporal method shows the lowest errors; In 50% of days, the network usage error of temporal, spatial, and spatio-temporal methods were smaller than 32MB, 20MB, and 14MB, respectively. This result indicates that the network usage of mobile users is more related to the location rather than the time of day. The lower similar errors of the spatial and spatio-temporal methods indicate that network usage at different times of day is trivial when a user is stationary at certain locations.

To investigate why the temporal method was less effective than the spatio-temporal and spatial methods, we further analyzed the network usage patterns of the participants. Since the prediction accuracy is highly related with the regularity of network usage patterns, we estimated the standard deviation of network traffic according to time

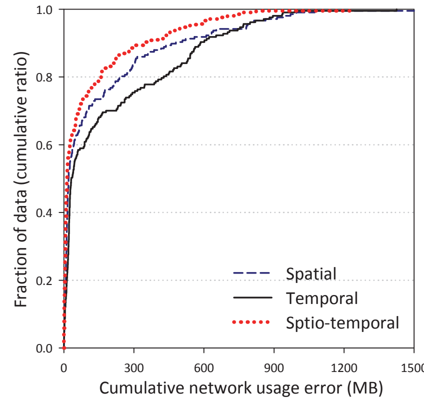


Fig. 6. Cumulative error of the network usage predictor for all users.

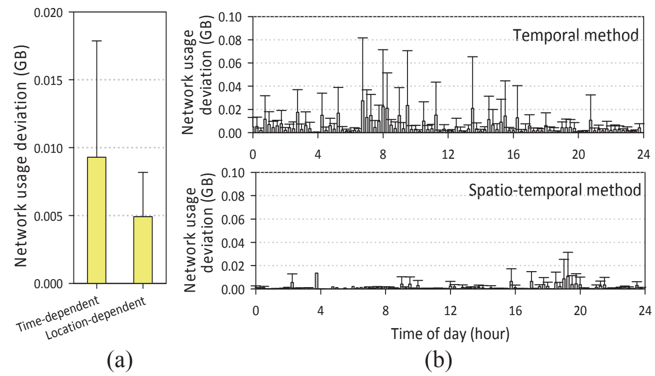


Fig. 7. Deviation of network usage (a) of all users; (b) of one user. The upper figure in (b) is the deviation generated using the temporal method while the lower one is generated using the spatio-temporal method.

and location, respectively, as shown in Fig. 7(a). The time-dependent deviation was much larger than the location-dependent deviation. The result indicates that the temporal method is not suitable for network usage prediction. Fig. 7(b) shows the standard deviation of the network usage for each period. The upper graph shows the average deviation of each time regardless of location, while the lower graph shows the average deviation of each time in the same location. The average deviation of the temporal network usage was larger than that of the spatio-temporal network usage. In the spatio-temporal network usage, the large temporal deviation was reduced by the low spatial deviation. This result indicates that the proposed spatio-temporal method has a similar performance to the spatial method because of the fallback mechanism. In summary, the spatio-temporal and spatial methods are more accurate for network usage prediction than the temporal method, since the network usage pattern is highly correlated with location rather than with time.

3) *Evaluation of the Offloading Scheduler*

We evaluated the performance of our offloading scheduler by estimating the used throughput according to the type of prediction method: without offloading, temporal, spatial, spatio-temporal, and the adaptive method. The adaptive method is differentiated from the spatio-temporal method by the adaptive selection scheme in the network

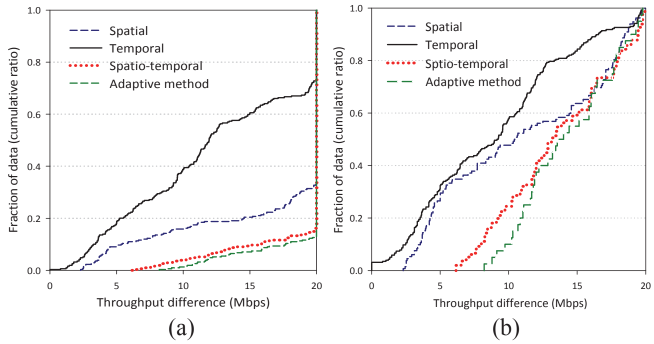


Fig. 8. (a) Throughput of both cellular and WiFi provided by the different methods and (b) throughput of WiFi provided by the different methods.

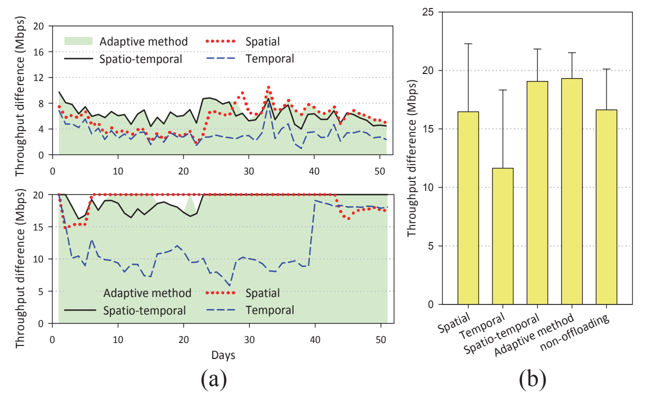


Fig. 9. (a) Throughput provided by the different methods in two users; (b) Throughput difference in a day.

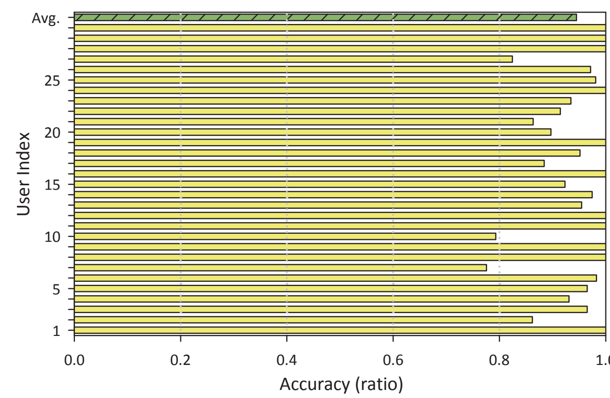


Fig. 10. Average usage time of WiFi according to WiFi usage type of participants.

usage predictor. Fig. 8(a) shows the CDF graph of the average throughput difference from both LTE and WiFi usage according to the method used. Here, the throughput difference represents the average throughput while the participants use a WiFi or a cellular network for a day. We set the throughput of LTE as 20Mbps since 88% of WiFi throughput is less than 20Mbps in the collected traces. The throughput difference in both LTE and WiFi is shown in Fig. 8(a) while Fig. 8(b) shows one in only WiFi networks. The throughput difference with the temporal method was much lower than other methods; 22% of the throughput difference of the temporal method was 20 Mbps, while it was 62% for the spatial methods. The adaptive method clearly outperforms the related methods; 85% of throughput gain was 20 Mbps. This is due to the high accuracy of network usage prediction with the adaptive selection.

Fig. 9(a) shows the trace of throughput difference of two different participants. The proposed method shows the best performance for most days by adaptively changing the network usage prediction method. Without the adaptive selection scheme, the spatio-temporal method shows high throughput since it handles the variation of throughput and network usage according to location and time of day. The high throughput in the adaptive method indicates that the network usage predictor adequately selected the optimal predictor for each user. We transformed the cumulative throughput gain into the daily throughput, as shown in Fig. 9(b). The result showed that the proposed scheme outperforms the related works. Considering that our scheme did not charge additional data fee, the results are acceptable to mobile users.

We investigated how accurately our method selects an optimal method for network usage prediction. Our offloading scheduler always chose the best prediction method with an efficiency of more than 75% regardless of user, and the average accuracy in selection was 94%. This means that the proposed method can choose the most suitable method for network usage prediction accurately.

We finally investigated the WiFi usage time in the proposed method. Note that more than half of mobile users tended to turn off WiFi in daily life due to battery issue as well as low quality of WiFi networks (see Section II). Fig. 10 shows the ratio of WiFi usage time in the proposed method over typical usage time of mobile users. The participants originally activated WiFi for 14.6±7 hours in a day, whereas the usage time is reduced to 2.67±6 hours with the proposed method. Such reduction of usage time increases the daily battery lifetime by 48±20% in Galaxy S3 when the user used the phone for 9 hours. Considering that the proposed method still provides high throughput of WiFi networks, the result indicates that the method not only derives high throughput but also minimizes the redundant energy usage of WiFi.

C. Micro Benchmarks

For the micro benchmarks, we analyzed the performance of the proposed system according to the cellular budget and learning period.

1) Effect of Cellular Budget

We evaluated the throughput gain of the offloading scheduler according to the five common cellular budgets (500MB, 1GB, 2.5GB, 5GB, and 10GB per month). The optimal model provides high throughput within the cellular budget even if the allocated cellular budget is small. As shown in Fig. 11, the constant use of a single method from the three methods (temporal, spatial, and spatio-temporal) could not guarantee the best performance regardless of the cellular budget. In contrast, the proposed offloading scheduler always exhibits the largest throughput gain despite of the data plan by adaptively selecting the most suitable

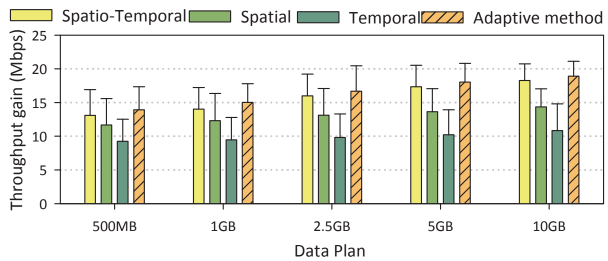


Fig. 11. Average throughput according to different monthly budgets.

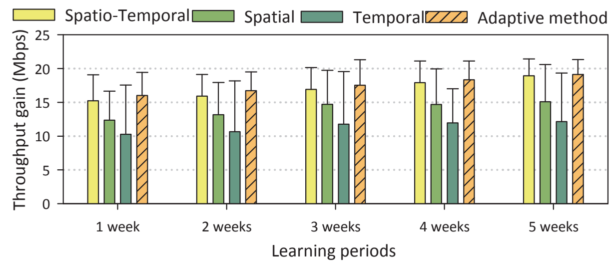


Fig. 12. Average throughput according to different learning periods.

prediction method. This result shows that the proposed scheme provides the highest throughput by offloading the data through WiFi when the given cellular budget is same.

2) *Effect of Learning Period*

We investigated the performance of the proposed offloading scheduler according to different learning periods. The optimal model should provide robust results despite a short learning period. Fig. 12 shows the provided throughput according to learning period from one to five weeks. The throughput gain with the proposed method was always larger than the other methods regardless of learning period. This is due to the benefit of the adaptive method selection in the network usage predictor. The high throughput during the first week is because of the large portion of LTE use. However, the high throughput during the fifth week with higher WiFi usage indicates that the adaptive method provides the most robust throughput, despite a large portion of WiFi use as the learning period increases.

VI. Conclusion

In this paper, we propose a personalized offloading system to provide maximum throughput within monthly cellular budgets. To the best of our knowledge, we are the first to design an offloading policy that adaptively manages diverse patterns of throughput and network usage in practice. We designed an adaptive offloading scheduler using MDP with a weighted cellular budget. The experiments with 30 mobile users show that our approach provides 29% higher throughput than previous studies with diverse cellular budgets. We believe that the proposed scheme would improve the quality of network usage in smartphones by automatically switching the network (i.e., LTE and WiFi) without user intervention.

VII. Acknowledgement

This work was supported by a grant from the National Research Foundation of Korea (NRF), funded by the Korean government, Ministry of Education, Science and Technology under Grant (NRF-2014R1A2A1A11049979).

References

[1] Cisco, Cisco Visual Networking Index: Forecast and methodology, Cisco Public, 2014.

[2] A. Balasubramanian, R. Mahajan, and A. Venkataramani, Augmenting mobile 3G using WiFi, in Proceedings of the 8th MobiSys, ACM, San Francisco, CA, USA (2010) 209-222

[3] K. Lee, J. Lee, Y. Yi, I. Rhee, and S. Chong, Mobile data offloading: How much can WiFi deliver?, in IEEE/ACM Transactions on Networking, Vol. 21, no. 2, April (2013) 536-550

[4] B. Han *et al.*, Cellular traffic offloading through opportunistic communications: A case study, in Proceedings of the 5th ACM workshop on CHANTS, ACM, USA (2010) 31-38

[5] O. B. Yetim and M. Martonosi, Adaptive usage of cellular and WiFi bandwidth: An optimal scheduling formulation, in Proceedings of the 7th ACM International Workshop on CHANTS, ACM (2012) 69-72

[6] S. Nirjon, A. Nicoara, C.-H. Hsu, J. Singh, and J. Stankovic, Multi-Nets: Policy oriented real-time switching of wireless interfaces on mobile devices, in Proceedings of IEEE 18th RTAS, IEEE, Beijing, China (2012) 251-260

[7] Y. Im *et al.*, AMUSE: Empowering users for cost-aware offloading with throughput delay tradeoffs, in Proceedings of the 32nd INFOCOM, IEEE, Turin, Italy (2013)

[8] H. Liu *et al.*, Efficient 3G budget utilization in mobile participatory sensing applications, in Proceedings of the 32nd INFOCOM, IEEE, Turin, Italy (2013), 1411-1419

[9] Wonbo Lee, Jonghoe Koo, Sunghyun Choi, and Yong Seok Park, ESPA: Energy, usage ($), and performance-aware LTE-WiFi adaptive activation scheme for smartphones, in Proceedings of WoWMoM, IEEE, Sydney, Australia (2014)

[10] A. Aijaz, H. Aghvami, and M. Amani, A survey on mobile data o ffloading: technical and business perspectives, IEEE Wireless Communications, 20 (2) (2013) 104–112

[11] A. Y. Ding *et al.*, Enabling energy-aware collaborative mobile data offloading for smartphones, in Proceedings of the 10th SECON, IEEE, New Orleans, USA (2013), 487-495

[12] P. Deshpande, X. Hou, and S. R. Das, Performance comparison of 3G and metro-scale WiFi for vehicular network access, in Proceedings of the 10th ACM SIGCOMM conference on Internet Measurement (IMC), ACM, Melbourne, Australia (2010), 301-307

[13] S. Dimatteo *et al.*, Cellular traffic offloading through WiFi networks, in Proceedings of the IEEE 8th MASS, IEEE, Wuhan, China (2011), 192-201

[14] M. Cheung and J Huang, Optimal delayed Wi-Fi offloading, in Proceedings of the 11th WiOpt, IEEE, Tsukuba, Japan (2013), 564-571

[15] B. D. Higgins, J. Flinn, T. J. Giuli, B. Noble, C. Peplin, and D. Watson, Informed Mobile Prefetching, in Proceedings of 2012 MobiSys, ACM, New York, USA (2012) 155-168.

[16] L. Song, D. Kotz, R. Jain, and X. He. Evaluating next-cell predictors with extensive wi-fi mobility data. IEEE Trans. Mobile Computing, 5(12), Dec. (2006) 1633-1649

[17] Y. Chon, H. Shin, E. Talipov, and H. Cha, Evaluating mobility models for temporal prediction with high-granularity mobility data, in Proceedings of 10th PerCom, IEEE, Lugano, Switzerland (2012) 206-212

[18] W. Jung, Y. Chon, D. Kim, and H. Cha, Powerlet: An active battery interface for smartphones, in Proceeding of 2014 UbiComp, ACM, Seattle, USA (2014)