MOBILE TRAFFIC OFFLOADING BY EXPLOITING SOCIAL NETWORK SERVICES AND LEVERAGING OPPORTUNISTIC DEVICE-TO-DEVICE SHARING

XIAOFEI WANG, MIN CHEN, TED "TAEKYOUNG" KWON, LIANGHAI JIN, AND VICTOR C. M. LEUNG

Xiaofei Wang is with Huazhong University of Science and Technology and the University of British Columbia

Min Chen and Lianghai Jin (corresponding author) are with Huazhong University of Science and Technology.

Ted "Taekyoung" Kwon is with Seoul National University.

Victor C. M. Leung is with the University of British Columbia.

ABSTRACT

The ever increasing traffic demand is becomming a serious concern for mobile network operators. In order to solve the traffic explosion problem, there have been research efforts on offloading the traffic from cellular links to local communications among users. In this article, we leverage opportunistic device-to-device sharing, exploit the social impact among users in social network services (SNSs), and propose the novel framework of traffic offloading assisted by SNSs via opportunistic sharing in mobile social networks, called TOSS. In TOSS, initially a subset of mobile users are selected as seeds depending on their content spreading impact on online SNSs and their mobility patterns in offline MSNs. Then users share content objects via opportunistic local connectivity (e.g., WiFi Direct) with each other. Furthermore, the observation of SNS user activities reveals that individual users have distinct access patterns, which allows TOSS to utilize the user-dependent access delays between the content generation time and each user's access time for opportunistic sharing purposes. By trace-driven evaluation, we demonstrate that TOSS can reduce by 63.8–86.5 percent of cellular traffic while satisfying the access delay requirements of all users. Therefore, traffic offloading by leveraging opportunistic device-to-device sharing based on SNSs can be quite effective and efficient as a promising content delivery service for traffic reduction in future mobile networks.

INTRODUCTION

Due to the fast development of mobile communication technologies, more and more users tend to download content on mobile devices(e.g., reading articles and watching videos on phones and tablets). The ever increasing traffic load is becoming a serious concern of mobile network operators (MNOs), but the study in [1] pointed out that a large portion of the traffic load is due to duplicated downloads of the same popular files. For instance, the top 10 percent of popular videos in YouTube account for nearly 80 percent of all views [1]. Hence, how to effectively reduce duplicated downloads via cellular links by offloading certain data traffic via other types of networking connectivities has become a hot research topic.

In practice, mobile users can often be surrounded by many adjacent neighbor users. One can easily discover neighboring users in proximity and set up temporary local connectivities with them [2], via WiFi Direct and device-to-device (D2D) in Long Term Evolution (LTE)-Advanced, and near-field-communication (NFC) for content sharing. Note that throughout this article, we use the term D2D sharing synonymously for all the user-to-user, device-to-device, and people-to-people sharing techniques mentioned above for simplicity.

Therefore, many studies are being carried out to exploit opportunistic sharing (i.e., epidemic sharing) by intermittent meetings among mobile users for traffic offloading in mobile social networks (MSNs) [3, 4]. Some users can initially download content objects via the cellular link and later share with nearby mobile users while moving. It is proposed that, by selecting an appropriate initial pushing set of seeds and utilizing D2D sharing, content can be disseminated efficiently while cellular traffic can be reduced significantly [4]. However, there are still several important issues in related research that have not been fully elaborated, such as:

- *How can we know or predict the dissemination delay of each user for each content?* We cannot simply assume the same dissemination deadline, while users indeed have various delay requirements [5].
- *How do we design the seeding strategy to minimize cellular traffic while satisfying the delay requirements of all users?* Strategies for selecting initial seeds are discussed in prior work [3], but user mobility is mostly considered by ignoring the real social relationships among users.
- *How can we efficiently make mobile users share*

content with others? Studies in [3, 4] assume people will always exchange content gratuitously. However, in reality, people mostly share information by "word-of-mouth" propagation [6]; hence, the realistic social relationship among users should be exploited.

Recently it was discovered that there is a dramatic rise in the number of mobile users who participate in online SNSs (Facebook, Twitter, Tumblr, Sina Weibo, etc.), where more and more media content has been shared and propagated rapidly and widely [6]. Therefore, we seek to exploit the social relationship of users in both the offline MSNs and the online SNSs. Note that the "online" here indicates the SNSs in the Internet formed by the virtual accounts used by real people, while the "offline" MSN here means the realistic MSNs (e.g., meetings and groups) formed by people in real life. By investigating related measurement and modeling studies of MSNs and SNSs, we point out the following key points, which can be utilized for content sharing and traffic offloading:

- In MSNs, the mobility patterns of users can be measured and modeled [4], and hence each user can have a different offline mobility impact regarding the dissemination of content files.
- Due to the effect of word of mouth [6], users significantly impact information spreading to friends. The influence, or spreading impact, can be modeled based on the analysis of social behavior histories, while the user's historical measurement of spreading impact can be used for forecasting future sharing activities.
- In SNSs, the access pattern of each user can be measured, statistically modeled by fitting and thus mostly predicted [5, 6]. That is, we can analyze and utilize the access delay between the content generation time and the user access time, which is per-user-dependent mainly due to people's different lifestyles. Hence, we can disseminate the content of interest to users considering their different access delay sensitivities.
- User relationships and interests in online SNSs have significant homophily and locality properties, which are similar to those of offline MSNs [7]. Homophily means the "birds of a feather" effect; that is, people with similar interests like to share interesting content with each other. Locality means that people who are graphically close may have similar trends of accessing and sharing content with each other [7].

Since users are mostly clustered by geographical regions and interest demands, there have been some studies to accelerate the content sharing in MSNs by investigating users' social relationships, such as, BUBBLE Rap [8], that utilizes social grouping characteristics in the real world. In this article, we are motivated to exploit the real social relationships mined from online SNSs, and hence propose a novel framework of traffic offloading assisted by SNSs via opportunistic sharing in mobile social networks, called TOSS. TOSS pushes the content object to a properly selected group of seed users, who will opportunistically meet and share the content with others, depending on their spreading impact

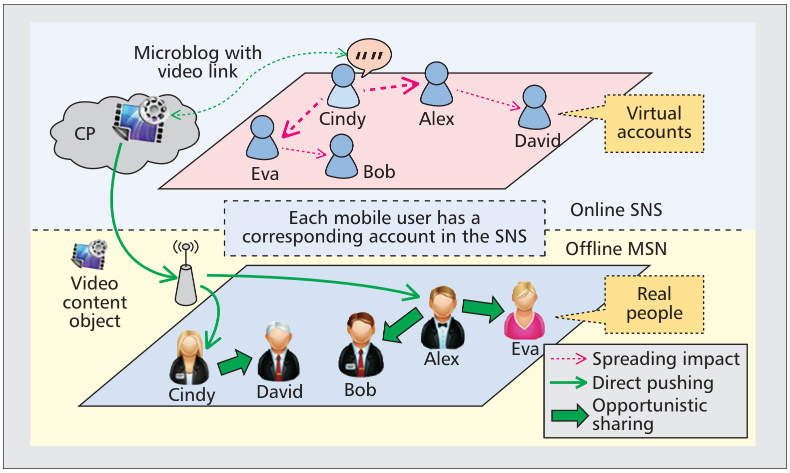

Figure 1. Illustration of an example of the TOSS framework.

on the SNS and their mobility impacts in the MSN. TOSS further exploits the user-dependent access delay between the content generation time and each user's access time. From trace-driven evaluation and model-based analysis, TOSS lessens the cellular traffic by 63.8–86.5 percent, while still satisfying the delay requirements of all users. To the best of our knowledge, this is the first study that seeks to combine online SNSs with offline MSNs for traffic offloading considering user access delay sensitivities. Note that this article is based on our study in [4], in which more technical details can be found.

The rest of the article is organized as follows. We show the TOSS preliminaries, and explain the details of the mobility impact factor, spreading impact factor, and access delays, respectively. Then the optimization issue is discussed, and evaluation results are shown. Finally we describe concluding remarks and future work.

The TOSS Framework Preliminaries

As shown in Fig. 1, the TOSS framework entails both an online SNS and an offline MSN.. Supposing there are total N mobile users, u_i, $i = 1, \dots, N$, who have corresponding SNS identities, for any two users, u_i and u_j, if u_j follows u_i, u_j is one follower of u_i and u_i is one followee of u_j. At any time, a user may find or create a new interesting article, image, or video, and share it in the SNS as an initiator of the content. We define a short message posted by a user containing the content (or link to the content) as a microblog (e.g., a tweet in Twitter or a post on Facebook), and a content file is called a content object. All his/her followers will then be able to access the content, and some of them will further reshare on their timelines. Making comments will not induce any information spread; thus, we only consider resharing. Afterward, while the microblog is being spread to other users in the online SNS, the content object will be accessed and spread in the offline MSN.

In the TOSS framework, we define several factors for each user, u_i: two for the online SNS, (1) the outgoing spreading impact, $I_i^{S\rightarrow}$, and (2)

The Infocom trace has the highest contact rate because users are clustered at a conference spot. The MIT trace also has high contact rate since users are friends within the campus. The Beijing and the SUVnet traces have large inter-contact intervals because they have relatively low frequency of GPS records and large user base.

the incoming spreading impact, $I_i^{S\leftarrow}$, both of which indicate how important the user is in propagating the microblog (to others or from others); one for the offline MSN, the mobility impact, I_i^M, which indicates how important the user is in sharing the content object (to others or from others) via physical encounters.

Due to the symmetric nature of meetings (contacts), there is no need to have outgoing or incoming mobility impact. In later sections, we briefly discuss how to calculate them. Therefore, considering the above factors, TOSS can seek to select a proper subset of users as seeds for pushing the content object directly via cellular links, and exploit the user-to-user sharing in the offline MSN. We thus define a vector $\vec{p}$ to indicate whether or not to push the content object to a user via cellular links (e.g., $p_i = 1$ means pushing the content object directly to user u_i). Note that the TOSS framework is not confined strictly to the dissemination of one popular content to all the users, but applies to general deliveries of any content to a group of potential recipients of any size.

From the illustrated scenario of TOSS in Fig. 1, in the online SNS, Cindy shares a video (link) with Eva and Alex, who may in turn share with Bob and David, respectively. Meanwhile, the video content is first downloaded via a cellular link and stored in Cindy's phone. However, in the offline MSN, Cindy is geographically distant from other people but David is in proximity. Although David may not know Cindy, TOSS detects that the $I^{S\rightarrow}$ impact of Cindy to David via Alex is also very strong, and thus lets Cindy share the video with David via local WiFi connectivity. Furthermore, TOSS evaluates the I^M impact of Alex and pushes another copy to him via a cellular link, because Alex is likely to meet Bob and Eva in the offline MSN frequently, and Bob and Eva often access content with some delays. Then the content object will be propagated by local connectivities from Alex to Bob and to Eva later. TOSS reduces 3/5 of the cellular traffic in this scenario.

Mobility Impact in the Offline MSN

Modeling of Mobility Impact

It has been shown that mobile users in offline MSNs have different mobility patterns [4] and hence different potentials for sharing content. Thus, the mobility impact, I^M, is defined to quantify the capability of a mobile user to share a content object with other users via opportunistic meetings, or contacts, while roaming in the MSN. The temporary connectivity with nearby users mostly relies on active discovery mechanisms; thus, we assume all mobile users are synchronized with a low duty cycle for probing as proposed by eDiscovery [2]. Referring to [3, 4] we use λ_{ij} to denote the opportunistic contact rate of user u_i with user u_j. Hereby, we define the mobility impact factor I_i^M for u_i to the whole user base as

$$I_i^M = \sum_{j=1}^{N} \lambda_{ij}. \quad (1)$$

Finally, regarding the delivery of one content object, given the initial pushing vector $\vec{p}$, as well as the inter-contact rate λ_{ij} of any two users u_i and u_j, we can utilize epidemic modeling to model the opportunistic sharing procedure, and thus calculate how long it will take for any user u_i to obtain the content object, defined as the *content obtaining delay* of u_i, denoted by t_i^*. The calculation details are skipped in this article, but readers may refer to related studies in [4]. Note that TOSS does not need to shorten the content obtaining delays for all users, but seeks the optimal $\vec{p}$ to induce proper content obtaining delays to match with the delay sensitivity of each user, which is detailed later.

Measurement of Mobility Impact

We choose to analyze four popular mobility traces, MIT [9], Infocom [10], Beijing [11], and SUVnet [12], which record either direct contacts among users carrying mobile devices or the GPS coordinates of each user's mobile route. The four traces differ in their scales, durations, and mobility patterns. The MIT and Infocom traces are collected by normal people, but the Bejing and SUVnet traces are collected by vehicles.

From the traces, we obtain the inter-contact intervals, $1/\lambda$, of all user pairs, as shown in Fig. 2a. The Infocom trace has the highest contact rate because users are clustered at a conference spot. The MIT trace also has a high contact rate since users are friends within the campus. The Beijing and SUVnet traces have large inter-contact intervals because they have relatively low frequency of GPS records and large user bases. Also, I^M values of all users of the traces are plotted in Fig. 2b (values smaller than 0.001 are ignored), which indicates similar trends as the traces as discussed above. Users in the Infocom trace have the highest potential to obtain the content by D2D sharing, and users in the Beijing trace have the weakest potentials.

Spreading Impact in the Online SNS

Calculation of Spreading Impact

There are many ways to evaluate user impact on information spreading in SNSs, and we decide to utilize probabilistic models to quantify the content spreading impact. Hereby we define the spreading impact factor ($I^{S\rightarrow}$) of user u_i to user u_j, denoted by γ_{ij}, $0 \le \gamma_{ij} \le 1$ (inversely $I^{S\leftarrow}$ factor defined by γ_{ji}), is the ratio of the number of the microblogs of u_i that u_j accesses and reshares to the number of all microblogs in u_j's timeline. From another perspective, γ_{ij} can indicate how often u_i's microblogs will appear in u_j's timeline. From u_j's point of view over a certain period, we need to consider

- The number of microblogs u_j has created by herself
- The number of reshared microblogs by u_j from u_i
- The number of re-shared microblogs from all followees, for calculating the 1-hop impact by

$$\gamma_{ij}^1 = \frac{\#\text{ reshares from } u_i \text{ to } u_j}{\#\text{ microblogs created by } u_j + \sum_{1-\text{hop followees}} \#\text{ reshared from followees to } u_j}$$

By recursively accumulating the spreading impact, we can obtain two-hop impact by:

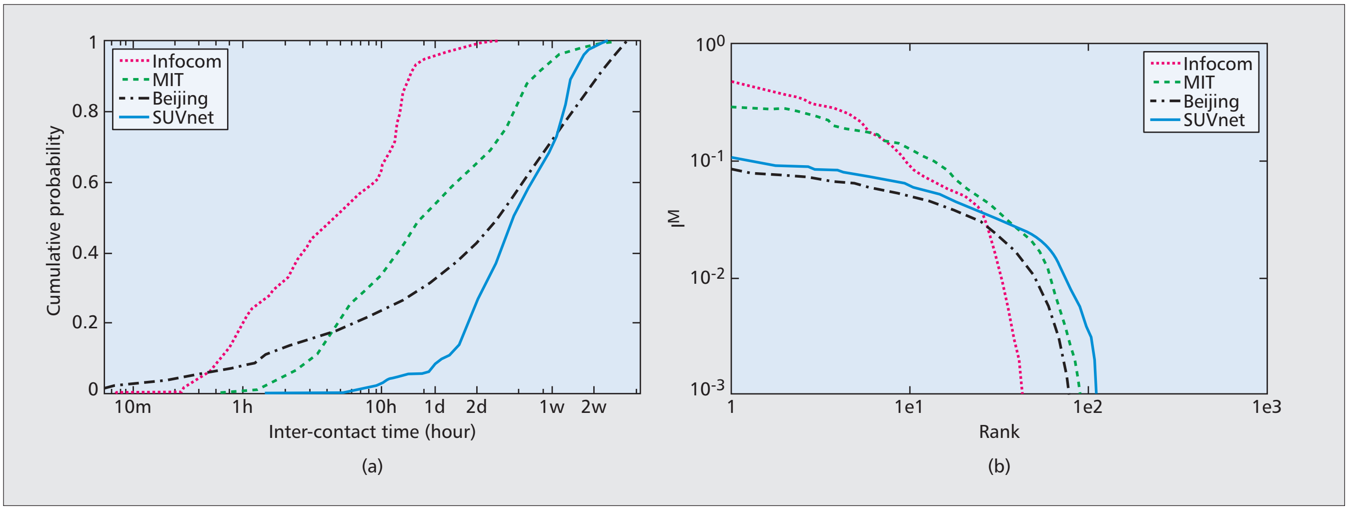

Figure 2. Measurement of $1/\lambda_{ij}$ and I^M: a) analysis of $1/\lambda_{ij}$; b) analysis of I^M.

$$\gamma_{ij}^2 = 1 - \prod_{\substack{u_k \in 1-\text{hop followers of } u_i \\ \&\&\ 1-\text{hop followees of } u_j}} (1 - \gamma_{kj}^1 * \gamma_{kj}^1),$$

and so on. In this article, we skip the calculation details, but readers can refer to [4]. Then, finally, we use γ_{ij}^* to denote the impact from user u_i to user u_j via all possible paths with less than or equal to H hops.

$$\gamma_{ij}^* = 1 - \prod_{n=1}^{H} (1 - \gamma_{ij}^n), \tag{2}$$

Note that the average path length in SNSs is normally 4.12, as studied in [6], so we set $H = 4$. Then, finally, $I^{S\rightarrow}$ and $I^{S\leftarrow}$ of u_i to and from the whole user base can be respectively calculated by

$$I_i^{S\rightarrow} = \sum_{j=1}^{N} \gamma_{ij}^*, I_i^{S\leftarrow} = \sum_{j=1}^{N} \gamma_{ji}^*. \tag{3}$$

For spreading impact, γ_{ij} is considered as the sharing probability that u_j hopes u_i will share the content object when u_i and u_j meet. Consequently, we can finally get the mobility impact factor of any two users, λ_{ij}, obtained from mobility traces, the spreading impact of any two users, γ_{ij}, and the content obtaining delay PDF, t_i^*, of any user u_i, given the initial pushing vector $\vec{p}$, which can be obtained from SNS traces.

$$t_i^* = \textbf{\textit{ContentObtainingDelay}}\ (\{\gamma_{ij}^*\}, \{\lambda_{ij}\}, \vec{p}).$$

Measurement of Spreading Impact

We selected the most popular online SNS in China, Sina Weibo, and keep track of 2,223,294 users for four weeks. We collected a total of 37,267,512 microblogs generated (and partially reshared) by the users, and further obtained the list of all resharing activities for each microblog. However, calculating the overall spreading impact, I^S, for the whole user base takes a substantially long time. Thus, we analyze the sub-graphs of a corresponding number of users from the whole social graph by random walking according to the scale of the mobility traces (discussed in detail later). There have been some related measurement studies in [6, 13, 14] pointing out that the SNS is a scale-free complex network, in which node strength distribution follows the power law, at least asymptotically. That is, a small number of nodes have dominant impacts on the network, while many nodes have very little impact if we consider the node degree or the spreading impact (resharing ratio) as the strength of a node to the network [13, 14]. Due to the characteristics of scale-free complex networks, subgraphs from the whole network graph (when not too small) by the random walking method can still obtain similar power law characteristics.

We draw the log-log plots for $I^{S\rightarrow}$ and $I^{S\leftarrow}$ of the nodes from the sampled sub-graphs as shown in Fig. 3. We can see that a smaller number of people have significant outgoing impact ($I^{S\rightarrow}$) to the whole SNS, while many users have very little impact. We also see that many users are more likely to be impacted rather than impacting others ($I^{S\rightarrow} < I^{S\leftarrow}$). All of the figures are able to reflect the asymptotical power law trend. Thus, conclusively, all of the sub-graphs with different sizes can still represent the SNS characteristics, and it will be an acceptable methodology to map the SNS sub-graphs to the mobility traces.

Access Delays of Users in the Online SNS

In the SNS, while some users may access the SNS frequently, others access the SNS at relatively longer intervals. Thus, the access delay between the content generation time and a user's access time becomes different for each user [5, 6]. As illustrated in Fig. 4, user A creates a microblog for an interesting video in the SNS at t_0. One of A's followers, B, happens to see A's microblog after a certain delay at t_1 due to B's personal business. Once B clicks to play it, a buffering delay is needed until t_2; B will reshare

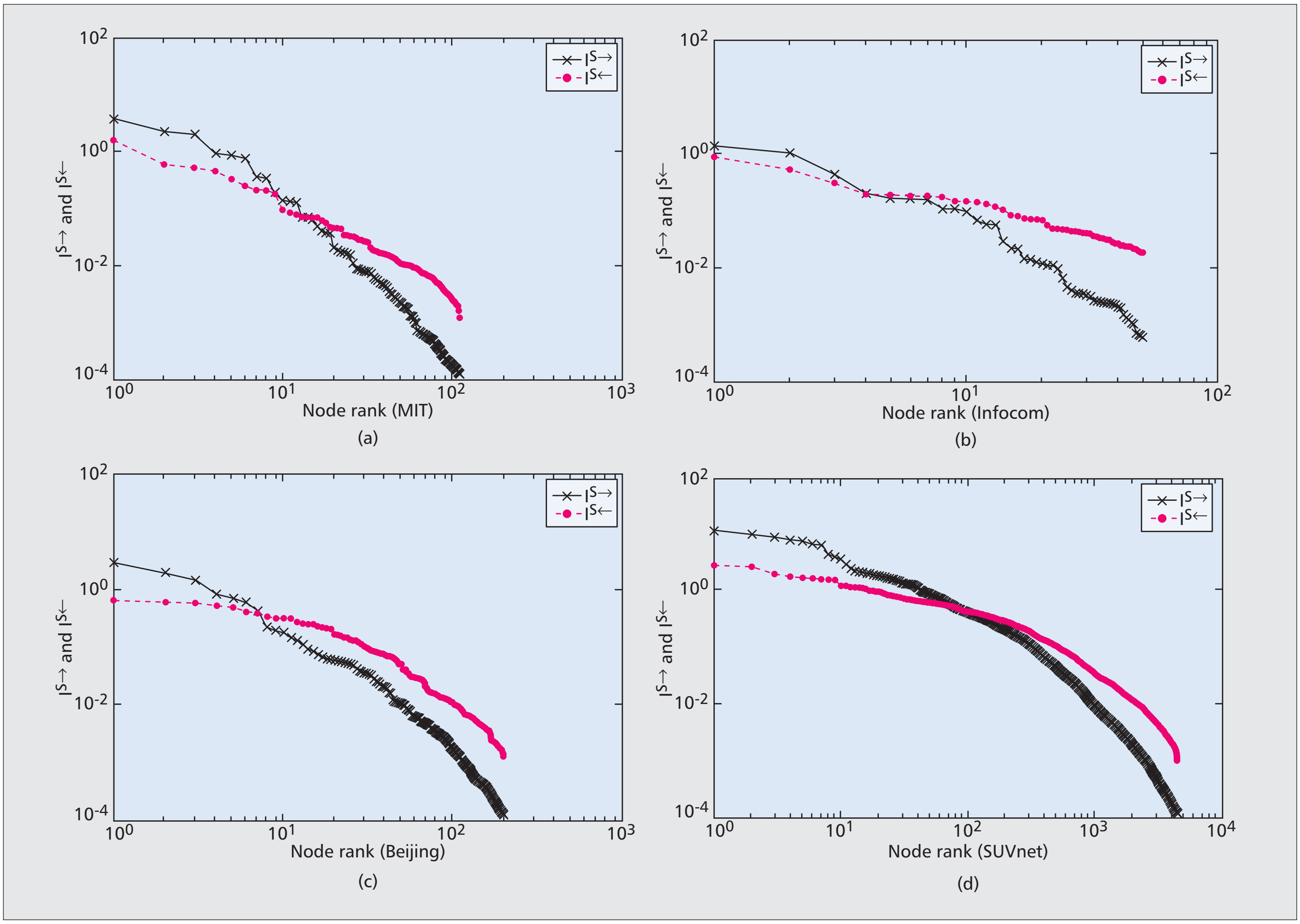

Figure 3. Measurement of I^S for sub-graphs sampled from the SNS graph with different sizes corresponding to the mobility traces: (a) MIT (100 users); (b) Infocom (41 users); c) Beijing (182 users); d) SUVnet (4311 users).

the video at t_3 after watching it. According to the definition of access delay given above, the access delay for B should hence be $t_2 - t_0$. However, in practice, only SNS providers can have access to t_1 and t_2 data. Since for texts, images, and most videos, t_1, t_2, and t_3 may be near, in our study, we consider B's access delay as $t_3 - t_0$, which can be captured from the SNS measurement trace by checking A's creating time and B's resharing time.

We still use the collected SNS trace data from Sina Weibo, and the access delay is gathered as the time difference between the generation time of the original microblog and the time of resharing by a follower. First, we pick up four real users from the trace and plot their access delays by probability distribution function (PDF), as shown in Fig. 5b. Users u_1 and u_2 like to access the content frequently with short delays. However, users u_3 and u_4 have significant access delays on the order of hours and even days.

From the cumulative distribution function (CDF) of the average of all the access delays of each user based on the whole user base in Fig. 5b, half of the users have the average access delays larger than 23,880 s, which is about 6 h and 38 min. Taking a closer look, we find that:
- 3.67 percent of users have average access delay shorter than 10 min.
- 20.38 percent of users have delay shorter than 1 h.
- 26.79 percent of users access the SNS with average delay longer than a day.

Therefore, we verify that a substantial number of users access the SNSs with large delays, which is a sufficiently large portion of users to allow TOSS to disseminate the content object by offline opportunistic sharing.

We use a PDF to model the access delays of each user, say u_i, in terms of the probability of accessing the content at t, denoted as ***Satisfactory***$_i(t)$, which can be considered the access satisfactory function. If the content object is already obtained locally in the user's device when he/she has the highest probability to access the content, he/she will be mostly satisfied. In order to model the various distributions of access delays with different shapes of PDF curves, we choose to use Weibull distribution for fitting, which is commonly used for profiling user behaviors in SNSs [15]:

$$\textbf{\textit{Satisfactory}}_i(t,\beta_i,k_i) = \frac{k_i}{\beta_i}\left(\frac{t}{\beta_i}\right)^{k_i-1} e^{-\left(\frac{t}{\beta_i}\right)^{k_i}}, t \geq 0, \quad (4)$$

where the Weibull fitting parameters β_i and k_i can identify the access pattern of user u_i.

System Optimization and Heuristic Algorithm

By evaluating I^S (both incoming and outgoing) and I^M values of all users, how to choose proper set of seeds for initial pushing, $\vec{p}$, to get the obtaining delay t^* for each user in order to maximize the sum of the access satisfactory for all users becomes the objective of TOSS. Thus, the sum of access satisfactory functions of all users should be maximized:

$$\begin{aligned} &\underset{\vec{p}}{\text{Maximize}} := \sum_{i=1}^{N} \boldsymbol{Satisfactory}_i(t_i^*, \beta_i, k_i) \\ &\text{Subject to} : \left|\vec{p}\right| \le C, \end{aligned} \tag{5}$$

where the total amount of initial pushing seeds is constrained by C, and we call $\Sigma \boldsymbol{Satisfactory}_i(t)$ the total access satisfactory function of the whole user base.

It is hard to solve the above optimization problem analytically, since all related equations are not closed-form. With power series approximations, we can find the maximum values by general numerical methods. Also, we can even tune and find the needed C given a target total access satisfactory value. Furthermore, it is possible to design a heuristic algorithm to find the near-optimal solution $\vec{p}$ for maximizing $\Sigma \boldsymbol{Satisfactory}_i(t)$ numerically by a general hill-climbing method, but due to the limited space, we skip discussing it.

TOSS Evaluation Results

Mapping the MSN and SNS

Due to the lack of a trace that contains the activities of the same users in both online SNSs and offline MSNs, we consider three choices for mapping SNS users (the Sina Weibo data set) to MSN users (the four mobility traces):

- Random: SNS users are randomly mapped to MSN users.
- h-h: both SNS and MSN users are sorted in descending order of $I^{S\rightarrow}$ and I^M respectively, and then are mapped correspondingly;
- h-l: Both SNS and MSN users are sorted, but an SNS user with high $I^{S\rightarrow}$ is mapped to an MSN user with low I^M.

Because different mapping schemes show marginal differences, in the following, we calculate the average values of evaluation results across the three mapping schemes to reflect a more general and realistic scenario with heterogenous user types.

Initial Pushing Strategies

To select the users who will be initial seeds, $\vec{p}$, constrained by the allowed total number of seeds, C, we consider the following five pushing strategies based on the impact factors:

- **p**-λ: We sort users by I^M in descending order and choose the top C ones.
- **p**-$\gamma^{\rightarrow}$: We sort users by $I^{S\rightarrow}$ in descending order and choose the top C ones.
- **p**-$\gamma^{\leftarrow}$: We sort users by $I^{S\leftarrow}$ in descending order and choose the top C ones.

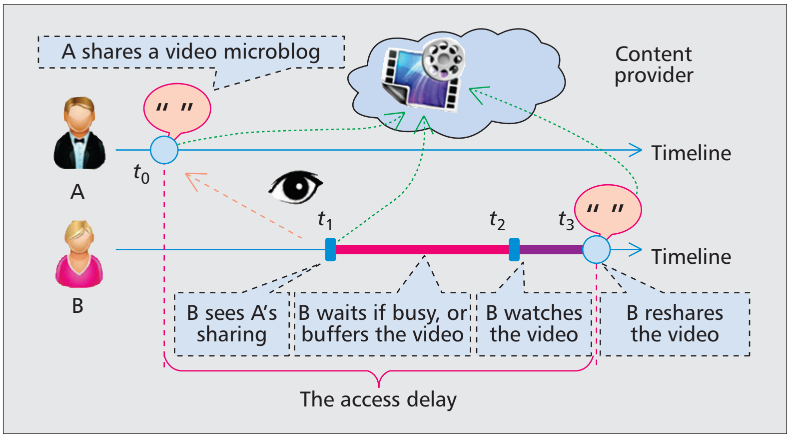

Figure 4. Illustration of content access delay.

- **p**-$\lambda^*\gamma^{\rightarrow}$: We sort users by $I^M * I^{S\rightarrow}$ conjunctively in descending order and choose the top C ones.
- **p**-$\lambda^*\gamma^{\rightarrow}$: We sort users by $I^M * I^{S\leftarrow}$ conjunctively in descending order and choose the top C ones.

In current literature, there are actually many viral marketing methods to evaluate an SNS user's strength regarding information spreading; for example, we can easily qualify by node degree including outgoing degree (number of followees) and incoming degree (number of followers). Note that here the arrow direction is the "following/followed" relationship, the reverse of the spreading direction. Furthermore, Google's PageRank algorithm [16] is applied to the selected SNS sub-graphs for getting the PageRank scores of all nodes. We also consider random pushing and the heuristic algorithm, and hence have five more initial pushing strategies based on the graphs:

- **p-R**: We randomly choose C users.
- **p**-$D^{\rightarrow}$: We sort users by outgoing node degree in descending order and choose C users.
- **p**-$D^{\leftarrow}$: We sort users by incoming node degree in descending order and choose C users.
- **p**-Pr: We sort users by PageRank score in descending order and choose the top C users.
- **p-H**: We run the hill-climbing heuristic algorithm to obtain the near-optimal pushing vector.

Note that we call these nine pushing strategies above, except **p-H**, "simple" pushing strategies.

Traffic Reduction for Satisfying 100 Percent, 90 Percent, and 80 Percent of Users

Recall that the access satisfactory function of u_i is $\boldsymbol{Satisfactory}_i(t)$. A user is **satisfied** if he/she can obtain the content when his/her access probability ($\boldsymbol{Satisfactory}_i(t)$) approaches its maximum (in the fitting Weibull pdf). We investigate what percentage of users (initial pushing ratio) should be initial seeds to satisfy the access delay requirements of 100 percent, 90 percent, and 80 percent of users depending on different pushing strategies.

From earlier, $\boldsymbol{Satisfactory}_i(t)$ is an increasing

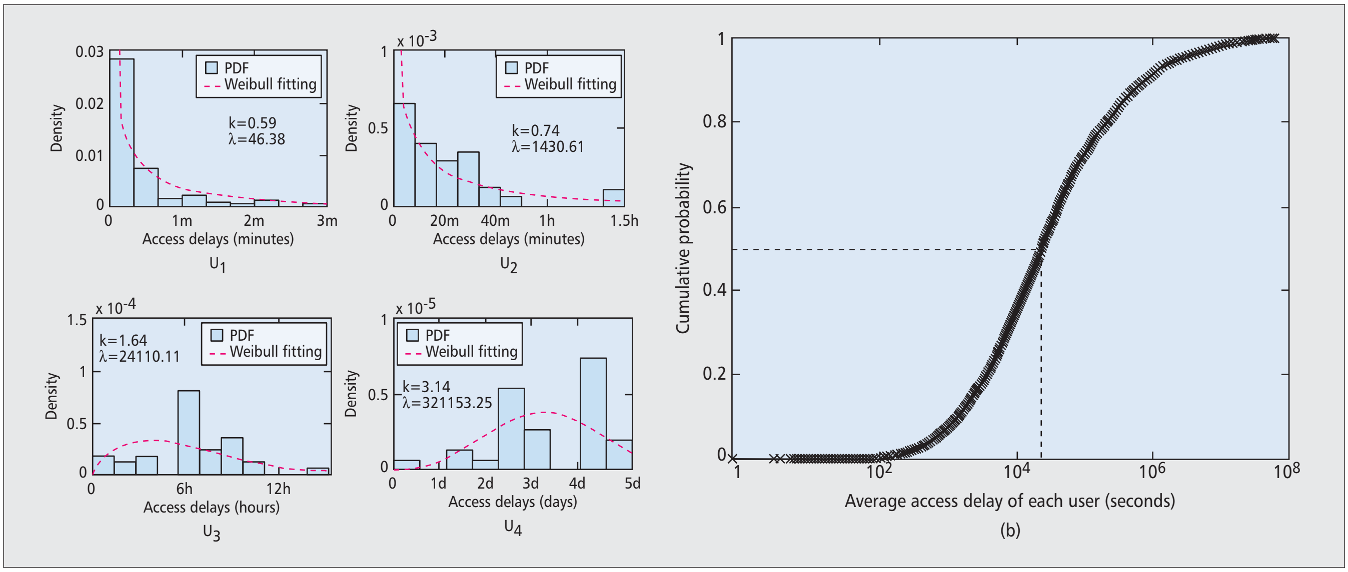

Figure 5. Measurement of access delays: a) four example users; b) access delay statistics of all users.

function of C ($|\vec{p}|$), and the number of satisfied users is also an increasing function of C. The C value that makes $\Sigma \textbf{\textit{Satisfactory}}_i(t)$ approach its maximum will be the standard number of initial pushing seeds for satisfying 100 percent of users. We examine how C can be reduced (for higher offloading gains) if we target the satisfaction of 90 and 80 percent of users.

From Fig. 6, to satisfy 100 percent of all users, **p-H** always finds the best initial pushing vector (i.e., the least number of seeds), and **p-R** performs the poorest, while **p**-$D^{\rightarrow}$ and **p**-$D^{\leftarrow}$ also perform poorly, so simply pushing by node degree is not preferred. In most cases, **p**-$\lambda^*\gamma^{\rightarrow}$ and **p**-$\lambda^*\gamma^{\leftarrow}$ perform the second best, which implies that we can conjunctively consider the I^S and I^M factor by simple multiplication to achieve near-optimal performance. **p**-*Pr* achieves not very good performance compared with strategies by impact factors, as it focuses the connections of the network graph but ignores the historical spreading impact, while our proposed factors (γ) make better sense. In MIT and Infocom traces, λ-based strategies perform better than γ-based ones, which means the mobility factor decides more on the sharing process when nodes are with high mobility. In Beijing and SUVnet traces, γ-base ones perform better, which means the social factor has more control over when nodes have low mobility.

When we target satisfying 90 percent of all users, the required initial pushing ratio is reduced significantly. With simple pushing strategies, for the MIT and the Infocom traces, only 15.4 and 10.5 percent of users need to be the initial seeds on average. The number of initial seeds is further dramatically reduced when satisfying 80 percent of users. Approximately 10 percent initial pushing ratio is needed for all traces except the Beijing trace, which requires about 17 percent initial pushing ratio. The Beijing and SUVnet traces always need relatively larger numbers of initial seeds due to their low contact rates and large user bases.

The implication from Fig. 6 is that when users have relatively higher mobility patterns, the mobility impact will mostly decide the content obtaining delays, but when people are not moving and meeting frequently within a large user base, the online spreading impact needs to be enhanced for initial pushing to offload traffic effectively. Generally, **p-H** is about 15–24 percent better than **p-R** and 12–16 percent better than **p**-λ and **p**-γ; this is a good practical solution to evaluate the multiplication of **p**-λ and **p**-γ for selecting seeds.

Complementary On-Demand Delivery

If a user who has not obtained the content (by initial pushing or sharing) until he/she actually accesses it, we have to deliver it over a cellular link, which is called **on-demand** delivery, and thus the on-demand delivered traffic is not offloaded. We now compare the three target percentages of satisfied users (investigated above) in terms of total offloaded traffic. For example, in the case of 90 percent of satisfied users, 10 percent of remaining users (i.e., those who have not received the content) will access the content via cellular links. Table 1 shows how much traffic is offloaded from cellular links for the three cases where the offloaded traffic ratios of the nine single pushing strategies are averaged, followed with that of **p-H** after /. Note that boldfaced numbers are the highest amount of traffic reduction for each trace across the three target satisfaction cases (i.e., 100, 90and 80 percent). When lowering the percentage of satisfied users from 100 to 90 and to 80 percent, although the initial pushing ratios become reduced, in some cases, the on-demand delivery for the remaining 10 and 20 percent of users may induce the increment of the total cellular traffic instead. Overall, TOSS can reduce from 63.8 to 86.5 percent of the cellular traffic load while satisfying the access delay requirements of all users.

We notice a balance between the traffic reduction due to the initial pushing and the traffic increment by on-demand delivery, as the sat-

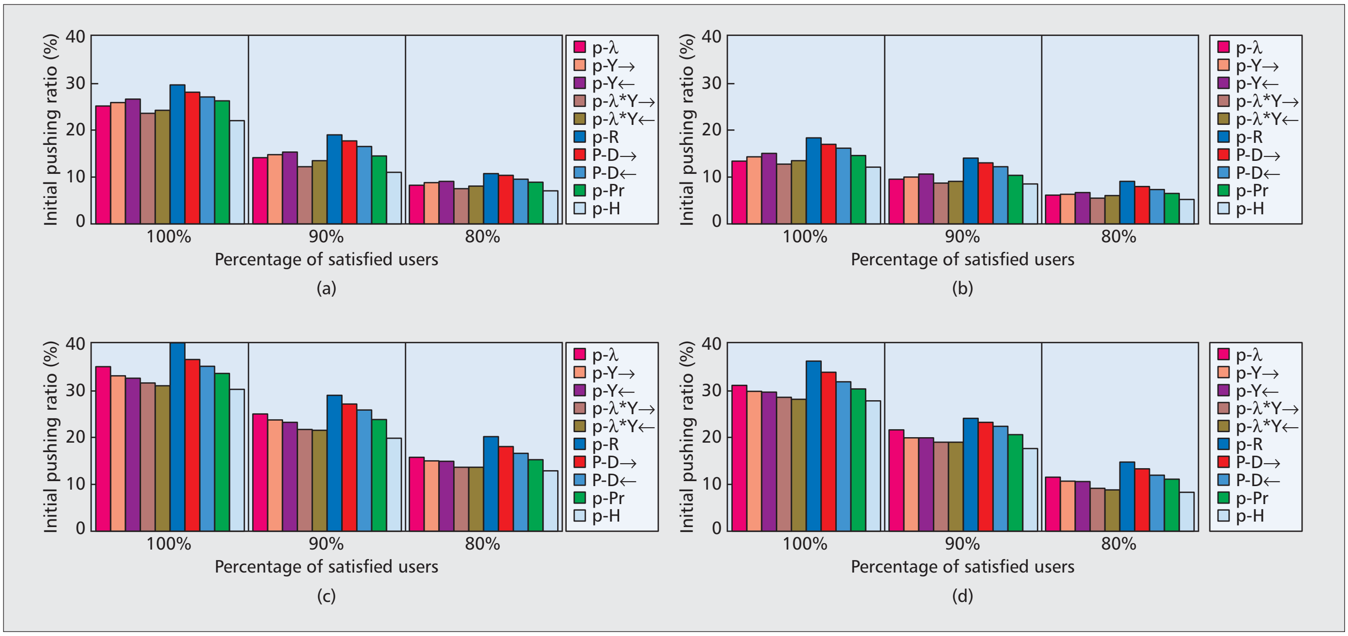

Figure 6. Initial pushing ratios to satisfy 100%, 90%, and 80% of all users: a) MIT; b) Infocom; c) Beijing; d) SUVnet.

isfaction percentage of users changes. It is about how to deal with those users with low I^S and I^M as well as large access delays, which will burden TOSS with selecting the optimal initial seeds, as they are hard to reach by even many hops. Instead, it is better for TOSS to exclude them for a better solution of initial seeds to satisfy other users, and finally carry out on-demand delivery.

Conclusion

In this article, we have introduced the TOSS framework to leverage SNSs for offloading the mobile cellular traffic by opportunistic device-to-device sharing. We have discussed strategies to select the appropriate initial seeds depending on their SNS spreading impact, content access delays, and MSN mobility impact. Trace-driven evaluation reveals that TOSS can reduce 63.8 to 86.5 percent of cellular traffic.

In the future, practical deployment of the TOSS framework mostly relies on cooperation among MNOs, CPs, and SNS providers. Incentive-based business models for the three entities will be very important, by which mobile users can also benefit from sharing content with others. Also, issues of privacy, security, and system scalability need further in-depth investigation.

Acknowledgment

This work is supported by the National Natural Science Foundation of China Project (Grant No. 61300224), the International Science and Technology Collaboration Program (2014DFT10070) funded by China Ministry of Science and Technology (MOST), the Canadian Natural Sciences and Engineering Research Council, and Basic Science Research Program through the "National Research Foundation of Korea (NRF)" funded by the Ministry of Science, ICT & Future Planning (2013R1A2A2A01016562). Professor Lianghai Jin is the corresponding author.

Trace	100%	90%	80%
MIT [9]	73.6/76.3	**74.6/76.9**	70.9/72.2
Infocom [10]	**85.3/86.5**	79.5/80.4	73.4/74.1
Beijing [11]	65.3/68.4	**65.0/68.9**	63.8/65.2
SUVnet [12]	68.5/70.3	**68.7/71.0**	68.3 /70.7

Table 1. Percentage of traffic reduction with on-demand delivery: average of nine simple pushing strategies/heuristic.

References

[1] M. Cha *et al.*, "I Tube, You Tube, Everybody Tubes: Analyzing the World's Largest User Generated Content Video System," *Proc. 7th ACM Internet Measurement Conf.*, San Diego, CA, 2007.

[2] B. Han and A. Srinivasan, "eDiscovery: Energy Efficient Device Discovery for Mobile Opportunistic Communications," *Proc. 20th IEEE Int'l. Conf. Network Protocols*, Austin, TX, 2012.

[3] S. Ioannidis, A. Chaintreau, and L. Massoulie, "Optimal and Scalable Distribution of Content Updates over a Mobile Social Network," *Proc. 28th IEEE INFOCOM*, Rio de Janeiro, Brazil, 2009.

[4] X. Wang *et al.*, "TOSS: Traffic Offloading by Social Network Service-Based Opportunistic Sharing in Mobile Social Networks," *Proc. 33rd IEEE INFOCOM*, Toronto, Canada, Apr. 2014.

[5] J. Yang and J. Leskovec, "Patterns of Temporal Variation in Online Media," *Proc. 4th ACM Int'l. Conf. Web Search and Data Mining*, Hong Kong, 2011.

[6] H. Kwak *et al.*, "What Is Twitter, a Social Network or a News Media?," *Proc. 19th Int'l. World Wide Web Conf.*, Raleigh, VA, 2010.

[7] E. Jaho and I. Stavrakakis, "Joint Interest- and Locality-Aware Content Dissemination in Social Networks," *Proc. 6th Int'l. Conf. Wireless On-Demand Network Systems and Services*, Snowbird, UT, 2009.

[8] P. Hui, J. Crowcroft, and E. Yoneki "BUBBLE Rap: Social-Based Forwarding in Delay-Tolerant Networks," *IEEE Trans. Mobile Computing*, vol. 10, 2011, pp.1576–89.

[9] N. Eagle and A. Pentland, "Reality Mining: Sensing Complex Social systems," *Springer J. Personal and Ubiquitous Computing*, vol. 10, no. 4, 2006, pp. 255–68.

[10] A. Chaintreau *et al.*, "Impact of Human Mobility on Opportunistic Forwarding Algorithms," *IEEE Trans. Mobile Computing*, vol. 6, no. 6, 2007, pp. 606–20.
[11] Y. Zheng *et al.*, "Mining Interesting Locations and Travel Sequences from GPS Trajectories," *Proc. 18st Int'l. World Wide Web Conf.*, Madrid, Spain, 2009.
[12] H. Huang *et al.*, "Performance Evaluation of SUVnet with Real-Time Traffic Data," *IEEE Trans. Vehic. Tech.*, vol. 56, no. 6, Nov. 2007, pp. 3381–96.
[13] R. P. Satorras and A. Vespignani, "Epidemic Spreading in Scale-Free Networks," *Physical Rev. Letters*, vol.86, no. 14, 2001.
[14] D. Towsley, "A Walk in the Dark: Random Walks and Network Discovery," *12th IFIP Int'l. Conf. Networking*, New York, NY, 2013.
[15] L. Gyarmati and T. A. Trinh, "Measuring User Behavior in Online Social Networks," *IEEE Network*, vol. 24, no. 5, 2010, pp. 26–31.
[16] S. Brin and L. Page, "The Anatomy of a Large-Scale Hypertextual Web Search Engine," *Computer Networks and ISDN Systems*, vol. 30, no. 1–7, 1998, pp. 107–17.

Biographies

Xiaofei Wang (xiaofeiwang@ieee.org) received his M.S. and Ph.D degrees from the School of Computer Science and Engineering, Seoul National University (SNU), in 2008 and 2013, respectively. He received his B.S. degree from the Department of Computer Science and Technology, Huazhong University of Science and Technology (HUST), in 2005. He is a recipient of the Korean Government Scholarship for Excellent Foreign Students in the IT Field by NIPA from 2008 to 2011, and also received the Global Outstanding Chinese Ph.D. Student Award in 2012. His current research interests are social-aware multimedia service in cloud computing, cooperative backhaul caching, and traffic offloading in mobile content-centric networks.

Min Chen (minchen@ieee.org)is a professor in the School of Computer Science and Technology at HUST. He was an assistant professor in the School of Computer Science and Engineering at SNU from September 2009 to February 2012. He worked as a post-doctoral fellow in the Department of Electrical and Computer Engineering at UBC for three years. Before joining UBC, he was a postdoctoral fellow at SNU for one and a half years. His research focuses on Internet of Things, machine-to machine communications, body area networks, body sensor networks, e-healthcare, mobile cloud computing, cloud-assisted mobile computing, ubiquitous networks and services, mobile agents, and multimedia transmission over wireless networks, among others.

Ted "Taekyoung" Kwon (tkkwon@snu.ac.kr) is currently a professor in the School of Computer Science and Engineering, SNU. Before joining SNU, he was a postdoctoral research associate at the University of California at Los Angeles (UCLA) and City University of New York (CUNY). He obtained B.S., M.S., and Ph.D. degrees from the Department of Computer Engineering, SNU, in 1993, 1995, and 2000, respectively. During his graduate program, he was a visiting student at IBM T. J. Watson Research Center and the University of North Texas. His research interest lies in sensor networks, wireless networks, IP mobility, and ubiquitous computing.

Lianghai Jin (lianghaijin@hust.edu.cn) received his B.S. degree in computer science from Central South University, Changsha, China, in 1988, his M.S. degree in computer science from Beijing Jiaotong University, China, in 2002, and his Ph.D. degree in pattern recognition and intelligent systems from HUST, Wuhan, China, in 2008. He is currently an associate professor with the School of Computer Science and Technology, HUST. His current research interests include color and video image processing.

Victor C. M. Leung [S'75, M'89, SM'97, F'03] (vleung@ece.ubc.ca) received B.A.Sc. (Hons.) and Ph.D. degrees, both in electrical engineering, from UBC, where he holds the positions of professor and TELUS Mobility Research Chair in the Department of Electrical and Computer Engineering. He is a Fellow of the Royal Society of Canada, Engineering Institute of Canada, and the Canadian Academy of Engineering. He has contributed more than 700 technical papers, 26 book chapters, and 5 books in the areas of wireless networks and mobile systems. He was a Distinguished Lecturer of the IEEE Communications Society. He has served or is serving on the Editorial Boards of *IEEE Transactions on Computers*, *Wireless Communications*, *Vehicular Technology*, *IEEE Wireless Communications Letters*, and several other journals, and has contributed to the Organizing and Technical Program Committees of numerous conferences. He was a winner of the 2012 UBC Killam Research Prize and the IEEE Vancouver Section Centennial Award.