# iDEAL: Incentivized Dynamic Cellular Offloading via Auctions

Wei Dong, Swati Rallapalli, Rittwik Jana, Lili Qiu, Senior Member, IEEE, K. K. Ramakrishnan, Fellow, IEEE, Leo Razoumov, Yin Zhang, Senior Member, IEEE, and Tae Won Cho

Abstract—The explosive growth of cellular traffic and its highly dynamic nature often make it increasingly expensive for a cellular service provider to provision enough cellular resources to support the peak traffic demands. In this paper, we propose iDEAL, a novel auction-based incentive framework that allows a cellular service provider to leverage resources from third-party resource owners on demand by buying capacity whenever needed through reverse auctions. iDEAL has several distinctive features: 1) iDEAL explicitly accounts for the diverse spatial coverage of different resources and can effectively foster competition among third-party resource owners in different regions, resulting in significant savings to the cellular service provider. 2) iDEAL provides revenue incentives for third-party resource owners to participate in the reverse auction and be truthful in the bidding process. 3) iDEAL is provably efficient. 4) iDEAL effectively guards against collusion. 5) iDEAL effectively copes with the dynamic nature of traffic demands. In addition, iDEAL has useful extensions that address important practical issues. Extensive evaluation based on real traces from a large US cellular service provider clearly demonstrates the effectiveness of our approach. We further demonstrate the feasibility of iDEAL using a prototype implementation.

*Index Terms*—Cellular networks, economics, optimization, wireless networks.

#### I. INTRODUCTION

S THE world embraces wireless and mobile technologies, cellular data traffic is growing exponentially, and this trend is expected to continue [13]. Given the scarcity of spectrum resources, it is becoming increasingly expensive for a single cellular service provider to provision sufficient cellular resources to support all its consumers all the time, especially given the significant variability in demand (e.g., cellular traffic follows strong diurnal and weekly patterns [31], [38]). The current best practice is for service providers to augment the cellular network capacity by deploying alternative wireless technologies (e.g., Wi-Fi and femtocells, which potentially have higher capacity but limited communication range) on their own. While this approach is helpful in alleviating the stress on the busiest cellular regions in the short term, it needs to be complemented in

Manuscript received February 06, 2013; revised June 08, 2013; accepted July 04, 2013; approved by IEEE/ACM TRANSACTIONS ON NETWORKING Editor X. Wang. Date of publication August 19, 2013; date of current version August 14, 2014. This work was supported in part by the NSF under Grants CNS-0916106 and CNS-1017549.

W. Dong, S. Rallapalli, L. Qiu, and Y. Zhang are with The University of Texas at Austin, Austin, TX 78712 USA (e-mail: wdong86@cs.utexas.edu).

R. Jana, K. K. Ramakrishnan, L. Razoumov, and T. W. Cho are with AT&T Labs—Research, Florham Park, NJ 07932 USA.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TNET.2013.2273766

the long-term by dynamically utilizing the many Wi-Fi hotspots and femtocells that have already been deployed by third parties.

We propose a solution that enables a service provider to leverage resources on demand from third-party resource owners by buying capacity whenever needed. Measurement studies show that many third-party Wi-Fi hotspots and femtocells have significant spare capacity even during busy hours (e.g., [29] analyzes Wi-Fi utilization in a wide range of scenarios and finds that in all cases the utilization is below 40%). We do not necessarily limit ourselves to considering Wi-Fi resources. For example, when one cellular network is under stress, other cellular service providers in the same area may have spare cellular resources to provide a dynamic roaming service. On-demand purchase of such third-party spare resources can potentially lead to a win-win solution: The cellular service provider achieves significant savings by not having to provision for the peak traffic demands; the third-party resource owners gain additional revenue from the otherwise wasted spare capacity; the overall user experience is also improved. In order for this approach to be successful, however, it is essential to have an incentive framework that can effectively foster collaboration while guarding against nontruthful and collusive behavior.

Our Approach: Incentivizing Cellular Offloading via Auctions: We propose iDEAL, a novel auction-based incentive framework to enable dynamic offloading of cellular traffic. In iDEAL, a cellular service provider purchases bandwidth on demand from third-party resource owners, who may be a Wi-Fi hotspot owner, a femtocell owner, or another cellular service provider. This auction problem is naturally formulated as a *reverse auction*, where the goods of interest are bandwidth resources, third-party hotspot owners serve as sellers (i.e., bidders or auctioneers) and submit their bids while provider Aor a trusted third party serves as an auctioneer, who evaluates the bids from all hotspot owners and makes decisions regarding whose services to purchase in order to satisfy A's traffic demands and minimize A's total cost. Each bidder submits a bid that specifies the total amount of bandwidth it offers in the next time interval and the unit price it asks for. After collecting all the bids, the cellular service provider determines: 1) an allocation, i.e., how to allocate its traffic between different third-party resource owners (depending on the region they cover) and its own cellular network; and 2) a price, i.e., how much it pays each third-party resource owner that offloads cellular traffic.

The use of reverse auction is motivated by the following observations. First, a key challenge in utilizing resources from third-party resource owners is that we do not know their cost function. Their cost function may be based on multiple considerations, some of which may not be revealed to the cellular

1063-6692 © 2013 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications standards/publications/rights/index.html for more information.

service provider. Reverse auctions provide a formal framework for third-party resource owners to express the price they demand and for the cellular service provider to optimize the allocation based on the received bids. Second, by using reverse auctions, the cellular service provider avoids having to negotiate a long-term bilateral agreement with each individual third-party resource owner. Negotiating such long-term agreements is difficult and possibly inefficient due to dynamic traffic demands and resource availability. Instead, the cellular service provider can now establish short-term contracts with third-party resource providers. It also potentially cuts costs by leveraging competition across third-party resource owners. Third, reverse auctions can be incrementally deployed today, yielding savings to the cellular service provider even when only a subset of third-party resource owners participate.

Unique Challenges: While reverse auction has been applied to cellular offloading in the past (e.g., [12]), our problem setting poses several unique challenges. Despite their importance, none of these challenges has been considered earlier.

- Diverse spatial coverage: Cellular resources can serve traffic anywhere in a cell sector (albeit at different rates depending on path loss etc.), whereas Wi-Fi hotspots and femtocells have a much more limited communication range, making it essential to consider the spatial coverage of different resources. However, one cannot simply partition resources into separate regions and launch independent reverse auctions within each region because the longer-range cellular resource introduces coupling between the Wi-Fi hotspots or femtocells in different regions. For example, buying more resources from a cheaper Wi-Fi hotspot in one region frees up more cellular resources, which reduces the amount of cellular traffic to be offloaded in regions with more expensive Wi-Fi hotspots.
- Traffic uncertainty: Cellular traffic is highly dynamic and unpredictable. Since the cellular service provider has to purchase third-party resources based on predicted traffic demands at a future time, it can easily result in underprovisioning or overprovisioning without an effective technique to cope with traffic uncertainties. In contrast, in conventional reverse auction settings, the total amount of goods that the buyer wants is typically known a priori.
- Nontruthful bidding and collusion. It is essential for us to explicitly guard against both nontruthful bidding and collusion. Due to the distributed nature of hotspot locations, collusion in our context is quite different from what was studied previously and calls for a new study to understand possible collusion strategies and mitigate them.

Contributions: Our paper makes three main contributions.

 We design the iDEAL incentive framework to effectively address the above unique challenges. Compared to conventional mechanisms for reverse auctions, iDEAL has the following distinctive features: 1) iDEAL explicitly accounts for the spatial coverage of different resources and can effectively foster competition among third-party resource owners in different regions, resulting in significant savings to the cellular service provider. 2) iDEAL incentivizes bidders (i.e., third-party resource owners) to participate in the reverse auction and to be truthful in their bidding. 3) iDEAL is provably efficient in that the winners are the bidders who have the lowest valuation of their resources. 4) iDEAL can effectively mitigate collusion. 5) iDEAL can effectively cope with the highly dynamic nature of traffic demands.

- 2) We present useful extensions to iDEAL: 1) support general bidding curves, which gives a hotspot owner the flexibility to submit its ask price in the form of a curve, such that different unit price is used when different amount of capacity is sold; 2) support femtocell offloading and dynamic roaming; 3) incorporate quality-of-service consideration (in addition to cost); 4) potentially delay demands that are delay-tolerant to further reduce cost and improve efficiency; and 5) determine which users' traffic to offload.
- 3) We extensively evaluate iDEAL using simulation based on real traces from one of the largest US cellular service providers. Our results clearly demonstrate the effectiveness of our approach. We further demonstrate the feasibility of our approach using a simple prototype implementation.

# II. PROBLEM FORMULATION

In this section, we formulate the problem of offloading cellular traffic as a *reverse auction*. The offloading is transparent to clients and does not affect cellular pricing (i.e., users pay for the data usage regardless of whether it is carried by the cellular provider or third-party resource owners).

*Basic Auction Settings:* Consider a cellular network *A* that is interested in purchasing and leveraging spare resources from third-party Wi-Fi hotspots to satisfy traffic demands from its customers. The third-party hotspot owners should be rewarded for opening up their services to *A*'s customers. To facilitate such cooperation, provider *A* can set up an auction to let third-party hotspot owners submit bids to offer their network resources, e.g., dollars per bit rate for unit time (e.g., 1 h) that a third-party hotspot owner offers.

This problem is naturally formulated as a *reverse auction*. Since the demand changes over time, e.g., due to diurnal variations [31], the auction takes place periodically or whenever demand changes. The auction frequency is chosen to balance the overhead and the accuracy of traffic demand estimation.

The cellular network is shared across a relatively large area typically called a cell site. A site is further subdivided into three or more sectors. Cellular resource in different sectors is relatively independent, so we only consider a single sector. The same solution can be applied independently to other sectors. The sector can be considered to be divided into *m* small regions based on locations of Wi-Fi hotspots and Wi-Fi range as shown in Fig. 1. A Wi-Fi hotspot can satisfy traffic demands only in its region.

*Naive Solution:* A simple approach is to statically partition the cellular resource into different regions and determine the amount of Wi-Fi resource needed in each region based on the amount of user demand in the region. Then, we conduct a local auction within a region to utilize the cellular resource and Wi-Fi resources dedicated to the region. We call it *static local auction*. While simple, this approach has several important limitations: 1) Due to limited Wi-Fi coverage, the number of hotspots in a region is limited, i.e., the competition is limited. However, adequate competition is essential for an auction-based approach to be effective. 2) This formulation treats different regions equally,

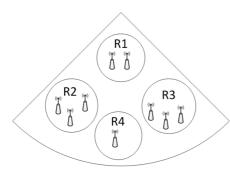


Fig. 1. Sample cellular sector and its Wi-Fi regions.

however the service provider may view different regions differently because different regions may have different spectrum efficiencies due to different signal-to-interference-plus-noise ratio (SINR) from the base station. 3) The static allocation cannot effectively take into account the available Wi-Fi resources and their bids across different regions. For example, even when a region has higher traffic demand, we may or may not need to allocate more cellular resources to the region depending on: a) how many Wi-Fi hotspots are in the region; b) what are their prices; and c) how the Wi-Fi hotspots and their prices compare with those in other regions. If there are more Wi-Fi hotspots in a region offering cheaper bids than in the other regions, we can allocate less cellular resources.

Design Goals: We seek an auction scheme to: 1) account for different spatial coverage of resources, which has not been considered in existing work; 2) cope with dynamic traffic demands; 3) achieve high efficiency, where the winners in the auction are the hotspot owners who really can provide the service at a cheaper price, thereby improving the overall system efficiency and social welfare; 4) promote truthful bidding to prevent bidders from gaming the system, effectively discover price to ensure that the overall system is efficient, and avoid unnecessary system fluctuation due to gaming, as unwanted switching between Wi-Fi and 3G can negatively impact user experience [17]; 5) lower cost, which is natural but is challenging to achieve simultaneously with truthfulness; and 6) guard against collusion.

#### III. OUR SOLUTION: iDEAL

In this section, we introduce our solution: iDEAL. We start by designing the auction setting that fosters more competition and captures the service provider's regional preferences. Then, we describe the two stages of iDEAL: 1) *allocation*, i.e., determine how to allocate traffic among third-party resource owners and the cellular network itself to minimize cost given the bids; 2) *pricing*, i.e., decide how much should be paid to individual third-party resource owners in order to provide enough incentives for them to be truthful. Optimal allocation does not depend on pricing, but assumes all sellers are truthful. Pricing depends on the allocation and is designed so that staying truthful is the seller's optimal strategy. Table I summarizes the key notations.

# A. iDEAL Auction Setting

*Third-Party Wi-Fi Resources and Bids:* Suppose *n* thirdparty hotspot owners offer their resources to the cellular service provider by submitting their bids. Let  $A_j = \{\lambda_j, p_j\}$  denote

TABLE I NOTATIONS number of regions in a cellular sector mnumber of sellers in a cellular sector n $d_i$ traffic demand in region icellular capacity in region i Ci spectrum efficiency of cellular network in region i $e_i$ total cellular spectrum usage:  $z = \sum_{i=1}^{m} c_i/e_i$ z $x_{i}$ total capacity bought from seller jthe unit price seller j asks for  $p_j$  $\overline{\lambda}$ the Wi-Fi capacity offered by seller j F(z)cellular cost function the region that seller j belongs to

hotspot owner j's bid, which indicates hotspot owner j wants to sell  $\lambda_j$  amount of bandwidth at a price  $p_j$  per bit per second. The bids are *nonatomic* (i.e., a hotspot owner is willing to sell a part of the capacity it offers). Function f(j) returns the region where hotspot owner j sells its capacity (e.g., f(j) = i means hotspot owner j sells its capacity in region i). For simplicity, we assume that each hotspot owner j sells capacity in a single region, i.e., regions do not overlap. In Section IV, we show how to extend our approach to support overlapping regions. As Wi-Fi may not cover the whole sector, areas without Wi-Fi coverage can be treated as special regions with no Wi-Fi bids.

Cellular Resources as a Virtual Bid: Let the traffic demand vector be  $D = \{d_1, d_2, \ldots, d_m\}$ , where  $d_i$  is the demand in region *i*. In order to effectively leverage both third-party and cellular resources, we let the service provider also participate in the auction by submitting a virtual bid. The virtual bid is in the form of a cost function F(z), where z is the total amount of spectrum used in the entire cellular sector. Let  $c_i$  be the cellular capacity in region *i*, and let  $x_j$  be the total capacity bought from hotspot owner *j*. To satisfy the cellular traffic demand  $d_i$  in each region *i*, we must have:  $c_i + \sum_{j:f(j)=i} x_j \ge d_i$ . Since different regions may have different spectrum efficiency, we denote the actual spectrum usage in region *i* as  $c_i/e_i$ , where  $e_i$  is the spectrum efficiency in region *i*. Thus, the total spectrum usage is  $z = \sum_{i=1}^m c_i/e_i$ .

We consider F(z) to be a piecewise linear convex function, capturing the fact that, below a certain value, the cost (reflecting sunk cost [34]) is very low because the service provider has already invested in buying the spectrum and needs to keep the system running; as the cellular network becomes more loaded, the cost increases; and once it is overloaded, the cost increases sharply to capture the high cost of congestion. Thus, z is not limited by the available spectrum and can go to infinity. A similar convex cost function has been widely used in modeling congestion cost in the Internet (e.g., [18], [28], and [30]).

Because the cellular resource in the virtual bid can be used in any region in the sector, it introduces coupling between the regions. The entire sector can now be viewed as one auction instead of several independent ones as in the naïve solution. Even if the number of hotspots in one region is small, its hotspots are not guaranteed to win since the auction may buy more Wi-Fi from other regions and save the cellular resource for this region, i.e., hotspots compete not only within their regions, but also across regions. We now see a new type of competition, which we call *interregion competition* in addition to *intraregion competition*.

*Auction Objective:* The goal of the cellular service provider is to minimize the total Wi-Fi and cellular cost, while satisfying

Fig. 2. Problem formulation to optimize allocation.

the customers' demands (i.e.,  $c_i + \sum_{j:f(j)=i} x_j \ge d_i$ ) and offering appropriate incentives to the third-party Wi-Fi hotspot owners to share their resources.

#### B. Preparation: (Static) Global Allocation

We first ignore traffic variations and develop techniques to effectively utilize both cellular and Wi-Fi resources in serving user traffic demands.

We formulate a global resource allocation problem as a linear program in Fig. 2. The formulation effectively captures global cellular resources and local Wi-Fi resources by treating the cellular resource as a single resource with a single bid. As shown, our goal is to minimize the sum of total Wi-Fi cost (based on their bids) plus cellular cost F(z). The constraint [C1] ensures that we have enough Wi-Fi and cellular resources to satisfy traffic demands in each region *i*. The constraint [C2] relates the cellular capacity with the cellular spectrum. The constraints [C3] and [C4] put upper and lower bounds on  $x_j$ and  $c_i$ . Since there is no upper bound on *z*, there is always a feasible solution. When *z* increases beyond the available spectrum, F(z) grows rapidly to reflect high congestion cost. This problem can be solved efficiently using linear program solvers (e.g., CPLEX).

#### C. iDEAL Dynamic Global Allocation

Traffic demand changes over time and is challenging to predict accurately. Based on the history of observed demand vectors, we can optimize for the representative demand vectors that are likely to occur in the next time interval. Our goal is to find the allocation to minimize the worst-case cost for these representative demand vectors.

Algorithm: Formally, suppose there are K historical demand vectors, denoted as  $D_k = (d_{k1}, d_{k2}, \dots, d_{km})$   $(k = 1, \dots, K)$ , where  $d_{ki}$  denotes the kth possible demand in region i  $(i = 1, \dots, m)$ . While it is difficult to predict accurately the demand vector for the next time interval, it is common in robust traffic engineering to assume that the demand vector for the next time interval is covered by the convex hull of all the historical demand vectors  $D_k$  [30]. Under this assumption, we can minimize the worst-case cost while satisfying all possible demands that may arise in the next time interval. We formulate this dynamic global allocation problem by modifying the LP formulation in Fig. 2. In particular, we change [C1] and [C2] to the following:

$$\begin{bmatrix} \text{C1-dynamic} \end{bmatrix} \qquad \sum_{f(j)=i} x_j + c_{ki} \ge d_{ki} \qquad \forall \ k \text{ and } i$$
$$\begin{bmatrix} \text{C2-dynamic} \end{bmatrix} \qquad \sum_i \left(\frac{c_{ki}}{e_i}\right) = z \qquad \forall \ k$$

to ensure that we have enough cellular and Wi-Fi resources to satisfy all possible demand vectors. This is much more efficient than provisioning for the peak demand in each region.

From now on, we will refer to our dynamic global allocation algorithm as *iDEAL*, and the static global allocation algorithm as *iDEAL* (*static*).

*Property:* A nice property of this dynamic global allocation is that it effectively leverages the global cellular resource on demand to satisfy different possible traffic demands. In particular, while the total cellular resource is fixed, the amount of cellular resource used in each region can change according to the real demand. When demand shifts from one region to another over time, the same global cellular resource can be used instead of provisioning for the peak demand in each region. Therefore, global cellular resource has a distinctive advantage over local Wi-Fi resources in satisfying time-varying demand, which we explicitly leverage in our formulation.

# D. iDEAL Pricing Solution

As discussed in Section II, we want the pricing scheme to be truthful and efficient. Meanwhile, we want the pricing scheme to fully benefit from the interregion competition. For example, when hotspots in one region lower their bids and offload more traffic, this would reduce the demand for third-party resources in other regions and cause hotspots in other regions to sell less. To capture this unique interaction between intraregion and interregion competition, we cannot treat auctions in different regions as separate auctions and compute pricing separately; instead, we must consider them as a single auction and explicitly incorporate interregion competition into the payment computation.

The Vickrey–Clarke–Groves [35] auction is well known. It is both truthful and efficient. It pays a winner the opportunity cost that the presence of the winner introduces on the other players. VCG has a major weakness—its cost is generally high [7]. However, in our setting, VCG is able to capture the interregion competition, which lowers the cost. Thus, to preserve the nice properties of VCG (i.e., truthfulness and efficiency) while achieving low cost, we apply the VCG principle globally over the whole cellular sector and compute the *global opportunity cost* to capture both interregion and intraregion competition.

Algorithm: We follow the general VCG principle and compute the global opportunity cost as follows. Let V(D, N)denote the valuation consumed in the optimal allocation. D is a demand matrix containing K demand vectors  $D_k = \{d_{k1}, d_{k2}, \cdots, d_{km}\}$   $(k = 1, \cdots, K)$ , which specify possible demands in each region. N is the set of bidders (including the cellular service provider). Given the result of the allocation scheme, if we buy t capacity from winner bin region r, the amount of money we pay to b will be  $V(D, N \setminus \{b\}) - V(D^1, N \setminus \{b\})$ , where  $D^1$  is derived from D by setting  $d_{kr} = \max(0, d_{kr} - t)$  for each k and  $N \setminus \{b\}$  is the set of remaining bidders after removing bidder b. Thus,  $V(D^1, N \setminus \{b\})$  is the total value sold by other bidders under the current optimal allocation;  $V(D, N \setminus \{b\})$  is the total valuation optimized after removing b. The difference is the global opportunity cost b imposed on other bidders.

Next, we use an example to show how global opportunity cost is computed and how the interregion competition can help reduce cost. Fig. 3 shows two regions R1 and R2, each with 1 unit

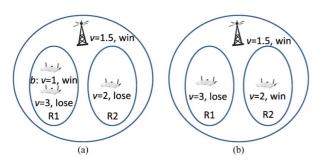


Fig. 3. Global opportunity cost example. (a) Optimal allocation. (b) Optimal allocation after removing the winning hotspot.

demand. R1 has two hotspots with valuations 1 and 3, respectively. R2 has one hotspot with valuation 2. Each hotspot has 1 unit resource. The cellular resource is 1 unit and is worth 1.5. The optimal allocation is shown in Fig. 3(a): 1 unit of Wi-Fi in region 1 with valuation 1 and 1 unit of cellular resource in region 2. To compute the global opportunity cost for the Wi-Fi winner b, we remove b and compute the optimal allocation without the winner as shown in Fig. 3(b). The new allocation should use all the cellular resource in region 1 and the Wi-Fi resource with valuation 2 in region 2. The total valuation sold by other bidders is thus 1.5 + 2 = 3.5, while in the original allocation the number is 1.5. Thus, the global opportunity cost we pay to b is 3.5 - 1.5 = 2. In comparison, with the same allocation, if we apply VCG in each region separately, the local opportunity cost is 3 since region 1 has only the Wi-Fi resource with valuation of 3 after we remove b. This shows that global opportunity cost is lower since it effectively takes into account resources across all regions. Note that this notion of global opportunity cost and its computation work for both static and dynamic global allocations. The two versions only differ in the allocation as described in Sections III-B and III-C.

*Properties:* iDEAL inherits the following three important properties from VCG: 1) bidders have incentives to be *truthful*; 2) the outcome of the auction is *efficient*; and 3) the auction is *individually rational*, meaning third-party resource owners have incentives to participate in the auction. Formally, we have the following three theorems.

*Theorem 1:* In iDEAL, truth telling is an optimal strategy.

*Proof:* Pick arbitrary bidder  $b_1$  in region  $r_1$ , and let its valuation for unit capacity (e.g. 1 b/s) be  $v_1$ . Suppose when  $b_1$  bids truthfully, it sells  $t_1$  amount of capacity. Its utility is

$$U1 = \left(V\left(D, N \setminus \{b_1\}\right) - V\left(D^1, N \setminus \{b_1\}\right)\right) - v_1 \cdot t_1$$

i.e., the difference between the payment it receives and its valuation for the amount it sells, where  $D^1$  is derived from D by setting  $d_{kr_1} = \max(0, d_{kr_1} - t_1)$  for each snapshot k.

If  $b_1$  bids untruthfully and sells  $t_2$  amount of capacity, its utility is:  $U2 = (V(D, N \setminus \{b_1\}) - V(D^2, N \setminus \{b_1\})) - v_1 \cdot t_2$ , where  $D^2$  is derived from D by setting  $d_{kr_1} = \max(0, d_{kr_1} - t_2)$  for each k.

In order to prove truth telling is an optimal strategy, we need  $U1 \ge U2$ . It is evident that

$$U1 - U2 = [v_1 \cdot t_2 + V(D^2, N \setminus \{b_1\})] - [v_1 \cdot t_1 + V(D^1, N \setminus \{b_1\})]$$

Here, the first term is the minimum total valuation needed when we buy  $t_2$  from  $b_1$ , and the second term is the total valuation used in the optimal allocation. Since the second term is optimal, the first term cannot be smaller, and thus  $U1 \ge U2$ .

*Theorem 2:* iDEAL is efficient, which means when bidders are rational, the winners are the bidders whose valuation for their resources is the least.

*Theorem 3:* iDEAL is individually rational, i.e., bidders of the auction will get nonnegative utility, assuming a bidder does not bid lower than his valuation.

Theorem 1 indicates that it is beneficial for a bidder to bid truthfully regardless of other bidders' strategies. Theorem 2 follows from the truthfulness property and our allocation, which minimizes the total valuation assuming everyone bids truthfully. Theorem 3 guarantees that winners will be paid no less than their valuation.

While Theorem 3 is easy to see in normal settings, it is less straightforward with our dynamic allocation because, in the dynamic allocation, the total amount of resource we buy is not fixed. Specifically, when computing the opportunity cost, we remove a winner and compute a new allocation and use the bid(s) of the newly admitted winner(s) as the payment. While the unit prices of the newly admitted bids are not lower than the winner's, the total amount of capacity we buy in the new allocation might reduce. This is because the new allocation may buy more cellular resource, which can be used everywhere and may reduce the need for Wi-Fi in all regions. That makes it hard to tell if the opportunity cost is higher than the winner's valuation. We prove the theorem using contradiction: If we remove a winner w and the amount of increased valuation we buy from others (i.e., the opportunity cost) is less than what w sells, then w should not have won.

# E. Understand and Guard Against Collusion

In this section, we first identify potential collusion strategies in iDEAL and show how they differ from those in normal VCG settings. We then discuss how to mitigate such strategies. We call a set of hotspots colluding together a *bidding ring*. A bidding ring colludes by adopting a certain bidding strategy to maximize utility, i.e., the difference between the payment and the true valuation of the resource sold.

1) Collusion Strategies: Due to the distributed nature of hotspot locations, collusion in our context is quite different from collusion in normal settings, where the optimal collusion strategy is to let one proxy bidder buy (or sell in an reverse auction) for the whole bidding ring [8]. However, in our system each hotspot submits a separate bid. This forbids hotspots to collude optimally and thus may resort to other collusion strategies identified below. In particular, we consider two types of collusion: 1) single seller collusion, whose objective is to maximize the total utility of all hotspots owned by this seller, and 2) multiseller collusion, where each seller colludes with other sellers, but tries to solely maximize its own utility.

In both types of collusion, a bidding ring can drive up the price and increase its utility by *supply reduction* (i.e., drop losing bids or reduce the capacity offered in winning bids, which is equivalent to bidding an extremely high price for the capacity that is removed from bidding). Supply reduction can drive up price because it increases the opportunity cost, which is determined by the immediate losing bids.

2) Mitigating Collusion: We mitigate collusion as follows.

Bidding as a Group to Address Single-Seller Collusion: A single seller with multiple hotspots has an incentive to reduce supply because its hotspots submit separate bids. The opportunity cost of one hotspot can be affected by the price/availability of its other hotspots. Thus by strategically dropping some of its hotspots or raising their prices, it can increase its revenue. This strategy is especially harmful as it may also increase the opportunity cost of other sellers' hotspots. Ultimately, it incurs a higher cost to the service provider.

To address the issue, we let the hotspots owned by the same entity bid as a group, i.e., the seller who owns multiple hotspots discloses all its hotspots, and we consider them as a single bidder in the auction. The seller has an incentive to choose this option, since bidding truthfully is an optimal strategy (Theorem 1). It is also preferred from the service provider's perspective because it only removes competition within the group. The hotspots in this group still compete with hotspots of other sellers, which helps to bring down the cost.

Dynamic Demands in Multiseller Collusion: In order to benefit from supply reduction, a bidding ring needs to accurately predict which bids may lose and drop them. Without that, supply reduction can cause harm by letting the bidding ring miss opportunities to win. Making such predictions is challenging due to the dynamic nature of the traffic demand and Wi-Fi availability. Therefore, in practice, supply reduction does not necessarily increase the utility of the hotspots, which can discourage them from colluding.

*Stability of Multiseller Collusion:* When multiple parties are involved in collusion, a natural question is whether the collusion is stable (i.e., all members of the bidding ring have incentives to stay in the ring [8], [11]). Reference [8] shows that in normal settings, collusion in VCG is stable under certain assumptions. However, their conclusion does not apply to our context because of the difference in collusion strategies. Specifically, we make the following two observations.

First, without utility sharing, members of a bidding ring have an incentive to leave the ring (i.e., do not conduct supply reduction). Formally, we have the following lemma.

*Lemma 1:* Without utility sharing, for bidding ring members, no supply reduction is a (weakly) dominant strategy (i.e., no worse than supply reduction).

This follows from the truthfulness of VCG and the fact that different sellers submit separate, sealed bids and cannot pose as one entity in our system.

Second, the condition of "no utility sharing" is likely to hold in practice due to difficulties of estimating utility obtained from collusion in our system. One reason is that traffic demands and Wi-Fi availabilities are highly dynamic, which makes it hard to attribute utility changes to collusion. Moreover, using sealed bids makes it hard to validate the behavior of other members in the bidding ring. We can make it even harder through system design such as delayed payment (e.g., paying the hotspots every week even though the auction is conducted hourly), which further obfuscates the utility.

# **IV. PRACTICAL CONSIDERATIONS**

Allowing General Bidding Curves: Section III assumes the cellular cost F(z) is convex, and hotspot owners can bid only one price value  $p_j$  for the entire capacity they offer and are

willing to sell part of the capacity at a prorated amount. Now we consider more general cost functions, including convex, concave (capturing economy of scale), or a combination. For example, a third-party hotspot owner may want the flexibility of specifying an entire bidding curve characterizing the ask price with respect to the amount of offloaded traffic. The formulation shown in Fig. 2 remains the same, except that the optimization is no longer a linear program or convex program due to the more flexible  $p_i$  and F(z).

When only F(z) is nonconvex and hotspot owners still bid only one price value for the entire capacity they offer, the problem is easy to solve. We approximate F(z) as a piecewise linear function and solve the optimization problem by enumerating all possible line segments that z belongs to (because within each segment F(z) is still a linear function). Essentially, we solve a linear programming problem for each possible line segment, and then pick the best result from all the corresponding linear programs.

To support general Wi-Fi hotspot cost functions, we use dynamic programming. We introduce discrete step sizes s and s'that quantify the smallest cellular and Wi-Fi capacity unit to purchase, respectively. Small step sizes give better solution but increase running time. We build a table T, where each entry T(k, z) gives the cost of satisfying the demands of the first k regions:  $1, 2, \ldots, k$  using cellular capacity equal to z

$$T(k,z) = \min_{v} \left\{ T(k-1, z-v) + F(z) - F(z-v) + auctionCost(k, d_k - v \cdot e_k) \right\}$$

where auctionCost(k, x) is the minimum cost for satisfying x amount of demand using Wi-Fi in region k,  $d_k$  is the demand in region k,  $e_k$  is the spectral efficiency in region k, and both z and v are multiples of step size s.

We compute *auctionCost* using Dynamic Programming as follows:

$$W_k(i, y) = \min \{ W_k(i - 1, y - u) + cost(i, u) \}$$

 $W_k$  is the Wi-Fi cost table for region k. Here,  $i = 1, 2, ..., N_k$ , where  $N_k$  is the number of bidders in region k. The amount of capacity we seek to satisfy, y, varies from 0 to x, where x is the total amount of capacity we want to satisfy with Wi-Fi in region k. cost(i, u) is simply the cost of satisfying u amount of demand using only bidder i at his bid price. The expression indicates the cost of using i hotspots to satisfy y demand in region k is the minimum cost of using i - 1 hotspots to satisfy y - u demand plus the cost of using ith hotspot to serve u demand. y and y are multiples of step size s'. We then have  $auctionCost(k, x) = W_k(N_k, x)$ .

Supporting Offloading to Femtocells and Dynamic Roaming: In addition to third-party Wi-Fi hotspots, femtocells and other cellular networks can also be used for offloading. Roaming to other cellular networks considered here is different from traditional roaming. Traditional roaming is enabled only outside the current cellular provider's coverage area, whereas dynamic roaming in our context can take place within the coverage area to reduce congestion. In order to support offloading to different types of technologies, we need to effectively handle partially overlapping spatial coverage, as different resources have different coverage ranges. This requires changes to the allocation algorithm. We extend our approach to support these scenarios by dividing overlapping regions into multiple nonoverlapping regions and allowing one provider to belong to multiple regions. The constraint [C1] in Fig. 2 is then replaced by the following two new constraints:

$$\sum_{j:i\in f(j)} x_{ji} + c_i = d_i \qquad \forall i = 1, 2, \dots, m$$
$$\sum_i x_{ji} = x_j \qquad \forall j = 1, 2, \dots, n$$

where  $x_{ji}$  is the amount of capacity bought from seller j and used in region i. This extension can not only support offloading to different types of networks, but also allow a hotspot provider to use its resources across different regions (e.g., hotspots belonging to a single restaurant chain spread across different regions but sharing the same bottleneck capacity).

Incorporating Quality Score: The cellular service provider may prefer some hotspots over others due to different quality (e.g., to avoid hotspots that do not guarantee the amount of capacity they offer). In this case, we can differentiate which hotspots to use based on the quality score  $q_i$   $(0 < q_i \leq 1)$ of hotspot *i*. The higher the score, the better the quality and the easier the hotspot can win in future auctions. To achieve that and ensure the auction is still truthful and individually rational, we change the objective function in the allocation phase to  $\sum_{i} (x_i \cdot p_i/q_i) + F(z)$ , which essentially increases the bid of hotspots with low-quality scores and makes it harder for them to win. We also change the payment for winner j to  $q_j$  times the opportunity cost so that individual rationality is still preserved because the opportunity cost is no less than  $x_i \cdot p_i/q_i$ . It is not difficult to see that the auction is still truthful since the quality scores are bid-independent.

Benefiting From Delay-Tolerant Demands: Some application traffic (e.g., e-mails) is delay-tolerant. A natural way to take advantage of such traffic is to delay them when it is too costly to satisfy them immediately (e.g., when the current traffic load is very heavy or when most traffic demands originate from outside the Wi-Fi coverage areas and have to be satisfied by only the cellular network).

Our framework is flexible enough to support this new optimization task. Consider traffic demands in m snapshots, namely  $D = \{d_{k,i}, d_{2,i}, \ldots, d_{k,i} \ldots\}$ . We can optionally delay a certain demand in snapshot *i* to snapshot *j*, where i < j. The resulting demand (called final demand) becomes  $D' = \{d_1 - \delta_1, d_2 - \delta_2 +$  $\delta_{1,2}, \ldots, d_i - \delta_i + \sum_{j < i} \delta_{j,i}, \ldots, d_k + \sum_{j < m} \delta_{j,m}$ , where  $\delta_i$  denotes the total amount of traffic in snapshot *i* delayed to future snapshots and  $\delta_{i,j}$  denotes the amount of traffic in snapshot i delayed to snapshot j. It is easy to see  $\delta_i = \sum_{j>i} \delta_{i,j}$ . Our goal is to satisfy D' during every interval *i* while minimizing the total cost of satisfying the demands over all intervals plus the penalty incurred in delaying traffic. This can be formulated as an LP problem shown in Fig. 4. The objective reflects the cost of satisfying the final traffic demand in all snapshots plus the penalty associated with delaying traffic. [C1] enforces that the final traffic demand is satisfied during every snapshot. [C2] captures total cellular spectrum usage in snapshot k. [C3] and [C4] provide bounds on  $x_{k,j}$  and  $c_{k,i}$ . This extension changes the allocation algorithm in iDEAL, but the same pricing algorithms and proofs in Section III-D still apply.

$$[Input: d_{k,i}, e_i, \lambda_{k,j}, p_{k,j}, F(z)$$

$$Output: x_{k,j}, c_{k,i}, z_k, \delta_i, \delta_{j,i}$$

$$minimize: \sum_k (\sum_j p_{k,j} x_{k,j} + F(z_k)) + penalty \sum_{k>j} (k-j) \delta_{j,k}$$

$$subject to:$$

$$[C1] \sum_{\substack{i:f(i)=i}} x_{k,j} + c_{k,i} = d_{k,i} - \delta_{k,i} + \sum_{j \le k} \delta_{j,k,i} \quad \forall k, i$$

 $\begin{array}{ll} [C2] & \sum_{i=1}^{j < j < \kappa} & \forall k, i \\ [C3] & 0 \le x_{k,j} \le \lambda_{k,j} & \forall k, j \end{array}$ 

$$[C4] \quad 0 \le c_{k,i} \qquad \qquad \forall k,$$

Fig. 4. Problem formulation to optimize allocation. The variables are defined in Table I, except that now some variables have the additional subscript k to denote their values in snapshot k.

Selecting Users to Offload: So far, we have considered how much capacity to buy from each Wi-Fi hotspot. Now we study which users should be offloaded to a particular Wi-Fi hotspot. We prefer a scheme that minimizes the switching time (e.g., avoid offloading users who will soon leave the hotspot or have little traffic to send) while fully utilizing the purchased capacity at the hotspot. We analyze the Wi-Fi traces from the large cellular provider in the US, which also provides Wi-Fi services, and find that a user who has stayed at a hotspot for a longer time in the past is more likely to stay longer in the future. Therefore, we propose a simple heuristic, which is to offload a user to the hotspot if: 1) the user has already stayed there for at least a threshold amount of time (e.g., 5 s); and 2) his estimated traffic demand is lower than the residual purchased capacity at the hotspot. The condition 2) tries to avoid switching the user back and forth between the Wi-Fi hotspot and the cellular network. Thus, users who are likely to soon leave the hotspot do not have to incur the Wi-Fi switching overhead.

Supporting Unsplittable Demand: A single device usually can only connect to one hotspot at a time. To capture such unsplittable demand, we add the following constraint to the formulation in Fig. 2: The demand of each device has an indicator  $t_{kj}$ , where  $t_{kj} = 1$  when demand  $b_k$  is assigned to hotspot j, and 0 otherwise. Therefore, for each demand  $b_k$ ,  $\sum_j t_{kj} \leq 1$ . For each hotspot j,  $\sum_k b_k \times t_{kj} \leq x_j$ . To solve this integer programming, we can first relax integer constraint on  $t_{kj}$  and solve the relaxed LP. Then, we round  $t_{kj}$  to 0 or 1. Similar rounding approach has been successfully used to find an approximate solution to several integer programming problems (e.g., [33]).

#### V. EVALUATION METHODOLOGY

We evaluate our approach using trace-driven simulations. We first describe the traces and how they are used.

*Traces:* We use the following traces: 1) Locations of cell towers and femtocells from a large cellular provider in the US, and hotspot locations from [36]. 2) Detailed network data with periodic (every 2 s) reports of which sectors mobile devices are using for their data communication. We use one-week data from 2011 and pick the busiest sectors out of thousands of sectors. We then use this data to estimate the number of users in a sector during one hour, and the amount of time they stay in that sector. 3) 3G HTTP traces report detailed HTTP session information, such as HTTP duration, downloaded bytes, and type of the download during all 24 h on a single day in 2011. This is aggregated over several sectors and does not have information

*Generating Regions:* We pick a sector from Manhattan downtown, where the cellular network in the area often gets overloaded. We find 144 hotspots in this sector. We generate regions by clustering the Wi-Fi hotspots using k-means [25]. We use six regions as it minimizes partition index [10], which is a commonly used clustering metric. We run the clustering algorithm 100 times and pick the clustering that minimizes the average distance of Wi-Fi hotspots to the centers of their assigned regions. In this case, the average distance is 32.9 m. We assume a hotspot can only serve the demand of its own region.

Network Configuration: Based on the typical cell tower spacing of 400–500 m in busy urban areas [3], we use 250 m as the communication range for a cell sector. The communication ranges for Wi-Fi and femtocell are set to 100 and 40 m [6], respectively. To calculate spectrum efficiency, we use the distance between the centroid of the region and the cell tower, and the distance between the centroid of the region and the interfering cell towers, and compute path loss using Hata model [19]. We consider six nearest base stations as interfering base stations to calculate the SINR. We account for self-interference and compute the resulting SINR' as:  $SINR' = SINR/(1 + \alpha * SINR)$  where SINR' and SINRdenote the signal-to-interference-plus-noise ratios with and without self interference, respectively, and  $\alpha = 0.005$  [5]. We get the spectrum efficiency by applying the Shannon's Law. Since the Shannon capacity is an overestimate of the real capacity, we scale down the result to match the maximum efficiency that is generally observed in a cellular network (2 b/s/Hz).

Generating Traffic Demands: To generate the demand for an hour, we determine the number of users from the detailed network data during that hour and pick all the HTTP requests of the corresponding number of users from the 3G HTTP trace. The detailed network data and the HTTP trace are both anonymized, and we only use the aggregated demand information in our evaluation. We replace the data rate in the trace with the desired demand rate according to the application types: video 350 kb/s, audio 128 kb/s, application (e.g., download binary files) 350 kb/s, text 150 kb/s, and image 165 kb/s. We determine the rates of applications, text, and images according to the 90th percentile rate that users receive from the 3G HTTP trace and determine the video and audio rates using the data from a large service provider. The data rates in the traces are not used since they are limited by the current cellular capacity and may not indicate the real demand.

We place users randomly inside the sector and assign them to regions according to their locations. When a single demand vector is used, we use the peak demand from each region as the final traffic demand. When dynamic allocation is used, we use all the demand vectors corresponding to the time when any region has peak demand. In this way, both static and dynamic allocation schemes can sustain the peak loads in all regions.

Generating Bids: We use the distribution of backhaul data rates and pick the available data rate uniformly distributed between 25%-75% of the backhaul data rate. The Wi-Fi bids are then generated based on the pricing plan of a major service provider. We uniformly choose 50%-150% of the price as a hotspot's valuation for a given backhaul capacity to capture varying costs from different service providers. We then determine the hourly Wi-Fi valuations according to its capacity and monthly bills assuming 30 days/month and 8 h/day. The real bids depend on their bidding strategies and may differ from their valuations. The cellular bid F(z) is set to 0 (reflecting the sunk costs) when z is below 80% of cellular capacity (which is set to three carriers i.e., 3 times 3.84 MHz), and set to c times estimated maximum Wi-Fi valuation when z exceeds 80%. Note the Wi-Fi valuation is per bit per second, whereas the spectrum cost is per hertz. Thus, we translate the maximum Wi-Fi valuation to price per hertz using the lowest spectrum efficiency so that Wi-Fi is always preferred when the cellular network is overloaded. We set c to 1.25 by default and vary it to evaluate its impact.

*Performance Metrics:* We compare different schemes using valuation and cost. Valuation is the total valuation of all resources consumed and reflects efficiency. Cost is the total price of cellular and Wi-Fi resources the service provider pays. For both metrics, lower values are preferred.

#### VI. EVALUATION RESULTS

# A. Comparison of Truthful Auctions

We first compare the cost incurred under different auction schemes, including iDEAL, iDEAL (static), per-region VCG with global allocation, and per-region VCG with local allocation. All the auctions are truthful except per-region VCG with global allocation, which is included to show how VCG will perform without interregion competition. In addition, we also compare to fixed pricing, where the service provider pays the hotspots a fixed price and uses the global allocation to determine which hotspots to buy. A hotspot with higher valuation than the fixed price we may generate as the fixed price. The result of using average Wi-Fi price as the fixed price is similar and omitted for brevity.

Fig. 5 shows the cost incurred under different schemes. We first observe that auction-based approaches work much better than the fixed pricing when there is enough competition. When the number of hotspots is small, fixed pricing can perform better than most auction-based approaches. However, iDEAL achieves lower cost than the fixed pricing even when the number of hotspots is 40. With 130 hotspots, iDEAL is almost an order of magnitude better than the fixed pricing. Second, iDEAL outperforms iDEAL (static), which outperforms both versions of per region VCG. Per-region VCG fails to capture the interregion competition and thus may suffer from limited competition and lead to high cost. In comparison, both versions of iDEAL fully benefit from interregion competition. iDEAL further reduces cost by leveraging the flexibility of using cellular resource in different regions on demand, thus reducing the demands for third-party resources. Therefore, iDEAL and

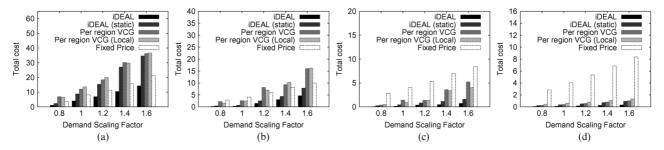


Fig. 5. Total cost comparison with truthful bids. (a) Number of hotspots = 40. (b) Number of hotspots = 70. (c) Number of hotspots = 100. (d) Number of hotspots = 130.

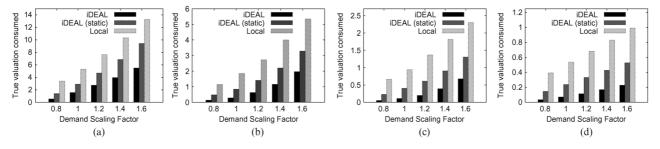


Fig. 6. Comparison of total true valuation consumed. (a) Number of hotspots = 40. (b) Number of hotspots = 70. (c) Number of hotspots = 100. (d) Number of hotspots = 130.

iDEAL (static) outperform per-region VCG by 63%–80% and 10%–61%, respectively.

We further compare the efficiency of the following allocations, all with truthful bids: 1) iDEAL, which can optimize allocation according to multiple possible demands; 2) iDEAL (static), which optimizes allocation according to a single traffic demand; 3) local allocation, which statically allocates cellular resources to different regions based on the traffic demands in these regions. Note here we omit the fixed pricing because it is not an auction and it makes allocation decisions solely based on the fixed price instead of the valuation. Fig. 6 shows the total true valuation of different allocation schemes as we scale the traffic demands by a constant factor from 0.8 to 1.6 and vary the total number of hotspots participating in the auction. As before, iDEAL outperforms its static counterpart, iDEAL (static), which further outperforms the local allocation. iDEAL reduces the total valuation to only 8%-42% of local allocation since it can effectively adapt the cellular allocation to different regions based on real demand. Even iDEAL (static) performs very well: Its total valuation consumed is only 34%-72% of local allocation.

Fig. 7 further compares the cost of different auction schemes as we vary the cellular cost F(z) by changing its parameter cfrom 1 to 2, where the cellular bid is set to c times estimated maximum Wi-Fi valuation when z exceeds 80%. The absolute cost increases with c, as we would expect. The relative performance across different schemes is similar for all values of c we use. The total cost reduces as competition increases (i.e., when the number of hotspots goes up from 40 to 130).

# B. Comparison to Nontruthful Auctions

In this section, we study the impact of individual hotspot gaming in nontruthful auctions. We compare iDEAL to the first price and regional uniform price, both of which are widely used [15], [22]. The first price pays winners the amount of their

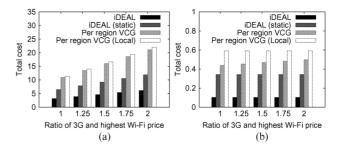


Fig. 7. Total cost comparison with varying cellular cost function. (a) Number of hotspots = 40. (b) Number of hotspots = 130.

bids, and the regional uniform price pays all the winners in a region at the first losing bid in the region. We do not compare to generalized second-price auction (GSP) because, unlike GSP, iDEAL does not differentiate between winning slots. If GSP were used, everyone would game to be the highest paid winner as in the first price. We use the static global allocation for all schemes, except that iDEAL uses dynamic global allocation. There are many possible gaming strategies. In our evaluation, we consider simple gaming strategies as examples and show that even these simple strategies can significantly degrade performance. In the first price, we assume a bidder can observe some fraction of bids from other bidders in his region. We call this fraction Knowledge Factor (KF). He then uses that information to guide his bid in the next round by bidding the maximum among: 1) his valuation; 2) the average of the lowest losing price he sees; and 3) the highest winning price he sees (including his own bid in the last round). In the first round, bidders start by bidding uniformly randomly between one time and two times their valuation. In the uniform price auction, bidders can game by supply reduction. Thus, we let the winners who do not sell all their capacity reduce their capacity to slightly below the amount they sell in the hope of admitting new winners and

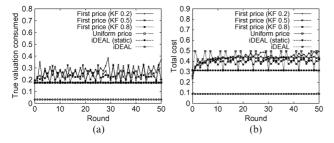


Fig. 8. Cost of gaming. (a) True valuation consumed. (b) Cost.

potentially increasing the price. When they do sell all their capacity, they will try to increase their offered capacity. In reality, bidders can be more aggressive. For example, all bidders may attempt to reduce supply (e.g., even when they sell all they offer, they can potentially still gain by supply reduction), which may harm the system even further. We conduct multiple runs and show the results from one run since they are all similar.

Fig. 8(a) shows how gaming affects efficiency. We make a few observations. First, both versions of iDEAL consume less total valuation. The total valuation of iDEAL is as low as 8% of the first price due to more effective use of cellular resources in presence of multiple demands. The total valuation of iDEAL (static) is only 45% of the first price. Second, both versions of iDEAL are stable as bidders are truthful. In comparison, the total value consumption fluctuates considerably in the first price auction because the bidders adapt their bids according to the others' bids. The uniform price performs close to iDEAL (static) because the bidders in our simulation only reduce supply slightly and they do not game by asking higher. In reality, the damage can only be worse.

Fig. 8(b) further compares the total cost to the provider. Similar to the case of total valuation, both iDEAL versions yield significantly lower cost. Specifically, iDEAL reduces the cost to 18% of the first price and regional uniform price. Moreover, even iDEAL (static) reduces the total cost to 63% of first price and regional uniform price. This result shows that with the help of interregion competition, using VCG does not incur higher cost than first price or regional uniform price.

#### C. Collusion

Collusion Under Dynamic Demands: We first study how often a bidding ring can improve its utility by supply reduction. We use two different sizes of the bidding rings: 20 and 50 out of 144 hotspots. For each size, we run the experiment 10 times with different random sets of hotspots. Each run consists of 50 rounds. In each round, the bidding ring drops all losing hotspots from the previous round. If there is no losing hotspot, it brings back the cheapest previously dropped hotspot. We vary the demand during each round, but keep Wi-Fi bids constant. We confirm the degree of traffic variation in the hourly traffic traces in multiple cellular sectors from a major cellular provider is comparable to the traces used for our evaluation. Fig. 9 plots the cumulative distribution function (CDF) of percentage of utility change of the bidding ring due to collusion. We find that for the bidding ring of size 50, collusion reduces the hotspots' utility for 13% of time and improves the utility for 28% of the time. For the bidding ring of size 20, the numbers are 20% and only 5%, respectively. When collusion reduces utility, it

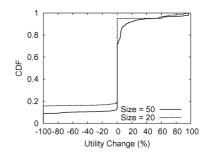


Fig. 9. CDF of utility change due to collusion.

reduces by 79% on average, while the number for improvement is only 30%. These results suggest dynamic demand significantly reduces the incentive to collude. In reality, when Wi-Fi bids are also dynamic, it is even harder to predict which set of hotspots will lose.

*Bidding as a Group:* Next, we compare bidding as a group with collusion using the same strategy mentioned above. Fig. 10 plots the average cost as we vary the total number of sellers and the total number of hotspots they own and perform 100 random runs for each configuration, where each configuration generates 10 sets of sellers and 10 sets of hotspots. The results are consistent with our expectation: A single seller collusion does not always improve utility, but it always incurs a higher cost to the service provider, especially when each seller has a large number of hotspots. In comparison, the group option, which is preferred by sellers, reduces the total cost by as much as 36% and 96% when the number of hotspots is 40 and 130, respectively. The damage of collusion reduces as the number of sellers increases since increasing the number of sellers means each seller controls fewer hotspots.

# D. Extensions

Allowing Bidding Curves: If only F(z) is nonconvex, we can approximate F(z) using t linear segments and optimize allocation by solving t LPs, one corresponding to each line segment. The running time increases by a factor of t. When the Wi-Fi bids are nonconvex functions, we need to use the dynamic programming (DP) formulation in Section IV to optimize allocation.

To quantify the computation cost and quality of DP solutions, we compare them to those of the LP when the Wi-Fi bids are convex (since LP can only handle convex functions). iDEAL static allocation is used in all cases. We performed the computation on a 7-core Intel Xeon 2.83 GHz CPU, with 32 GB RAM. Each result is an average of five runs. Fig. 11(a) shows that DP increases cost by 19.9%–45.1% compared to the LP due to discretization. As one would expect, smaller step sizes in the DP (defined in Section IV) yield closer results to the LP. This is achieved at the cost of increasing running time. As shown in Fig. 11(b), step sizes of 50 kHz for cellular spectrum and 40 kB/s for Wi-Fi capacity achieve close-to-optimal solution and take around 1 min for 130 bidders, which is affordable in practice.

Supporting Femtocell Offload: In Fig. 12(a), we let both Wi-Fi hotspots and femtocells participate in the auction. We vary the number of Wi-Fi bidders while keeping 16 femtocells. As expected, the benefit of femtocells is larger when we have fewer Wi-Fi hotspots. For example, the femtocells reduce the cost by 32% when there are only 40 hotspots. As the number

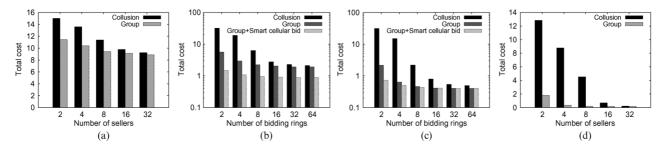


Fig. 10. Auction cost under collusion and with group option. (a) Number of hotspots = 40. (b) Number of hotspots = 70. (c) Number of hotspots = 100. (d) Number of hotspots = 130.

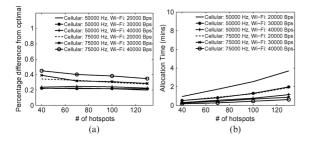


Fig. 11. Performance of dynamic programming. (a) Cost. (b) Allocation time

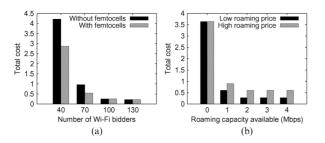
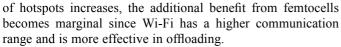


Fig. 12. Benefit of femtocell offloading and roaming. (a) Femtocell offloading. (b) Roaming.



Supporting Dynamic Roaming: Fig. 12(b) shows the total cost as the roaming capacity available varies from 0 to 4 Mb/s, where 0 corresponds to no roaming. The evaluation has 40 hotspots. In this case, since the Wi-Fi resource is insufficient, even having 1 Mb/s of available roaming capacity (around 10% total cellular traffic in the sector) can significantly cut down cost. Dynamic roaming reduces the cost to 17% of that when only Wi-Fi is used with the low roaming price (which is set to the maximum winning Wi-Fi bid we observe in the default settings), and 25% under the high roaming price (which is the maximum Wi-Fi bid we may generate based on the distribution we use). Further increasing roaming capacity leads to an even lower cost, but the improvement tapers off as the capacity increases beyond 2 Mb/s.

Delaying Delay-Tolerant Demands: We use two scenarios to demonstrate the benefit of delaying some traffic demands. Fig. 13(a) shows the total cost to serve both rounds under the first scenario, where we pick the highest demand of all the hours as the demand in the first round and use the average demand in an hour in the second round. We vary the penalty factor from 0

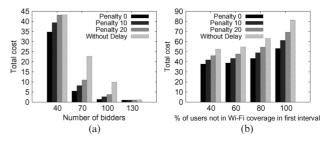


Fig. 13. Benefit of delay-tolerant demand. (a) Scenario 1. (b) Scenario 2.

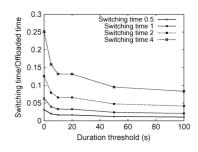


Fig. 14. Strategically selecting users to offload reduces the switching cost.

to infinity, where infinity means no demand will be delayed. We find delaying some demands to the second round is beneficial in all cases when penalty is smaller than infinity. The benefit is largest when the resource we have is not sufficient for the first round but sufficient for the second round. As the figure shows, the savings are 63.6% and 72.6% for 70 and 100 hotspots, respectively, when penalty factor is 10. The corresponding numbers are 51.4% and 61.2% when the penalty factor is 20.

In Fig. 13(b), we consider the scenario where in the first round many users are not in Wi-Fi coverage, and we vary the fraction of such users. In this evaluation, we only use hotspots from three regions and leave the other three regions not covered by Wi-Fi. In the second round, the users are placed uniformly across the sector. The results show that delaying some demands in this case yields significant saving. When 100% of the users are not in Wi-Fi coverage in the first round, the saving can be up to 24% under penalty factor 10, and 14.8% under penalty factor 20. When 40% of the users are outside Wi-Fi coverage, the numbers become 20.2% and 12.2%, respectively.

Selecting Users to Offload: We use 1 h of the HTTP trace in one region and determine the total Wi-Fi capacity to buy using iDEAL. We vary the Wi-Fi switching time from 0.5 to 4 s. Fig. 14 plots the ratio of total switching time and total offloaded time as we vary the threshold duration we use to offload a user to Wi-Fi (e.g., the user has to stay at the Wi-Fi hotspot for a period that is over the threshold time). In all cases, we ensure the Wi-Fi capacity that was purchased is fully utilized since most of the traffic come from the users who stayed at Wi-Fi much longer than the threshold time. As we can see, picking users who stay at Wi-Fi hotspot longer reduces the fraction of time spent in switching from 25.1% to 15.9% when using 4 s switching time and 5 s duration threshold.

# VII. IMPLEMENTATION

We describe how to offload in state-of-the-art commercial systems, and then present our implementation.

Offloading involves the following three issues: 1) identifying a network to offload; 2) automatic authentication; and 3) seamless offload so that the existing sessions are maintained during the offload. iDEAL already solves the first issue. Here, we describe the other two issues.

Automatic Authentication: 3GPP Release 7 uses the Extensible Authentication Protocol (EAP) for key distribution to WLANs that are owned by the cellular service provider. Hotspot 2.0 is developed to support authentication with externally owned hotspots. Specifically, Hotspot 2.0 uses 802.11u to support the Access Network Query Protocol (ANQP), which is a query-response protocol used by a mobile device to discover information including hotspot owner's domain name, roaming partners accessible via the hotspot, and EAP method used for authentication and IP address type availability [2], [4]. To support dynamic offloading in this paper, the roaming partners are updated dynamically according to the offloading decision of our algorithm.

Seamless Mobility: There are several strategies to perform data offloading. The simplest strategy is to use an application-based switch, which simply moves a connection between cellular and Wi-Fi networks but does not worry about preserving the sessions across the switchover. It can cause a disruption to most applications and result in negative user experience, especially for VoIP, VPN applications, and video streaming applications [1].

3GPP Release 8 uses Dual Stack Mobile IP (DSMIP) to enable seamless handover between 3G and Wi-Fi. The solution does not require any support from Wi-Fi hotspots. The cellular radio access network supports a Home Agent (HA) that binds the new IP address of the node to the permanent IP. Since the IP address is preserved in this case, it provides a better user experience compared to application-based switching [1]. Moreover, 3GPP Release 10 uses DSMIPv6, which allows mapping multiple IP addresses to a single permanent IP address and supporting simultaneous use of Wi-Fi and 3G according to application QoS requirements (e.g., keeping VoIP application on 3G and using Wi-Fi for bandwidth-intensive applications like video streaming). Reference [23] uses an implementation to quantify the efficiency of DSMIPv6 for managing handoffs between networks. It reports a 0.02-s interruption and three lost packets when switching between IPv6 to IPv6 connection using two interfaces and 0.09 s interruption, and 17 lost packets while switching between IPv4 to IPv6 network. Further optimizations are possible to reduce the switching delay and packet losses. For example, [32] shows packet loss can be reduced to near zero with buffering on the mobile nodes.

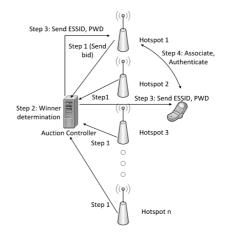


Fig. 15. System architecture.

*Our Implementation:* We develop a prototype implementation on Linux machines using a NetGear WAG511 NIC to demonstrate the feasibility of our solution. Fig. 15 shows our system architecture. Through a simple web interface, hotspot owners can submit their bids to the service provider machine, who controls the auction. Hotspots are configured using hostap [20]. Depending on who wins in the round, the service provider sends a message to the hotspot with the ssid and password it should use in the current round and also sends the ssid and password to the mobile client machine so that it can connect to the winning hotspot. This message is sent using TCP sockets. Authentication between mobile client and hotspot is done using WPA PSK through WPA Supplicant [37].

We collected performance statistics from the mobile client for billing and keeping track of hotspot quality score. We measured the upload and download statistics on the wireless interface using the Collectl tool [14] periodically (e.g., every 10 s) and send back the data to service provider PC for bookkeeping. iDEAL allocation and pricing take 43 and 74 ms, respectively, which are both small.

We further measure the association and authentication time in our implementation. After getting the scan results, it takes 18 ms to associate, 103 ms to perform four-way handshake (i.e., defining individual keys for unicast transmission), and 3 ms to perform the group handshake (i.e., defining keys for broadcast transmission). The authentication times can be further reduced (e.g., using techniques in [21] and [26]). Moreover, scan times can be shortened by selective scanning on fewer channels as proposed in [27]. Therefore, we can achieve a very low overhead for handoff, making offloading feasible.

#### VIII. RELATED WORK

The need to complement cellular networks with other forms of connectivity has been considered in the past. The authors in [9] conduct measurements in a vehicular testbed and report that Wi-Fi is available only 11% of the time and 3G is available 87% of the time. Moreover, they find that 3G and Wi-Fi availability are negatively correlated, e.g., Wi-Fi is available 50% of the times that 3G is not available. Lee *et al.* in [24] use daily mobility patterns of 100 iPhone users to measure the amount of data Wi-Fi can offload. They find that Wi-Fi can offload 65% of

data traffic without any delay; if 1-h or longer delay can be tolerated, the offload traffic increases further by 29%. Zhuo *et al.*, in [39], leverage VCG-based auction mechanism to incentivize mobile users to wait until they come in contact with a Wi-Fi AP. Authors in [16] quantify citywide Wi-Fi offloading gain. They show that even a sparse Wi-Fi network improves performance. Different from the above existing works, our paper focuses on how to incentivize third-party resource owners to offload cellular traffic. The work in [12] is closest to ours. It proposes a VCG reverse auction framework to buy femtocell resources. Their scheme is similar to the local allocation in spirit in that it statically determines the amount of third-party resource to buy in each region. As mentioned in Section I, it does not address the three unique challenges we focus on, namely, diverse spatial coverage, traffic uncertainty, and collusion.

# IX. CONCLUSION

How to sustain the explosive growth of cellular traffic without requiring prohibitive investment is a major challenge for cellular service providers. Our proposal, iDEAL, enables the cellular service provider to purchase and leverage third-party resources on demand through reverse auctions. iDEAL promises a win–win solution: The cellular service provider achieves significant savings by not having to provision for the peak traffic demands; third-party resource owners (e.g., Wi-Fi hotspots, Femtocells, or other cellular service providers) gain extra revenue from the otherwise wasted spare capacity; and the consumers' quality of experience is protected.

We showed that several of iDEAL's key features are critical to the superior performance of the reverse auction. iDEAL explicitly accounts for the diverse spatial coverage of different resources and copes with the dynamic nature of traffic demands. We showed that iDEAL effectively incentivizes third-party resource owners to be truthful to their true valuation in the bidding process and is provably efficient by choosing the bidders with the lowest valuation as the winners.

We evaluated iDEAL using simulations based on detailed traces from a large cellular service provider. Compared to the local allocation, iDEAL allocation yields up to 92% improvement. Compared to a variety of different auction schemes, iDEAL leads to up to 92% improvement. We also showed that iDEAL effectively mitigates collusion.

As future work, we plan to conduct trials of iDEAL in selected regions in one of the largest US cellular networks. Another interesting topic for future work is to support WiFi providers to sell to multiple cellular providers. In this case, we need to simultaneously capture competition among buyers as well as competition among sellers. Double auction is a potential solution, which we will explore.

#### REFERENCES

- Qualcomm, Inc., San Diego, CA, USA, "3G Wi-Fi seamless offload," 2010 [Online]. Available: http://www.qualcomm.com/ media/documents/files/3g-wifi-seamless-offload.pdf
- [2] "802.11u," 2013 [Online]. Available: http://en.wikipedia.org/wiki/ IEEE\_802.11u
- [3] "Cell site," 2013 [Online]. Available: http://en.wikipedia.org/wiki/ Cell\_site
- [4] S. Martin, "802.11u and Hotspot 2.0 promise Wi-Fi users a cellularlike experience," 2011 [Online]. Available: http://www.networkworld. com/news/tech/2011/090711-80211u-hotspot-250556.html

- [5] "3GPP2 C.R1002-A CDMA2000 evaluation methodology revision A," ver. 1.0, May 2009.
- [6] AT&T, Dallas, TX, USA, "Femto-cell range," 2013 [Online]. Available: http://www.att.com/shop/wireless/devices/ 3gmicrocell.jsp?fbid=jZ\_3W-pHS6d
- [7] L. M. Ausubel and P. Milgrom, "The lovely but lonely Vickrey auction," in *Combinatorial Auctions*. Cambridge, MA, USA: MIT Press, 1961, pp. 1–37.
- [8] Y. Bachrach, "Honor among thieves: Collusion in multi-unit auctions," in *Proc. AAMAS*, Richland, SC, USA, 2010, pp. 617–624.
- [9] A. Balasubramanian, R. Mahajan, and A. Venkataramani, "Augmenting mobile 3G using WiFi," in *Proc. MobiSys*, 2010, pp. 209–222.
- [10] J. C. Bezdek, "Numerical taxonomy with fuzzy sets," J. Math. Biol., vol. 1, pp. 57–71, 1974.
- [11] O. Biran and F. Forges, "Core-stable rings in auctions with independent private values," CESifo Group Munich, Germany, CESifo Working Paper Series 3067, 2010.
- [12] Y. Chen, J. Zhang, Q. Zhang, and J. Jia, "A reverse auction framework for access permission transaction to promote hybrid access in femtocell network," in *Proc. IEEE INFOCOM*, 2012, pp. 2761–2765.
- [13] Cisco, San Jose, CA, USA, "Cisco visual networking index: Global mobile data traffic forecast update, 2010–2015," Cisco White Paper, Feb. 2011.
- [14] "Collectl," 2013 [Online]. Available: http://collectl.sourceforge.net/
- [15] P. Cramton, "Competitive bidding behavior in uniform-price auction markets," in *Proc. Hawaii Int. Conf. Syst. Sciences*, 2004, vol. 2, pp. 1–11.
- [16] S. Dimatteo, P. Hui, B. Han, and V. O. Li, "Cellular traffic offloading through WiFi networks," in *Proc. IEEE MASS*, 2011, pp. 192–201.
- [17] W. Dong, Z. Ge, and S. Lee, "3G meets the Internet: Understanding the performance of hierarchical routing in 3G networks," in *Proc. ITC*, 2011, pp. 15–22.
- [18] B. Fortz, J. Rexford, and M. Thorup, "Traffic engineering with traditional IP routing protocols," *IEEE Commun. Mag.*, vol. 40, no. 10, pp. 118–124, Oct. 2002.
- [19] M. Hata, "Empirical formula for propagation loss in land mobile radio services," *IEEE Trans. Vehicular Tech.*, vol. VT-29, no. 3, pp. 317–325, Aug. 1980.
- [20] J. Malinen, "hostapd: IEEE 802.11 AP, IEEE 802.1X/WPA/WPA2/ EAP/RADIUS authenticator," 2013 [Online]. Available: http://hostap. epitest.fi/hostapd/
- [21] M. Kassab, A. Belghith, J.-M. Bonnin, and S. Sassi, "Fast pre-authentication based on proactive key distribution for 802.11 infrastructure networks," in *Proc. 1st ACM Workshop Wireless Multimedia Netw. Perform. Modeling*, 2005, pp. 46–53.
- [22] V. Krishna, Auction Theory. New York, NY, USA: Academic, 2002
- [23] R. Kuntz and J. Lorchat, "Versatile IPv6 mobility deployment with dual stack mobile IPv6," in *Proc. MobiArch*, 2008, pp. 49–54.
- [24] K. Lee, I. Rhee, J. Lee, S. Chong, and Y. Yi, "Mobile data offloading: How much can WiFi deliver?," in *Proc. CoNEXT*, 2010, Art. no. 26.
- [25] S. Lloyd, "Least squares quantization in PCM," *IEEE Trans. Inf. Theory*, vol. IT-28, no. 2, pp. 129–137, Mar. 1982.
- [26] A. M. Min-Ho, M. H. Shin, and W. A. Arbaugh, "Proactive key distribution using neighbor graphs," *IEEE Wireless Commun.*, vol. 11, no. 1, pp. 26–36, Feb. 2004.
- [27] S.-H. Park, H.-S. Kim, C.-S. Park, J.-W. Kim, and S.-J. Ko, "Selective channel scanning for fast handoff in wireless LAN using neighbor graph," in *Proc. PWC*, 2004, pp. 194–203.
- [28] L. Qiu, Y. R. Yang, Y. Zhang, and S. Shenker, "On selfish routing in internet-like environments," in *Proc. ACM SIGCOMM*, Aug. 2003, pp. 151–162.
- [29] R. Raghavendra, J. Padhye, and R. Mahajan, "Wi-Fi networks are underutilized," Microsoft Tech. Rep. MSR-TR-2009-108, 2009.
- [30] M. Roughan, M. Thorup, and Y. Zhang, "Traffic engineering with estimated traffic matrices," in *Proc. ACM SIGCOMM IMC*, Oct. 2003, pp. 248–258.
- [31] M. Shafiq, L. Ji, and A. Liu, "Characterizing and modeling Internet traffic dynamics of cellular devices," in *Proc. ACM SIGMETRICS*, 2011, pp. 305–316.
- [32] S. Sharma, I. Baek, and T.-C. Chiueh, "Omnicon: A mobile IP-based vertical handoff system for wireless LAN and GPRS links," *Softw. Pract. Exper.*, vol. 37, pp. 779–798, 2007.
- [33] D. Shmoys and E. Tardos, "An approximation algorithm for the generalized assignment problem," *Math. Program. A*, vol. 62, pp. 461–474, 1993.

- [34] "Sunk costs," 2013 [Online]. Available: http://en.wikipedia.org/wiki/ Sunk costs
- [35] W. Vickrey, "Counterspeculation, auctions and competitive sealed tenders," J. Finance, vol. 16, no. 1, pp. 8-37, 1961.
- [36] "Wireless geographic logging engine," 2013 [Online]. Available: http://wigle.net
- [37] J. Malinen, "Linux WPA/WPA2/IEEE 802.1X supplicant," 2013 [Online]. Available: http://hostap.epitest.fi/wpa\_supplicant/
- [38] Z. Zhu, G. Cao, R. Keralapura, and A. Nucci, "Characterizing data services in a 3G network: Usage, mobility and access issues," in Proc. IEEE ICC, 2011, pp. 1-6.
- [39] X. Zhuo, W. Gao, G. Cao, and Y. Dai, "Win-coupon: An incentive framework for 3G traffic offloading," in Proc. IEEE ICNP, 2011, pp. 206-215.



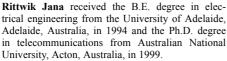
Wei Dong received the M.S. degree in computer science from The University of Texas at Austin, Austin, TX, USA, in 2011, and is currently pursuing the Ph.D. degree in computer science at the same university. His advisors are Prof. Yin Zhang and Prof. Lili Oiu.

His research interests include network economics and pricing, mobile systems and applications, security and privacy, and social networks.



Swati Rallapalli received the M.S. degree in computer science from The University of Texas at Austin, Austin, TX, USA, in 2010, and is currently pursuing the Ph.D. degree in computer science at the same university. Her advisor is Prof. Lili Qiu.

Her research interests lie in the area of protocol design for mobile networks. Some of her specific interests include localization and mobility prediction, network economics, and social networks.



He is a Principal Member of Technical Staff with AT&T Labs-Research, Florham Park, NJ, USA. He worked as an Engineer with the Defense Science and Technology Organization (DSTO), Australia, from 1996 to 1999, and as a Principal Member of Technical Staff with AT&T from 1999 to date. He

is a published author and has served as a reviewer and program committee member for numerous conferences and journals. His research expertise falls in the areas of IPTV, P2P, mobile middleware and wireless channel modeling.

Dr. Jana has also successfully co-chaired MobEA workshops at WWW 2003 to 2010.



Yin Zhang (M'02-SM'07) received the Ph.D. degree in computer science from Cornell University, Ithaca, NY, USA, in 2001.

He is an Associate Professor with the Department of Computer Sciences, University of Texas at Austin, Austin, TX, USA. Between 2001 and 2004, he was a member of the Networking Research Department, AT&T Labs-Research, Florham Park, NJ, USA. His research interests span several areas of computer networks, including network measurement, network management, and network security.

Prof. Zhang received the NSF CAREER Award in 2006.

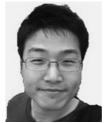


Award in 2006.

Lili Qiu (S'96-M'01-SM'07) received the Ph.D. degree in computer science from Cornell University, Ithaca, NY, USA, in 2001.

She is an Associate Professor with the Department of Computer Sciences, University of Texas at Austin, Austin, TX, USA. She was a Researcher with Microsoft Research, Seattle, WA, USA, from 2001 to 2004. Her research interests are in Internet and wireless networking, with special focuses on wireless network performance and management.

Prof. Qiu is a recipient of the NSF CAREER



Tae Won Cho received the doctoral degree from the University of Texas at Austin, Austin, TX, USA, in 2010

He is a Senior Member of Technical Staff with the AT&T Shannon Laboratory, Florham Park, NJ, USA. His interest spans mobility networks, computer network systems, social networks, and data mining.



K. K. Ramakrishnan (S'76-A'83-M'03-SM'04-F'05) received the M.S. degree in automation from the Indian Institute of Science, Bangalore, India, in 1978, and the M.S. and Ph.D. degrees in computer science from the University of Maryland, College Park, MD, USA, in 1981 and 1983, respectively.

He joined AT&T Bell Labs in 1994 and has been with AT&T Labs-Research, Florham Park, NJ, USA, since its inception in 1996. Prior to 1994, he was a Technical Director and Consulting Engineer in Networking with Digital Equipment Corpora-

Leo Razoumov received the M.S. degree in physics from Yaroslavl State University, Yaroslavl, Russia, in 1988, and the Ph.D. degree in mathematical physics from the P. N. Lebedev Physics Institute, Moscow,

Since 1996, his interests switched from physics to

information theory and wireless communications. He

has held technical leadership and R&D positions with

Qualcomm, Inc., San Diego, CA, USA; Intel Corpo-

ration, Woodbridge, NJ, USA; Wireless Information

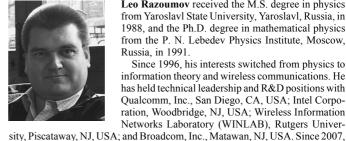
Networks Laboratory (WINLAB), Rutgers Univer-

tion, Littleton, MA, USA. Between 2000 and 2002, he was with TeraOptic Networks, Inc., Sunnyvale, CA, USA, as Founder and Vice President. A Distinguished Member of Technical Staff with AT&T Labs-Research, he is involved in several technical and strategic activities in networking and information distribution. He has published nearly 200 papers and has over 100 patents issued in his name.

Dr. Ramakrishnan is an AT&T Fellow, recognized for his contributions to communications networks, including congestion control, traffic management, and VPN services. He has been on the Editorial Board of the IEEE/ACM TRANSACTIONS ON NETWORKING and IEEE Network and has been a member of the National Research Council Panel on Information Technology for NIST.

Russia, in 1991

he has been a Principal Member of Technical Staff with AT&T Labs-Research, Florham Park, NJ, USA. His areas of his interest include advanced wireless networks, MIMO mobility, radio resource management, and wireless network



performance.