ABSTRACT

Opportunistic networking based on hand-held mobile devices is turning into a viable and efficient opportunity to locate, collect, route and share information within a swarm of collaborative nodes. In this paper we consider mobile (pedestrian and cars) and fixed terminals in a urban area that are interested in collecting the information originated from several sources. In particular, each terminal aims at retrieving the data items in a limited region of interest centered around the node position. Since data items may change over time all nodes must strive for having access to the latest version. Furthermore, for mobile terminals the region of interest is a time varying concept due to the dynamic behavior of nodes. The goal of the paper is to evaluate the amount of information each node is able to gather resorting to simple distributed data collection and sharing through opportunistic communications among neighboring nodes. In particular, we analyze the impact of node density, different mix of cars and pedestrian, and amount of node memory. Moreover, we evaluate the improvement of using location aware memory management policies as well as the effect of adding a few ideal nodes whose mobility is described by an unconstrained Brownian motion. To this end we develop a simulator based on mobility and radio propagation traces obtained from the UDelModels tools. The preliminary findings highlight that simple location aware memory management schemes effectively exploit nodes with limited amount of memory. Furthermore, increasing randomness of nodes movement has a beneficial impact on the average performance of all node types.

Categories and Subject Descriptors

C.2.1 [Computer-communication networks]: Network Architecture and Design — Store and forward networks

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

1. INTRODUCTION

Portable computing devices are nowadays available for both people, e.g., smartphones and netbooks, and cars. Wireless communication capabilities of these devices make it possible to devise innovative distributed applications based on the creation of opportunistic mobile ad-hoc networks in urban scenarios.

Applications can be based on the paradigm of data diffusion, collection, or both. In this paper we focus on the following abstraction: we consider mobile and fixed terminals that are interested in collecting data items originating from several sources in a specific urban area. In particular, each terminal aims at retrieving the data items in a limited region of interest (ROI) centered around the current node position. Since data items may change over time all nodes must strive for having access to the latest version. Furthermore, for mobile terminals ROI is a time varying concept due to the dynamic behavior of pedestrians and cars.

Terminals can basically perform two operations: acquire the data item originating in a specific location of the urban area when their current position is in its close proximity, e.g., autonomous sensing of environmental signals, reading from active sensors, passive information conveyed by RFID tag, commercial advertisements or traffic data, etc. Terminals can also exchange data items among them as long as the radio channel loss is lower than a given threshold. Terminals are equipped with memory to store the collected data items. The memory size depends on the type of terminal: hand-held computing devices are assumed to reserve a small memory to the application while terminals mounted on fixed terminals and vehicles are assumed to use a larger memory.

The analysis we present is based on the performance index defined as the fraction of updated data elements collected by a node with respect to the total amount of data elements within its ROI at a specific time. We present results for static and dynamic data items generation, various memory management schemes and for different classes of terminals, i.e., fixed nodes and mobile nodes (pedestrians and...
cans). Furthermore, we investigate the impact of increasing the randomness of mobility patterns in urban scenarios by adding artificial terminals that follow a completely random path within the considered simulation area characterized by Brownian motion with constant average speed.

The results we present are obtained from a simulator developed on top of the UDelModels tools [13, 24]. These tools are used to define three-dimensional maps of urban areas and to obtain traces of mobility and radio propagation. The mobility characteristics of the nodes are based on statistical studies of population and traffic dynamics. Mobility traces generated by UDelModels are very detailed: for instance, pedestrians exhibit a motion that is representative of people in an urban scenario, with different mobility distributions for outdoor and indoor walking, respectively. Moreover, a typical daily human activity cycle is taken into account. Cars mobility patterns take into account speed limits and traffic lights. The UDelModels propagation simulator is used to estimate the point to point channel loss between each pair of nodes in the three-dimensional space, taking into account the urban three-dimensional profile.

The paper is organized as follows: Section 2 describes related works, Section 3 defines the scenario and the abstraction we consider, Section 4 illustrates the simulator we developed, the tools we used and discusses the results we obtained. Finally, Section 5 draws conclusions and outlines future development of the current work.

2. RELATED WORKS

Several papers addressed the analysis of information diffusion in mobile nodes population. Closest to the spirit of our work is [17]: this paper considers the city-wide content distribution to a population of interested mobile nodes with the aid of a small number of fixed nodes (that also act as data sources) and other uninterested mobile nodes. Different distribution schemes are evaluated by simulation exploiting mobility traces collected by tracing movements of students from the Cambridge University.

In [5] Brownian mobility is studied to improve gossiping in a wireless system with the aim of local identification of objects and directional gossiping of queries. Theoretical bounds on gossip performance and information propagation speed in mobile systems have been obtained in [23,22] and [11]. Similarly, in [9] the capacity of ad hoc wireless networks is studied under general mobility conditions. Graph theory has been exploited to analyze mobility in wireless networks in [15] and [2].

Sophisticated memory management techniques have been analyzed in [16] and [7] to increase the hit probability of queries issued by mobile nodes searching for cached content in other nodes memory. The goal is to maximize the hit probability while minimizing the number of hops required by a query to reach the target mobile node. In [14] epidemic dissemination is considered and buffer management policies are studied with the aim to optimize average delivery rate and delay in case of limited size of buffer’s nodes.

Social networks concepts and analysis techniques have been exploited in several papers. In [8] the authors build a taxonomy of relations among nodes that can be used to facilitate the forwarding of information. Examining the frequency and duration of contacts among people, nodes are partitioned in different classes (friends, strangers, and others). Statistical analysis is used to define class dependent forwarding to maximize data dissemination. Similarly in [10] real traces are analyzed to cluster nodes in communities; the analysis of the distribution of the number of contacts is exploited to design an algorithm to forward information with good delivery ratio and very low resource utilization. Wireless sensor networks also show some similarities with our work. In particular, a three tiers architecture called MULE is proposed in [12]; it is composed of limited resources sensors worn by people, animals, and vehicles to carry and forward data that is gathered by access points. The main goal is to exploit the mobility of MULE nodes to forward data while saving sensor energy for communications. In [21] the authors design a distributed algorithm in wireless resource-limited network where nodes are interested in collecting data in a given area. A management scheme of nodes memory that maximize the mutual distance of stored data samples is devised to maximize area coverage. In [20] a probability distribution to query for a location in the area is considered and information theory is exploited to select the cache contents of mobile devices. In [1] the problem of efficiently answering real-time geo-centric ad-hoc queries on data stored by sensors is considered. Each zone has its local storage-sensor that is temporarily selected and may vary with time for energy load-balancing; a point-to-point routing scheme to deliver the reading of any sensor to its storage-sensor is also designed. Furthermore, in [18] the authors design a system to meet spatial and temporal constraints of queries of mobile users about their surrounding areas also in case of nodes with very low duty cycles.

3. OPPORTUNISTIC DATA COLLECTION

In this section the characteristics of the system under analysis are presented along with the assumptions we made to model the interactions among the fixed communication relays and the tetherless nomadic terminals.

Within our study we focus on the collection of data items located in a urban area $U$. We assume that a two-dimensional data set is associated with $U$. In particular, each $\delta \times \delta$ tile of $U$ is characterized by a data item $i(x,y,t)$, where the integer coordinates $(x,y)$ are the tile indexes and $t$ represents a time-stamp associated with the information. The information $i(x,y,t)$ can be composed by a set of environmental measurements taken autonomously by the mobile nodes or communicated by an infrastructure of active sensors, passive information conveyed by RFID tag, commercial advertisements or traffic data, etc. The area $U$ is populated by wireless fixed and mobile nodes whose objective is to retrieve the data in a limited ROI centered around the node position. For simplicity in the following we assume that the ROI is represented by all the information $i(x,y,t)$ within a $\Delta \times \Delta$ tile around each node.

Our model includes several classes of nodes, namely fixed (F) nodes and mobile nodes which are further divided into pedestrian (P), vehicular (V) and random walk (R) nodes. By F nodes we mean wireless relay stations placed at fixed location, e.g. at block intersections and building entrance. Nodes P,V and R are mobile wireless terminal characterized by different mobility patterns. P nodes move along streets at walking speed and typically concentrate within buildings, V nodes are cars moving faster along the streets. The R nodes are artificial entities that follow a completely random path with the area $U$; these nodes are characterized by Brownian motion with average speed equal to $v_R$. This latter class is
introduced to study the effects of nodes with unconstrained mobility on the performance of opportunistic data exchange and collection.

The interactions, i.e. the mobility and the radio coverage of F, P and V nodes, is modeled by the UDelModels tools [13]. UDelModels is a suite of tools for simulating urban mesh networks and includes a simulator of urban radio propagation and a simulator of realistic urban mobility. The mobility characteristics of the nodes are based on statistical studies of population and traffic dynamics. The UDelModels simulator takes as inputs a three-dimensional map of the urban area, the number of nodes in the different classes and the statistical parameters of the mobility. Pedestrians exhibit a motion that is representative of people in an urban scenario, with different mobility distributions for outside and inside walking, respectively. Moreover, a typical daily human activity cycle is taken into account.

Cars mobility patterns take into account speed limits and traffic lights.

The UDelModels propagation [24] simulator is used to estimate the point to point channel loss between each pair of nodes in the three-dimensional space, taking into account the urban three-dimensional profile. This estimates are used to predict the radio contacts that each node is able to establish with others. Two nodes are assumed to be able to communicate one another if the channel loss between them is lower that a threshold \( \alpha \). The only exception is represented by the \( R \) nodes that, being included in our model as ideal entities, do not use such precise attenuation model; instead an \( R \) node is assumed to be able to communicate with all the nodes within a fixed range \( A_R \) around its current position.

In Figure 1 an example of simulated urban area is shown where nodes ROI are represented as squares centered around sample P and V nodes.

### 3.1 Data collection strategies

The goal of the present study is the analysis of the efficiency of the data collection strategies used by the nodes to cover their respective ROI in presence of limited buffering capability \( B \), defined as the number of data items that a node is able to store in its local memory \( B \). Each node can collect all the information of interest, provided that \( B \geq (\frac{\Delta}{\delta})^2 \). The data stored in the buffer \( B \) of a given node come from two sources: the information \( i(x, y, t) \) directly acquired from the environment when the node steps inside the tile at coordinates \((x, y)\) and the information gossiped by other nodes during occasional radio contacts thanks to their mobility. In all our analysis we assume that the time is slotted. In each time slot the node acquires the information \( i(x, y, t) \) corresponding to its position and gets potentially interesting data pushed by other nodes in its radio range.

The data collection efficiency, i.e. the percentage of data collected by the node that fall within its current ROI, are clearly influenced by the strategy used to exchange data and the policy adopted to refresh the buffer.

The exchange policy we adopt is simple and completely random: each node randomly selects \( k_n \) nodes in its radio coverage and pushes to each of them \( k_{dt} \) randomly chosen data items stored in its \( B \). Therefore, during each time slot a given node receives a variable number of items from other nodes in the vicinity and collects one item associated to its position.

Different buffer management strategies have been included in our model. The simplest one is based on First In First Out (FIFO) approach with updating of the already known information at a given location \((x, y)\). With this strategy each received piece of information \( i(x, y, t) \) is pushed into \( B \) with a FIFO policy if no information about the position \((x, y)\) has been included yet; on the contrary, if \( i(x, y, t) \in B \) with \( t_0 < t \), then, the corresponding record is updated.

Clearly, this basic approach is quite simple but does not enforce any locality check on the content of \( B \). In this paper we analyze two improvements, referred in the following as selective dropping (SD) and selective insertion (SI), that aim at prioritizing the storage of the local data. When using SD the data can be popped out of \( B \) only if they refer to a location outside the node’s ROI. Similarly, the SI prevents from pushing into \( B \) any data outside the node ROI. It is worth pointing out that the locality check we introduced does not guarantee a priori a good performance to mobile nodes, since in that case the ROI is not statically known but varies according to the node mobility.

### 4. SIMULATION RESULTS

In this section we report on a set of experiments, whose goal is the analysis of the sensitivity of the proposed data collection policies to crucial feature of the system, i.e. the density and the kind of nodes mobility, the amount of nodes memory, and the usage of buffer management policies.

The performance index we are interested in is represented by the percentage of the ROI covered by the items stored in the local buffer \( B \) of each node. In the following we will refer to such percentage as coverage \((0 \leq C \leq 1)\). The value of the coverage clearly varies from node to node and it is time dependent. Indeed, when the simulation begins \( B = \emptyset \) and \( C = 0 \) for all the nodes. The goal of the collection policies we analyze is to keep \( C \) as high as possible all the time, excluding the initial transient phase. All our analysis will be based on the estimation of the coverage \( C \), defined as the average ROI coverage experienced in the different node classes: fixed, vehicular and pedestrian, respectively; the transient is excluded from the computation to avoid biasing the estimates. Finally, all the reported results are always averaged on 20 independent simulation trials so as to obtain statistically meaningful results.

As already mentioned, the simulation is based on the UDelModels tools [13, 24] for the computation of the nodes mobility and propagation. In particular, UDelModels is used to generate realistic node trajectories and to compute the communication channel loss between any two radio stations in the three-dimensional urban scenario with the granularity of 1 s. The UDelModels results are used as input by our ad-hoc simulator whose major functionalities are:

- to simulate the radio contacts among the nodes as far as the channel loss is below a threshold \( \alpha \);
- to implement the management of the buffer \( B \);
- to simulate the exchange of collected items among the nodes in radio contact;
- to simulate the mobility and radio propagation of \( R \) nodes, that are not available in UDelModels; the direction of the trajectory of the \( R \) nodes is updated randomly every second and their speed is drawn according to a uniform distribution;
Figure 1: A zoom of the simulated urban area: nodes ROI are depicted as squares.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>( U )</td>
<td>550 m × 500 m</td>
</tr>
<tr>
<td>( \delta )</td>
<td>25 m</td>
</tr>
<tr>
<td>( \Delta )</td>
<td>F nodes 400 m, V nodes 200 m, P, R nodes 100 m</td>
</tr>
<tr>
<td>( B )</td>
<td>( b \left( \frac{\Delta}{\delta} \right)^2 ), with ( b \geq 1 )</td>
</tr>
<tr>
<td>( V_R )</td>
<td>( U(9,11) ) m/s</td>
</tr>
<tr>
<td>( A_R )</td>
<td>20 m</td>
</tr>
<tr>
<td>( k_n )</td>
<td>5</td>
</tr>
<tr>
<td>( k_d )</td>
<td>5</td>
</tr>
<tr>
<td>Simulated time</td>
<td>from 7 to 10 a.m</td>
</tr>
</tbody>
</table>

Table 1: System settings.

The first phenomenon that we study is about the sensitivity of \( C \) with respect to the density of mobile nodes, in particular P nodes. This values are kept constant in all the following experiments. Locality awareness is not used and the simple FIFO buffer management policy is adopted. In Figure 2 the values of \( C \) for P, F and R nodes are reported versus the number of P nodes; no vehicles are simulated in this case. In Figure 2(a) an ideal situation when all nodes have no memory constraints (\( b = \infty \)) is shown, whereas Figure 2(b) refers to the opposite condition when all nodes use the minimum amount of memory (\( b = 1 \)). From the figures it can be noted that increasing the density of P nodes improves the performance of all node classes. Nonetheless, the performance gain turns out to be rather limited when not enough memory is available. In terms of absolute values P nodes always achieve the best coverage as such nodes have the smallest ROI (see Table 1).

In Figure 3 the coverage achieved by the nodes is studied as a function of the memory used by the P nodes. In this case we make the realistic assumption that F nodes do not have any memory limits whereas the pedestrians, being equipped with hand-held devices, can use a limited amount of buffering space for opportunistic communications. Figure 3(a) and Figure 3(b) refer to two different scenarios with 125 and 500 P nodes, respectively. This experiments show that all node classes benefit from a higher availability of memory at the mobile nodes.

We also consider the effect of different mobility patterns of the mobile nodes. To this end, in Figure 4 we show some experiments worked out in a scenario where the overall number of mobile nodes is kept fixed to 250, with a variable percentage of vehicles and pedestrians. In particular, we let the number of vehicles be the 0, 10 and 20 % of the mobile nodes. In Figure 4(a) no memory limits are used and it can be noted that vehicles help in improving the coverage of all node classes. In Figure 4(b) the same experiment is repeated in the case \( b = 1 \), i.e. when all the nodes use the minimum amount of memory. In this latter situation it can be noted that the coverage of P nodes tends to decrease 15 minutes are always considered as transient and therefore skipped in the computation of the average coverage \( C \).

The map data are available at http://udelmodels.eecis.udel.edu/
when a higher number of cars is around. This is clearly due to the fact that in all these experiments the nodes do not implement any form of locality awareness when managing their buffers. Since the V nodes move faster than P nodes, these latter are likely to store in their buffer useless data item transported by V nodes coming from a different area that in turns decreases their coverage.

The adoption of the SD and SI locality aware policies allows one to overcome the issue we have just commented on. Table 2 compares the values of $C$ obtained in the different node classes in the scenario with 200 pedestrians and 50 vehicles as a function of different system settings, namely the FIFO, SD and SI buffer management with and without memory limitation. The SD/SI column refers to the case when the two buffer management strategies are activated jointly. It is worth noting that the adoption of SD or SI highly improve the coverage yielding a good performance even in the presence of most severe memory constraints, i.e. $b = 1$ for all node classes. As an example, from the results in Table 2 it turns out that using SD or SI both P and V nodes are able to achieve a coverage in the case $b = 1$ that is very close to what they would get using unconstrained memory.

To complete the study on different mobility pattern we added to the system a limited number of R nodes with unconstrained mobility. In Table 3 the values of $C$ obtained by adding 1 and 5 R nodes when $b = 1$ are compared with those obtained without R nodes and $b = \infty$. The scenario is the same as in Table 2. By comparing the two tables, it can

<table>
<thead>
<tr>
<th>Class</th>
<th>Policy</th>
<th>FIFO</th>
<th>SD</th>
<th>SI</th>
<th>SD/SI</th>
<th>FIFO</th>
</tr>
</thead>
<tbody>
<tr>
<td>P</td>
<td>$b = 1$</td>
<td>0.30</td>
<td>0.55</td>
<td>0.57</td>
<td>0.58</td>
<td>0.68</td>
</tr>
<tr>
<td>V</td>
<td>$b = 1$</td>
<td>0.15</td>
<td>0.45</td>
<td>0.49</td>
<td>0.49</td>
<td>0.48</td>
</tr>
<tr>
<td>F</td>
<td>$b = 1$</td>
<td>0.53</td>
<td>0.56</td>
<td>0.57</td>
<td>0.57</td>
<td>0.56</td>
</tr>
</tbody>
</table>
be noted that adding a single R node increases the coverage of P, V and F nodes from 0.55, 0.45, 0.56 up to 0.71, 0.57, 0.68. Using 5 R nodes enhances the performance further.

In all previous experiments we assumed that the information $i(x, y, t)$ associated to the urban area is kept fixed for all the 3 hours simulation. Clearly, adding a temporal dynamic to the information has a negative impact on the overall system performance. To take this effects into account, we performed a last set of simulations where one information item is updated every $d_s$. The computation of the coverage index $C$ is slightly modified so as to consider only the updated information. In other words, the coverage is now defined as the percentage of updated information collected by the node buffer that falls within the ROI. The obtained results are shown in Figure 5, that reports the coverage as function of the update interval $d$ in the scenario with 200 P, 50 V and 5 R nodes using SD and $b = 1$. As expected, one can observe that frequent updating of the information negatively impact on $C$. Nonetheless, we can point out that in the presented scenario the proposed system is able to sustain 1 update every 30 s without a noticeable performance impairment.

5. CONCLUSIONS
In this paper we considered data gathering in urban scenarios using opportunistic networking. The area under study is discretized in tiles, each defining a data item that nodes can acquire and share through opportunistic transmissions to other nodes in their proximity. Nodes memory is limited and is proportional to the nodes computing resources, i.e., hand-held terminals are assumed to have small buffers while cars and fixed nodes enjoy greater storage capabilities. The distinctive features of our work lie in the exploitation of detailed mobility and radio propagation traces generated by the UDelModels tools and in the particular abstraction for data gathering: each node aims to retrieve the data items in its ROI centered around the current node position. Since data items may change over time all nodes must strive for having access to the latest version. Furthermore, for mobile terminals ROI is a time varying concept due to the dynamic behavior of pedestrians and cars.

We analyzed the performance of the system by defining and estimating the percentage of the ROI covered by the items stored in the buffer of each node (the node coverage) for both static and dynamic information. The preliminary findings highlight that simple location aware memory management schemes effectively exploit nodes with limited amount of memory. Furthermore, increasing randomness of nodes movement by adding a few ideal nodes whose mobility is described by an unconstrained Brownian motion proved to have a beneficial impact on the average coverage of all node types.
Future extensions of the current work are many: first of all we are conducting more simulation experiments to explore a wider range of values for the main system parameters. Longer term research will consider the impact of coding techniques [19] on the coverage of nodes as well as the use of compressive sensing [6, 3] concepts for sparsely defined information. Energy related issue will be taken into account to trade off node coverage versus battery life for hand-held devices. Finally, we are currently exploring the research area on spatio-temporal databases [4] to import ideas and techniques to improve our work.

6. REFERENCES


