Exploiting Schelling Behavior for Improving Data Accessibility in Mobile Peer-to-Peer Networks*

Long Vu, Klara Nahrstedt
University of Illinois at Urbana-Champaign
Illinois, USA
{longvu2,klara}@cs.uiuc.edu

Matthias Hollick
Technische Universität Darmstadt
Darmstadt, Germany
mhollick@kom.tu-darmstadt.de

ABSTRACT

In 1969, Thomas Schelling proposed one of the most cited models in economics to explain how similar people (e.g. people with the same race, education, community) group together in American neighborhoods. Interestingly, we observe that the analogies of this model indeed exist in numerous scenarios where co-located people communicate via their personal wireless devices in Peer-to-Peer (P2P) fashion. Schelling's model therefore can potentially serve as a mobility model and offer a unique opportunity to efficiently disseminate messages in mobile P2P networks. In this paper, we exploit the natural grouping and moving behavior of humans presented by Schelling to expedite data dissemination in such networks. Particularly, we design a push model for dense network areas to maximize data dissemination and a pull model for sparse network areas to utilize network bandwidth and node energy efficiently. We ensure that our scheme is lightweight since queries and responses are automatically limited within groups of mobile nodes carried by similar people. Moreover, we avoid broadcast storms by assigning each message a broadcast timer and applying overhearing mechanism to reduce redundant transmissions. Finally, our simulation results show that the proposed data dissemination scheme improves the query hit ratio significantly while utilizing network bandwidth efficiently.

1. INTRODUCTION

An extremely large percentage of personal devices (e.g. cell phones, PDAs, Zune) are now equipped with wireless network interfaces. This opens the door to a wide range of decentralized and ubiquitous communications in which personal wireless devices can collaboratively communicate in P2P fashion. Content exchanges and data dissemination in these P2P networks, therefore, become increasingly important and draws significant attention from the research community [3, 12, 14, 15, 20, 21, 22]. Previous studies on data dissemination in wireless networks fall into two main categories. The first category is for sparse and intermittently connected networks where mobile nodes use a store-carry-forward paradigm and only exchange messages once they have physical contact. This model follows the idea of Delay Tolerant Networks [2, 11, 23] but provides no guarantees about the quality such as coverage and delay of dissemination. The second category is for dense networks where mobile nodes are assumed to move according to a repetitive pattern [4] or in group [13, 19]. For mission-based networks such as military or first-responder networks, grouping methods work efficiently since nodes are pre-configured to move in groups, thus always staying in close proximity during their missions to cooperatively expedite data dissemination.

Although the strong grouping assumption is typical in mission-based network research, it becomes less realistic in civilian and commercial scenarios where people often move with no pre-configurations. As a result, the grouping of mobile devices carried by these people can occur instantly but not permanently. Let us consider a shopping street scenario, where customers walk to their interested shops and exchange messages via Bluetooth or 802.11 wireless interfaces of their personal wireless devices. Given two co-located customers A and B, according to previous grouping algorithms [19], A and B will be grouped and required to collaboratively forward messages. However, A and B may have different targeted shops, so they may move towards different directions in very near future, causing their wireless connection to break. Moreover, if A is interested in jewelry and B is interested in digital cameras, what is the immediate incentive for A to disseminate the packet about digital cameras from B, and vice versa? In this case A and B fail to collaborate, although grouped. We therefore believe that sharing mutual interests is crucial to motivate people (with their personal wireless devices) to collaboratively disseminate messages.

Interestingly, we observe numerous scenarios and applications where people motivate themselves to collaborate if they share interests on some topic. For example, audiences of an exhibition or students in a campus can group to exchange messages while heading towards the same destinations such as exhibition halls or classrooms. More interestingly, grouping and moving behavior of people who share similarities was presented in one of the most cited model in economics by Thomas Schelling [17]. According to the model, people move apart from each other if they have different interests; whereas, they group if they share mutual interests.

In this paper, we exploit the natural grouping and moving behaviors of humans presented by Schelling to expedite

---

*This work is supported by Boeing Trusted Software Center at the Information Trust Institute, University of Illinois.

MobiQuitous 2008, July 21-25, 2008, Dublin, Ireland

Digital Object Identifier: 10.4108/ICST.MOBIQUITOUS2008.3870
http://dx.doi.org/10.4108/ICST.MOBIQUITOUS2008.3870

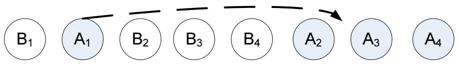


Figure 1: In Schelling's original model, people always move towards their similar neighborhoods

data dissemination in mobile P2P networks. Particularly, we first introduce two important properties of Schelling's model. Then, we leverage these properties to design a push model for dense network areas to maximize data delivery and a pull model for sparse network areas to utilize network bandwidth and node energy. Our scheme is lightweight since queries and responses are automatically limited within groups of mobile nodes carried by people with mutual interests. Moreover, we avoid broadcast storms by assigning each message a broadcast timer and applying overhearing mechanism to reduce redundant transmissions and collisions. Our scheme also allows leaving and arriving nodes, who share interest, to collaboratively answer queries and thus further improve data accessibility. Finally, our simulation results show that the proposed data dissemination scheme improves the query hit ratio significantly while utilizing network bandwidth efficiently, avoiding broadcast storms, and minimizing transmission collisions.

The paper is organized as follows. Section 2 presents Schelling's original model, its properties, and representative examples demonstrating the coexistence of Schelling's model and wireless technologies. Then, we discuss system models and system overview in Section 3. Next, we present our basic data dissemination protocol in Section 4 and the improvements for this basic protocol in Section 5. Section 6 evaluates our scheme and protocols while Section 7 summarizes related work. Finally, Section 8 concludes the paper.

2. SCHELLING BEHAVIOR

2.1 Schelling's Original Model

In 1969, Thomas Schelling, a Nobel-prize winner in economics, proposed one of the most cited models in economics to explain how similar people (i.e. people with same race, education, community) group in American neighborhoods [17]. According to Schelling, the grouping is created by movements of individuals who want at least a certain portion of similar neighbors. In other words, when a person is unsatisfied with his neighborhood, he moves towards a place where he has more similar neighbors. Such movements eventually create clusters of similar individuals. Figure 1 shows the idea of Schelling's model where a circle denotes an individual and shading patterns represent different interests. In this example, A_1 moves towards its closest and similar neighbor, A_2. When everyone is satisfied with their neighborhoods, the clustering reaches the stable equilibrium. In what follows, we use the terms group and cluster interchangeably.

2.2 Analysis of Schelling's Model

In this section, we introduce two important properties of the Schelling's model, which are later exploited by our protocols for improving data accessibility in mobile P2P networks.

2.2.1 *Density of similar individuals increases in proximity of clusters*

This property directly follows Schelling's original model since individuals move towards their desired neighborhoods and thus create clusters of similar individuals at these neighborhoods. As a result, the density of similar individuals increases significantly in proximity of these clusters.

2.2.2 *Similar individuals form small "moving" clusters during their movements*

According to Schelling, each individual always moves to his final cluster where he is satisfied with the neighborhood and stays. In Schelling's model, on the ways to their final clusters, similar individuals form small clusters. However, individuals at the boundary of these clusters may not be satisfied with their current mixed neighborhoods. Thus, they tend to move towards bigger clusters where they have better (similar) neighborhoods. When an individual at the boundary leaves, other inner individuals form the boundary; this again might cause them to leave. This process creates small "moving" clusters, which merge to bigger clusters.

2.2.3 *Schelling Behavior*

Schelling's original model focuses on economic and social phenomena where individuals gradually form groups on a very large timescale. For example, the formation of a China town in a city might take decades. However, in the context of wireless technologies, we observe numerous scenarios where mobile wireless devices carried by similar people (people share mutual interests on some topic such as books, music, movies) exhibit Schelling's model on a much smaller timescale (see Section 2.3)[1]. For instance, co-located customers can group for 20 minutes and exchange their mutually favorite product information via their wireless handheld devices, while heading towards the same mall. Further, Schelling's model originally focuses only on the outcome of the grouping process (or the final clusters). Meanwhile, we observe that in the context of wireless technologies, not only the outcome but also the grouping process itself can be exploited to expedite data dissemination. This motivates us to study the analogies of Schelling's model (instead of the original model), where the two above properties exist in a much smaller timescale and the grouping occurs during the physical movements of people carrying wireless devices. In what follows, we use the term Schelling behavior to denote the analogies of the Schelling's original model.

2.3 Schelling Behavior & Wireless Technology

There are many real world scenarios where mobile wireless technology and Schelling behavior indeed co-exist.

Our first scenario can be found in the commercial sector. Let us consider a shopping street where customers cluster while arriving at their targeted shops. In this scenario, wireless base stations at shops can broadcast product advertisements, hot sales, discounts. Meanwhile, customers are individuals in Schelling's model who walk to shops and can form groups (these groups of customers is moving towards the shops as second property) to exchange their opinions, reviews, and comments about their mutually interested products via their wireless personal devices. The density of customers gets maximum at the shops (the first property).

Our second scenario is a campus life where places such as book stores, libraries, and class rooms are visited frequently

---

[1] This behavior in mobile computing scenarios introduces higher oscillation in information availability; however, we introduce damping to ensure longer information access (see Sections 4 and 5).

Digital Object Identifier: 10.4108/ICST.MOBIQUITOUS2008.3870
http://dx.doi.org/10.4108/ICST.MOBIQUITOUS2008.3870

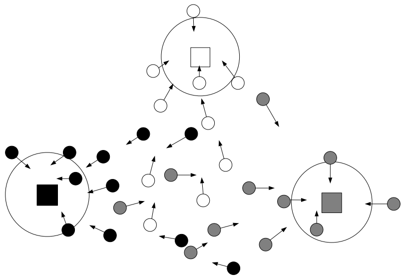


Figure 2: Network model

by university students. These places represent final clusters and students represents moving individuals in Schelling's model. Similar to shopping street scenario, wireless base stations at these places broadcast announcements and advertisements to the coming students. Meanwhile, coming students can form groups due to their co-locations and similar targeted places to exchange their information during their movements. Again, the density of students gets maximum at these places (the first property).

Our third scenario is a social event such as an art exhibition or a music concert in the downtown area of a city. The event "attracts" interested audiences and plays the role of a final cluster in Schelling's model. Wireless base station at the event can broadcast advertisements, content and showtimes of the event to arriving audiences. These audiences can form groups and exchange their opinions and comments about the event via their personal wireless devices.

For a generalized presentation, we denote the shops, bookstores, and exhibitions in the above scenarios as *Points of Interest* (*PoI*s). We believe these above examples represent a popular class of scenarios where these *PoI*s and mobile individuals essentially create mobile P2P networks and Schelling behavior occurs. In these networks, disseminating messages from *PoI*s to mobile individuals and among mobile individuals efficiently is challenging due to the dynamics and diversity. In this paper, we present a data dissemination scheme to address this challenge. In what follows, we use mobile individuals and mobile nodes interchangeably.

3. SYSTEM MODEL AND OVERVIEW

In this section, we first present our network model and data model. Then, we discuss our design objectives. Finally, we present the overview of our system.

3.1 Network Model

We focus on a hybrid mobile P2P network where each *Point of Interest* (*PoI*) has a wireless base station and a server which processes requests from mobile nodes. The base station periodically broadcasts messages from the server to the surrounding area of the *PoI*. We assume that Schelling behavior occurs and its two properties hold. That means, the network density at *PoI*s is very high and co-located nodes can group if they share mutual interests. Figure 2 shows our network where squares represent *PoI*s and circles are nodes whose arrows denote movement directions. In Figure 2, different shading patterns represent different interests. We assume that all wireless devices communicate via a common channel using IEEE 802.11 or Bluetooth. All mobile nodes have the same transmission range and distance between the two nodes within the transmission range can be estimated by various techniques [9] or using GPS. Each mobile node n can communicate with the server via the base station in the infrastructure mode and with other nodes in the ad hoc mode. Although it can change interest any time, a node n moves toward only one targeted PoI at one time. We assume there is a registration protocol that allows n to choose its targeted PoI. For example, at the gateway of the shopping or campus area, there exists a server broadcasting available PoIs to coming users and so their personal devices can capture these PoIs for future reference. n has a limited amount of cache (memory) to store messages. We also assume nodes with the same interest are cooperative.

3.2 Data Model

In this paper, we consider messages including text, images, short video clips created by the PoIs. For example, messages can be advertisements and discount in commercial applications, or lecture announcements in the campus life scenario. The broadcast frequency of messages depends on their popularity, which is determined by the PoIs. For example, for a hot sale with big discount, the shop will advertise/broadcast more frequently than other sales. Without loss of generality, we assume the popularity of messages follows a *Zipf like* distribution:

$$f(r;\theta;N) = \frac{\frac{1}{r^{\theta}}}{\sum_{i=1}^{N}\frac{1}{i^{\theta}}} \tag{1}$$

In Equation 1, N is the total number of messages created by one PoI and r is the rank of a message. When θ is equal to one, the *Zipf like* distribution becomes the classic *Zipf* distribution. We also assume the query/request of nodes follows the above *Zipf like* distribution since in reality people usually request information from more popular items [1].

3.3 Design Objectives

Our design objective is to develop a data dissemination that (i) allows a PoI to efficiently spread its advertisements, sales, and other information to as many coming nodes of its interest as possible, and (ii) allows coming nodes to query for their interested information as soon as they comes closer to the PoI. In particular, the data dissemination scheme should maximize data access of nodes following Schelling behavior, especially for ones in close proximity of the PoI. This is intuitive because when the mobile users arrive closer to a PoI, they expect to have a more timely access to the data of their targeted PoI. Moreover, the scheme should reduce redundant transmissions to save node energy, avoid broadcast storms, and minimize transmission collisions.

3.4 System Overview

In this section, we briefly present our system design. Figure 3 shows the overview of our network in which a square represents a PoI and circles represent mobile nodes. The dotted circles denote wireless broadcast (transmission) ranges. In our network, a mobile node moves toward its targeted PoI (e.g. nodes n_7 and n_8) and after it arrives at its targeted PoI, it stays (nodes n_3, n_4, n_6), and then leaves (nodes n_9, n_{10}). To cover the high density of nodes surrounding a PoI, we use a push model where the PoI creates a *Message Reachability Zone* by assigning each message m a Time-To-Live (TTL) value. This TTL specifies how many forwarding actions nodes in *Message Reachability Zone* perform on m.

Digital Object Identifier: 10.4108/ICST.MOBIQUITOUS2008.3870
http://dx.doi.org/10.4108/ICST.MOBIQUITOUS2008.3870

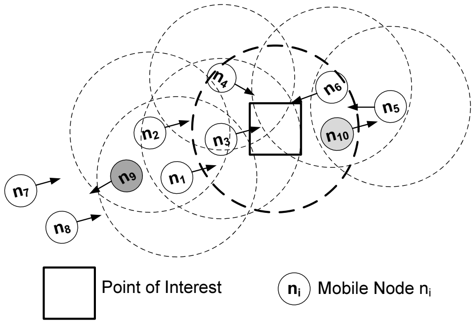


Figure 3: System overview

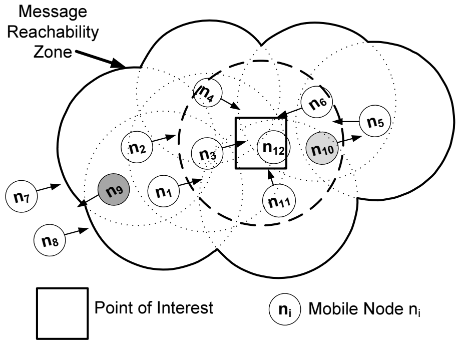


Figure 4: The *Message Reachability Zone* of a *PoI*

On receiving m, a node n computes a broadcast timer for m. Later, when m's timer expires and n did not hear any nodes broadcasting m, n re-broadcasts m. The timer is used to avoid broadcast storms and it is re-estimated whenever n overhears m being sent by other nodes. Outside the *Message Reachability Zone*, due to the sparse network, mobile nodes follow a pull model by sending queries to their neighbors to save network bandwidth and energy. To further improve data accessibility, similar nodes can automatically group and communicate in ad hoc mode to forward requests and answer queries. In the following sections, we present in detail the system and protocol designs.

4. DATA DISSEMINATION PROTOCOL

We present the basic dissemination protocol in this section and the improved protocol in Section 5.

4.1 Message Reachability Zone

According to the first property of Schelling behavior, density of similar nodes in proximity of a PoI increases significantly. Thus, we use a push model to disseminate messages to this dense area. Particularly, the base station periodically broadcasts messages and assigns a TTL value for each message m. Receiving m, nodes in proximity of the PoI cooperatively rebroadcast m to create a *Message Reachability Zone*. To be precise, "a *Message Reachability Zone (MRZ) of a* PoI *is an area covered by broadcasts of mobile nodes arriving and staying at the* PoI." According to this definition, the size of a MRZ is not fixed. Instead, it depends on TTL values of the messages, transmission ranges and speeds of relaying nodes (including *Arriving* and *Staying* nodes in Section 4.2). Particularly, if m has a larger TTL or relaying nodes have larger transmission ranges, the size of MRZ is bigger. In contrast, if the relaying nodes have higher speed, the size of the MRZ is smaller. Notice that n only rebroadcasts m if m's TTL is positive and anytime m is broadcast or overheard (see Section 5.1.2), its TTL decreases by 1. Figure 4 shows a MRZ with the solid curve boundary.

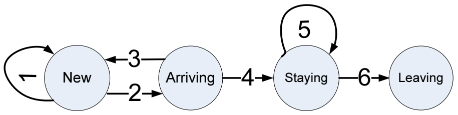


Figure 5: Mobile nodes can be in one of the four states: *New*, *Arriving*, *Staying*, and *Leaving*.

The above push model expedites messages for nodes inside the MRZ. In the next section, we present how mobile nodes in the entire network cooperatively disseminate messages.

4.2 Mobile Node States

Figure 5 shows a state machine where each circle is a state of a mobile node n in our network. In following sections, we discuss characteristics, the transitions, and corresponding protocols of mobile nodes at each state.

4.2.1 New Node

A mobile node n is in the *New* state if n is outside the MRZ of its targeted PoI. For example, in Figure 4, nodes n_7, n_8 are in *New* state since they are outside the MRZ. A node n can detect the existence of the MRZ by overhearing messages broadcast by nodes at the boundary of the MRZ.

The main communication mode of *New* nodes is ad hoc because they can not directly reach base stations. Whenever a *New* node n has a query q, n follows a pull model by broadcasting q to its similar neighbors and waiting for the answer from them. If n has a bigger similar neighborhood, q may be relayed by nodes in this neighborhood, to nodes inside the MRZ. Thus, q can be answered by nodes inside the MRZ or the PoI (see Section 4.3). The pull model is used because according to Schelling behavior, outside the proximity of the PoI, the density of similar nodes is very low. The pull model thus can save node energy, network bandwidth, and reduce interference. A *New* node n has two transitions: 1 and 2. The former occurs when n switches to a new interest and starts moving towards the new PoI. The latter occurs when n enters the MRZ of its current PoI and changes state to *Arriving*.

4.2.2 Arriving Node

At *Arriving* state, a mobile node is inside the MRZ. For instance, node n_1 in Figure 4 is an *Arriving* node. The role of *Arriving* nodes is to rebroadcast and relay messages to create the MRZ. All *Arriving* nodes receive a rich set of information via broadcasts of similar neighborhoods. To help *New* nodes detect the boundary of MRZ, in its broadcast messages, the *Arriving* node n adds a flag to mark its *Arriving* state. Using this flag, *New* nodes can distinguish the broadcast messages from other query/response messages and thus can detect the boundary of the MRZ.

Communication among *Arriving* nodes is ad hoc. Whenever an *Arriving* node n has a query q, n broadcasts q to its similar one-hop neighbors, which in turn can relay q to the PoI. Eventually, PoI will answer q if no nodes inside MRZ can answer. This means all queries of *Arriving* nodes have a very high chance to be answered. From *Arriving* state, n can switch to *New* or *Staying* state (transitions 3 and 4 in

Digital Object Identifier: 10.4108/ICST.MOBIQUITOUS2008.3870
http://dx.doi.org/10.4108/ICST.MOBIQUITOUS2008.3870

Figure 5). Switching to *New* state means the mobile node changes interest while switching to *Staying* state means the mobile node enters the transmission range of the base station of its targeted *PoI*.

4.2.3 Staying Node

At *Staying* state, a node n is inside the transmission range of the *PoI*'s base station, its queries will be answered instantly by its local cache or the *PoI*. For example, node n_3 in Figure 4 is a *Staying* node. A *Staying* node has two communication modes: infrastructure and ad hoc. The former is used to communicate with the *PoI* and other *Staying* nodes while the latter is for communication with *Arriving* nodes. *Staying* nodes can relay queries of *Arriving* or *New* nodes to the *PoI* and responses/answers from the *PoI* back to these nodes. *Staying* nodes also take part in creating MRZ by rebroadcasting messages. Similar to *Arriving* nodes, *Staying* nodes learn a rich set of information from the base station via periodic broadcasts. In Figure 5, a *Staying* node can switch to *Leaving* state by transition 6. This occurs when a node changes interest or leaves the network.

4.2.4 Leaving Node

When a *Staying* node switches interest or leaves the network (e.g. node n_9 in Figure 4), its state becomes *Leaving*. In the first case, a *Leaving* node of one interest becomes a *New* node of another interest. If n switches from the *Staying* state to the *Leaving* state, it resets all TTL values of messages in its cache. At the same time, it stops relaying queries/responses for nodes of its old interest. However, a *Leaving* node n can answer queries (via ad hoc mode) for coming nodes of its old interest (Section 5.2).

Algorithm 1 Query and Response

```
INPUT: a mobile node $n_1$ has a query $q$ for a message $m$;
$n.i$: interest of $n$
OUTPUT: $n_1$ has a query hit or a query miss
BEGIN
if ($m \in n_1$'s cache) then
  return query hit
else
  if ($n_1$ is a Staying node) then
    $n_1$ sends $q$ to the PoI, which can answer $q$.
  else
    $n_1$ broadcasts $q$ to its neighbors. Suppose $n_2$ receives $q$.
    if (($m \in n_2$'s cache)&($n_1.i = n_2.i$)) then
      if ($n_2$ is $n_1$'s one-hop neighbor) then
        $n_2$ returns the response message to $n_1$ directly
      else
        $n_2$ returns the response message to $n_1$ via an underlying routing protocol such as AODV or DSR.
      end if
    else
      if (($n_2.i = n_1.i$) & ($n_2$ hasn't broadcast $q$)) then
        $n_2$ broadcasts $q$ on behalf of $n_1$. The process repeats.
      else
        $n_2$ discards the query $q$
      end if
    end if
  end if
end if
$n_1$ waits for a timeout, if no answer then $n_1$ has a query miss
END
```

4.3 Limiting Query Scope

Because Schelling behavior exists in our network, the second property hold. That is, if co-located mobile nodes share mutual interests, they can automatically form groups during their movements towards the *PoI*s. We exploit this property to improve our query and response process. In particular, when a node n_1 has a query for a message m, it performs Algorithm 1. In this algorithm, whenever a node n_2 receives q if n_1 and n_2 have different interests, n_2 will not forward q. Thus, the query flooding is limited within the similar nodes of n_1. This is because Schelling behavior occurs, groups of similar nodes automatically exist in the network and limit the scope of query flooding. With this automatic grouping, our scheme does not need any group management protocols and thus reduces communication overhead. To avoid query broadcast storms, node n_2 only forwards the query q from n_1 once. If n_2 receives the same query q, it discards q.

4.4 Cache Management

Due to its limited cache size, when a node n receives message m, n performs the Algorithm 2 to keep most popular messages in its cache.

Algorithm 2 Cache Replacement

```
INPUT: $n$ receives $m$ and $m$ belongs to $n$'s interest
OUTPUT: $n$ finishes cache replacement
BEGIN
if ($m \in n$'s cache) then
  $n$ re-estimates $m$'s broadcast timer as in the Algorithm 3
else
  if ($n$'s cache is full) then
    if (all messages belong to $n$'s current interest) then
      $m_1$ = the least popular message in $n$'s cache
      if ($m_1$ is less popular than $m$) then
        $m$ replaces $m_1$
      else
        $n$ discards $m$
      end if
    else
      //There are several messages of $n$'s old interest
      $m$ replaces the least popular mess. of $n$'s old interest
    end if
  else
    $n$ adds $m$ into its cache
  end if
end if
END
```

5. IMPROVING DATA ACCESSIBILITY

To improve the efficiency of the basic scheme presented in Section 4, we present the broadcast storm avoidance and context-switching.

5.1 Broadcast Storm Avoidance

To reduce message overhead and avoid broadcast storms inside the MRZ, each message m will be assigned a broadcast timer by its receiver n. n will re-estimate the broadcast timer of m whenever n overhears m from its neighborhood.

5.1.1 Broadcast Timer Estimation

When an *Arriving* or *Staying* node n receives a message m, n estimates m's broadcast timer as follows:

$$m.broadcastTimer = \frac{TX}{Dist(sender, n)} \cdot TimeUnit \qquad (2)$$

In Equation 2, *m.broadcastTimer* denotes the value of m's broadcast timer estimated by n. Notice that n can later be the broadcaster of m when m's timer expires. TX is n's

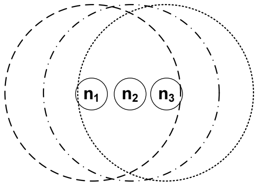


Figure 6: n_1 is the broadcaster, n_2, n_3 are receivers. The most distant node, n_3, will be the next broadcaster

transmission range and $TimeUnit$ specifies the scale of the broadcast timer (e.g. seconds or minutes). The distance between *sender* and n (i.e. $Dist(sender, n)$) can be estimated by various techniques [9] or using the GPS.

In Equation 2, if n_1 is the sender of m and n_2, n_3 are receivers, and if $Dist(n_1, n_2) < Dist(n_1, n_3)$, n_3 will assign a shorter broadcast timer to m. Then, n_3 will broadcast m prior to n_2, resulting in a larger region covered by m in the network. Figure 6 shows an example where n_1 is the original broadcaster of m. n_3 will be the next broadcaster because it is farther from n_1 than n_2.

5.1.2 Overhearing Mechanism

Besides broadcast timer, we use the overhearing mechanism to avoid broadcast storms, reduce transmission collisions and interference. In particular, whenever a node n_3 overhears or receives a message m, n_3 performs the Algorithm 3. By applying this algorithm, once n_3 overhears m, it re-estimates m's broadcast timer and decreases $m.TTL$ by 1. By re-estimating m's timer, n_3 can minimize redundant broadcasts, save node energy and reduce transmission collisions. By decreasing $m.TTL$ by 1, n_3 avoids updating m's broadcast timer, especially when n_3 is in a very dense area and its neighbors broadcast m so frequently. In Figure 6, when n_3 broadcasts m, n_2 re-estimates m's broadcast timer.

Algorithm 3 Broadcast timer update

```
INPUT: $n_3$ receives $m$ via broadcast of $n_1$. $n_3$ and $n_1$ have the same interest
OUTPUT: $n_3$ updates $m$'s broadcast timer.
BEGIN
$tmp = \frac{TX}{Dist(n_1,n_3)} \cdot TimeUnit$
if $m \notin n_3$'s cache then
    $m.broadcastTimer = tmp$;
    $n_3$ adds $m$ and applies cache replacement in Section 4.4
else
    if ($m \in n_3$'s cache) & ($m.TTL > 0$) then
        $m.broadcastTimer = tmp$;
        $m.TTL = m.TTL$-1;
    else
        $n_3$ discards $m$;
    end if
end if
END
```

5.2 Context-switching

Essentially, when leaving the old PoI, a node n can be a New node of another interest or n leaves the network. At the moment, n has a rich set of information about its old PoI, which can be used to improve query hit for coming nodes to n's old PoI. Let us consider the first case, n switches from its old interest PoI_1 to its new interest PoI_2. When n gets closer to PoI_2, n learns more about PoI_2 through its queries. Thus, n's cache content changes gradually, with more and more messages of PoI_2 replacing messages of PoI_1. This is where the context-switching occurs. In our scheme, when n switches interest to PoI_2, although n stops broadcasting messages of PoI_1 in its cache, n still answers queries from coming nodes to PoI_1 whenever its cache has the answers. The context-switching therefore depends on n's query for PoI_2 and n's cache replacement policy. The longer n keeps data of PoI_1, the better n can support coming nodes to PoI_1. For instance, in Figure 4, PoI_1 is the square, when n_9 is leaving PoI_1 for PoI_2, it meets n_7. If n_7 has a query q for a message m of PoI_1 and n_9 receives q, n_9 can answer if m still exists in n_9's cache. Likewise, n_9 may learn about its new interest PoI_2 from nodes leaving PoI_2 before n_9 enters the MRZ of PoI_2. Notice that in context-switching, leaving nodes and coming nodes are considered to share partially mutual interest. Similarly, if nodes leave the network, they also can help coming nodes to improve data accessibility.

6. EVALUATION

We implement a Java-based simulation in middleware layer to evaluate our scheme. Particularly, the simulation focuses on the design of the data dissemination protocol rather than the network stack and routing protocols. In this section, we first describe simulation settings, which result in Schelling behavior. Then, we rely on the Schelling behavior to evaluate our data dissemination scheme.

6.1 Existence of Schelling Behavior

Table 1 presents the settings we use to simulate Schelling behavior. We implement a mobility model operating on a Manhattan grid model where PoIs are at the intersections of streets. A mobile node n in our simulation works as follows. Initially, n obtains a position, a speed, and a PoI, all at random. Then, n starts moving along with streets and towards its $PoIs$. During its movement, n also can switch interest with probability p (in our simulation p=0.05 means 50 nodes change interests per second). When arriving at its PoI, n stays for a random period from 10 to 50 seconds (notice that if this period is longer, nodes cluster more at PoI and our scheme works better). After staying at the PoI, n changes interest to a new PoI, and repeats the entire process. In this paper, we study the *steady state* of our simulation. We expect that if we use settings in Table 1, at *steady state*, the simulation results in Schelling behavior.

Table 1: Network settings

Field	**Value/Unit**
Number of nodes	1000
Number of PoIs	6
Node speed	random [1.0,2.0](m/s)
Area of simulation	1000x1000(m^2)
Node and PoI tran. range (TX)	[50,75,100](m)
Staying period at a PoI	random [10,..,50](s)
Probability of changing interest (p)	random [0.01,0.08]

Figure 7 shows node distribution in *steady state*. Each shape in this figure denotes one interest, a big shape represents a PoI and a small shape is one mobile node. When a node changes to a new interest, its shape changes accordingly. Figure 8 shows node distribution of one particular interest. These two figures confirm that at *steady state* the Schelling behavior exists because (1) closer to a PoI, the density of nodes interested in this PoI increases and (2) similar nodes can group into clusters due to their close proximi-

Digital Object Identifier: 10.4108/ICST.MOBIQUITOUS2008.3870
http://dx.doi.org/10.4108/ICST.MOBIQUITOUS2008.3870

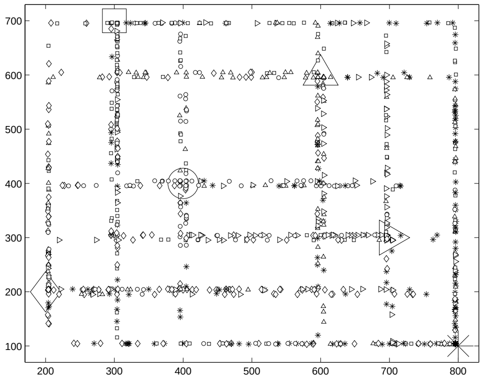


Figure 7: Schelling behavior exists in steady state. Big shapes are PoIs and small shapes are mobile users

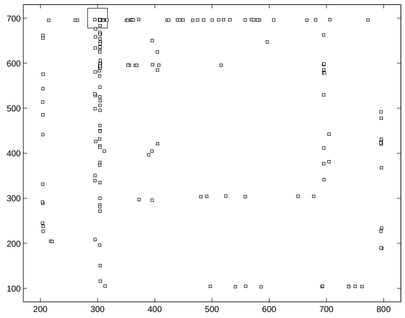


Figure 8: Schelling behavior for one interest

ties on the ways towards their mutual targeted PoIs. Figure 8 also shows that density of similar nodes gets maximized at the PoI and decreases gradually at farther distance. To further validate Schelling behavior in our simulation, we define the notion of *Interest Group* as follows:

1. **if** $(n_1.i = n_2.i)$ & $InTransmissionRange(n_1, n_2)$ **then** *InterestGroup*(n_1, n_2)

   InterestGroup(n_1,n_2) means n_1 and n_2 are in one *Interest Group* and $InTransmissionRange(n_1, n_2)$ means n_1 and n_2 are within transmission range of each other.

2. **if** *InterestGroup*(n_1,n_2) & *InterestGroup*(n_1,n_3) **then** *InterestGroup*(n_2,n_3)

We then use this *Interest Group* definition to group similar nodes and compute the group size. Figure 9 shows that when a node n is closer to its PoI, its group size increases. This confirms the first property of Schelling behavior. At farther distances, group size varies from 5 to 20. This confirms that mobile nodes can group into small clusters on the ways to their targeted PoIs; thus, the second property of Schelling behavior holds. Given the Schelling behavior, we simulate our data dissemination scheme with the settings in Table 2. Then, we use the metrics defined in Table 3 to evaluate our presented data dissemination scheme.

6.2 Evaluation of Data Dissemination

In our simulation, we vary θ (see Equation 1), node transmission range TX, node memory size M, number of messages in one broadcast of a PoI, TTL, and p to evaluate our proposed data dissemination scheme. Notice that we only consider messages created by the PoIs.

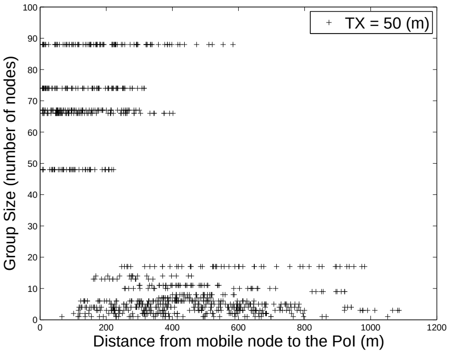


Figure 9: Distance to the PoI and the group size. Each plus sign (+) represents a mobile node

Table 2: Data Management Settings

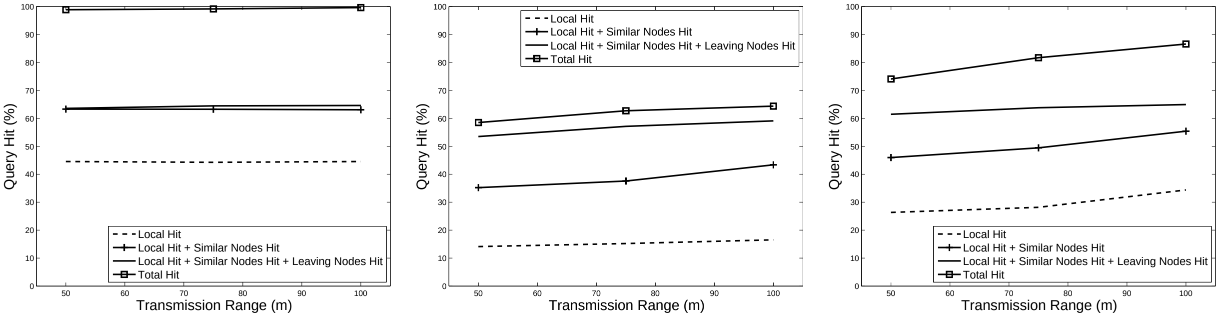
Field	**Value/Unit**
Number of messages created by a PoI	500
Number of messages in one broadcast of PoI	50,75,100
Node memory size M (all nodes are equal)	50,75,100
TTL	1,2,3,4
θ	[0.6,..,1.0]

6.2.1 *MRZ and Context-switching*

In Figure 10, the *Message Reachability Zone* (MRZ) and context-switching improve the "Total Hit". Particularly, Figure 10(a) shows that nodes inside the MRZ have a better "Local Hit" than that in Figure 10(b) because they are more informed by the similar neighborhoods. Meanwhile, "Local Hit + Similar Nodes Hit + Leaving Nodes Hit" of nodes outside the MRZ is slightly less than that of nodes inside the MRZ. This is because inside the MRZ, nodes have similar cache content as they tend to store most popular messages. Thus, when a node fails to answer a request, it is likely its neighboring nodes will fail. In contrast, nodes outside the MRZ have more diverse cache content; thus leaving nodes can contribute more to the query hit of coming nodes. In Figure 10(a), the two curves "Local Hit + Similar Nodes Hit + Leaving Nodes Hit" and "Local Hit + Similar Nodes Hit" look similar. This implies the context-switching is not very effective for nodes inside the MRZ because they can obtain information from their neighbors or the PoIs.

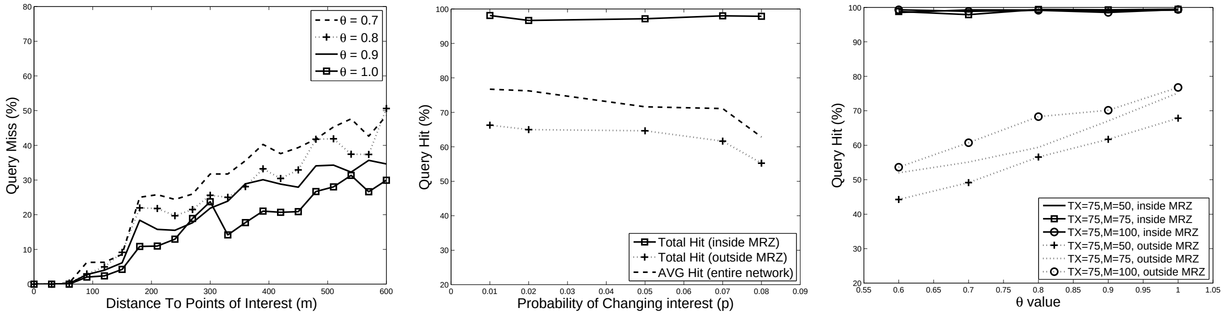
Figures 10(a) and 10(b) show the "Total Hit" of nodes inside the MRZ increases up to 100% while that of nodes outside the MRZ is 64%. This is because nodes inside the MRZ can get answers from the dense neighborhoods and the PoIs. Meanwhile, communication of nodes outside the MRZ is ad hoc in sparse areas. This also explains when TX increases, the "Total Hit" of nodes inside the MRZ varies slightly. Similarly, due to the sparse network, when TX increases the "Total Hit" of nodes outside the MRZ doesn't change significantly. However, for the average query hit of the entire network in Figure 10(c), the "Total Hit" increases about 15% (to $\sim$ 87%). This is because when TX increases, the MRZ becomes larger and thus more nodes are inside the MRZ, which improves average query hit ratio. In conclusion, for nodes inside the MRZ, our scheme is robust with

Digital Object Identifier: 10.4108/ICST.MOBIQUITOUS2008.3870
http://dx.doi.org/10.4108/ICST.MOBIQUITOUS2008.3870



(a) For nodes inside the MRZ (b) For nodes outside the MRZ (c) For the entire network

Figure 10: The MRZ and Context-switching improve the "Total Hit" significantly ($\theta = 0.8$,TTL=2,M=75)



(a) Distance sensitivity (b) Changing interest (c) θ sensitivity ($TTL = 2$)

Figure 11: System sensitivity

Table 3: Definitions of Metrics

Name	**Description/Unit(%)**
Local Hit (L_1)	Query answered by local memory
Similar Nodes Hit (S_1)	Query answered by similar nodes in the neighborhoods
Leaving Nodes Hit (L_2)	Query answered by leaving nodes during their context-switchings
Server Hit (S_2)	is contributed by (i) "Staying" nodes, who directly access the PoI and (ii) "Arriving" and "New" nodes through multi-hop relays
Total Hit	L_1 + S_1 + L_2 + S_2
Query Miss	100 - Query Hit

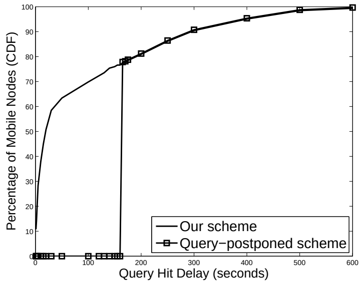


Figure 12: Query Hit Delay

respect to TX. The context-switching concept contributes more for nodes outside the MRZ and for smaller TX.

6.2.2 Impact of Distance and Changing Interest

Figure 11(a) (with TX=75,M=75,TTL=2) shows that at further distance to PoIs, query miss increases. This is expected because when a node n comes closer to its PoI, n observes higher data availability provided by similar nodes in its neighborhood. Thus, n should have higher query hit ratio. Figure 11(b) (with TX=75,M=75,$\theta = 0.8$,TTL=2) shows when the probability of changing interest increases, query hit ratio of nodes inside the MRZ is stable due to the rich set of information within the MRZ. Meanwhile, the query hit ratio of nodes outside the MRZ decreases gradually due to the sparse network density. This confirms the robustness of our scheme to p for nodes inside the MRZ.

6.2.3 Impact of θ and Node Memory Size

Figure 11(c) shows that inside the MRZ, increasing node memory size (M) does not improve the query hit much due to the high availability of data within this area. In contrast, outside the MRZ, increasing M improves query hit considerably, from 5% to 10%. Figure 11(c) also presents impact of θ on data accessibility. In particular, when θ increases to 1, the broadcasts of PoIs and queries of nodes follow a "more" $Zipf$ distribution (Equation 1). Thus, they have a better "match" and provide a higher query hit ratio. Again, impact of θ on nodes inside the MRZ is less significant than that on nodes outside the MRZ because nodes inside the MRZ are well informed by their neighborhoods. Meanwhile, nodes outside the MRZ communicate in ad hoc mode with limited memories. This result together with results in Sections 6.2.1 and 6.2.2 confirms that inside the MRZ, our presented scheme is robust to TX, θ, p and M.

6.2.4 Query Hit Delay and Context-switching

To further evaluate context-switching concept and automatic grouping of mobile nodes, we define the "Query Hit

Digital Object Identifier: 10.4108/ICST.MOBIQUITOUS2008.3870
http://dx.doi.org/10.4108/ICST.MOBIQUITOUS2008.3870

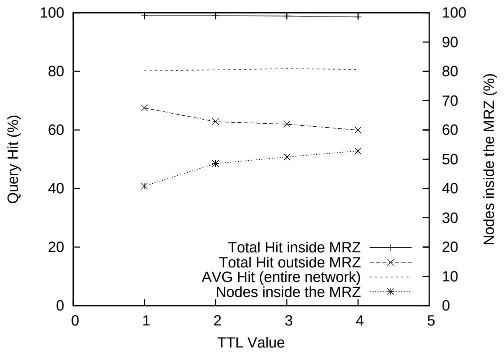


Figure 13: TTL sensitivity and number of nodes inside the MRZ ($TX = 75, M = 75, \theta = 0.8$)

Delay" metric of our scheme for a node n as follows:

$$QueryHitDelay(n) = t_2 - t_1 \quad (3)$$

In Equation 3, t_1 is the time at which n switches to a new interest (and thus n starts moving towards the corresponding PoI). t_2 is the time when n's first query for the new interest gets a hit. Figure 12 (with TX=75,M=75,θ = 0.8,TTL=2) compares our scheme with a query-postponed scheme where node n holds queries since n switches to a new interest until n arrives at the new PoI. At the PoI, n obtains 100% of query hit. In our simulation, "Query Hit Delay" metric of query-postponed scheme can be estimated using distance from n's current position to its new PoI, and the speed of n. Figure 12 shows that our scheme obtains shorter delay (less than 170 (s)) for about 78% of nodes. Meanwhile, the query-postponed scheme has a sudden change at 170 (s) since many nodes arrive at their new PoIs after 170 (s). Particularly, in 170(s) a node can travel from 170(m) to 340(m) and it gets into transmission range of the new PoIs because a few PoI pairs in our simulation are 400(m) and one pair is 300(m) apart. This concludes that our dissemination scheme obtains better access time for all nodes, regardless their distances to their targeted PoIs.

6.2.5 *TTL sensitivity and Message Overhead*

In Figure 13, when TTL increases, the "Total Hit" of nodes inside the MRZ is stable. In contrast, "Total Hit" of nodes outside the MRZ decreases because when TTL increases, the radius of the MRZ also increases (i.e. radius of $MRZ \sim TX \cdot (TTL + 1)$). Meanwhile, the node density is independent of the MRZ and decreases at further distance to the PoI. Thus, for a large MRZ, the node density at its edge is much lower than that of a small MRZ, resulting in lower query hit of nodes outside this large MRZ. However, for the entire network, the average (AVG) query hit is stable (∼82%) because larger TTL provides more nodes inside the MRZ (∼52% of nodes when TTL = 4). These nodes have better data accessibility to their targeted PoIs. This makes the AVG query hit for the entire network stable.

To further evaluate our scheme (TX=75,M=75,θ=0.8), we compare it with two other schemes: $PureFlooding$ and $LimitedFlooding$. These two schemes also rely on Schelling behavior of the network but they have different data dissemination strategies. $PureFlooding$ simply floods messages to the network. Queries are answered by node n itself and nodes within n's transmission range. $LimitedFlooding$ uses TTL to limit flooding. However, it does not have broadcast timers and context-switching concepts. Similar to $PureFlooding$, queries are answered by node n and nodes within n's transmission range. All other simulation settings of three schemes are exactly the same. Figure 14 presents the average (AVG) query hit ratio over the entire networks and average number of messages broadcast (overhead) by one node during the simulation. Particularly, our schemes improves the query hit ratio more than 20% because mobile nodes collaboratively relay and answer queries. Our scheme also outperforms $PureFlooding$ and $LimitedFlooding$ in terms of average message overhead. In particular, when TTL increases up to 4, each mobile nodes in our scheme sends less than 200 messages while the overhead of $LimitedFlooding$ and $PureFlooding$ are about 600 and 10000, respectively. In conclusion, our proposed scheme improves significantly average query hit ratio while minimizing message overhead.

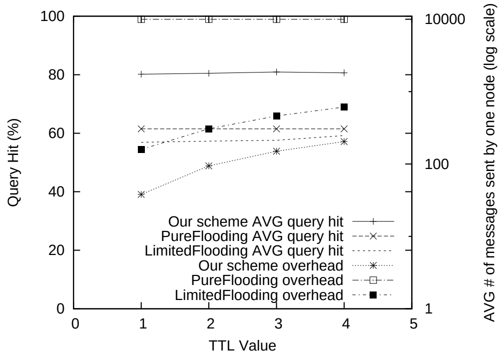


Figure 14: Our scheme improves significantly average query hit ratio while minimizing message overhead

7. RELATED WORK

We first summarize previous studies on data dissemination in wireless networks. Then, we discuss limitations of existing mobility models currently used to evaluate data dissemination schemes in mobile P2P networks.

7.1 Data Dissemination Strategies

The first approach is broadcast-based data dissemination [11, 14, 16, 20, 21], which tries to adapt the dynamic and unstable nature of wireless networks. The broadcast, therefore, is the intuitive way to disseminate messages. However, blind broadcast causes broadcast storms and hurts network bandwidth. To avoid this, numerous methods have been proposed [5, 10, 18], in which nodes broadcast messages and tune broadcast rate adaptively according to network condition. Although the broadcast is controlled, these schemes may still create broadcast storms in dense networks.

The second approach to data dissemination is for intermittently connected wireless networks [2, 23]. These schemes leverage the store-carry-forward paradigm to improve data delivery. They essentially work for sparse networks but may not work for denser networks where the quality of dissemination is expected. For example, deadline or coverage of data delivery may not be guaranteed with these schemes.

The third topic-based (interest-based) data dissemination [3, 8, 23] where co-located mobile users only exchange information if they share mutual interest on some topic. However, these schemes are only for ad hoc networks and don't exploit the big picture of the entire network where network

Digital Object Identifier: 10.4108/ICST.MOBIQUITOUS2008.3870
http://dx.doi.org/10.4108/ICST.MOBIQUITOUS2008.3870

density changes according to distance to PoIs as shown in Schelling behavior. Understanding the big picture of the network and the grouping behavior of users is the key to design an efficient data dissemination scheme.

Finally, geocasting is also a related data dissemination technique, where the broadcast is determined by the physical location of mobile nodes [6, 7]. Whereas our scheme uses the similar interests to expedite data dissemination.

7.2 Mobility Models and Data Dissemination

Mobility is a crucial factor in designing data dissemination schemes for mobile P2P networks. Currently, almost all existing data dissemination schemes are evaluated by either Random Way Point [5, 11, 18, 20, 21] or Group-based [13] mobility model. However, these mobility models have noticeable drawbacks. Obviously, Random Way Point Mobility model is unrealistic, especially in macroscopic scale, due to its repetitive movement patterns. Meanwhile, Group-based mobility model requires nodes to move in groups all the times and never change to new groups or moves independently. Group-based mobility model can be used in disaster/recovery, mission-critical scenarios where team members are required to collaborate in pre-configured groups. However, for civilian scenarios and daily activities, a strong grouping assumption may not hold. A recent work shows that human movements are affected by the needs of humans to socialize or cooperate [13]. Thus, people who share mutual interest can group to exchange messages.

8. CONCLUSION

In this paper, we leverage the natural grouping and moving behaviors of humans presented in Schelling's model, to design a novel data dissemination scheme for hybrid mobile P2P networks. First, we introduce two important properties of this model: (1) density of similar individuals increases in close proximity of clusters, and (2) on the ways to their mutually targeted clusters, individuals can group into small "moving" clusters. Then, we demonstrate various scenarios in real world where the analogies of Schelling's model (i.e. Schelling behavior) and wireless technologies indeed co-exist such as shopping streets, campus life and social events. This co-existence creates a unique opportunity to expedite data dissemination in mobile P2P networks.

Therefore, we propose an efficient and lightweight data dissemination scheme, which exploits Schelling behavior to provide timely data accessibility, especially for nodes inside the *Message Reachability Zone* (MRZ). Relying on the first property of Schelling's model, we design a push model by allowing nodes inside the MRZ to collaboratively rebroadcast messages. In order to avoid broadcast storms, save node energy and reduce transmission collisions, we assign each message m a broadcast timer and apply overhearing mechanism to re-estimate this timer. To exploit the second property, we apply a pull model, in which nodes outside the MRZ broadcast queries to their similar neighbors and queries/responses are automatically limited within the similar nodes, with no group management protocol requirement. To further improve data accessibility, leaving and coming nodes collaborate to answer queries. Our simulation results show that the proposed data dissemination scheme improves data accessibility significantly while minimizing network overhead.

In summary, we believe wireless technology and Schelling behavior co-exist in many scenarios and offer an ideal opportunity to expedite data dissemination in mobile P2P networks. Toward this end, our proposed data dissemination scheme is novel and widely applicable.

9. REFERENCES

[1] L. Breslau, P. Cao, L. Fan, G. Phillips, and S. Shenker. Web caching and zipf-like distributions: Evidence and implications. In *INFOCOM*, 1999.

[2] M. Grossglauser and D. N. C. Tse. Mobility increases the capacity of ad-hoc wireless networks. In *INFOCOM*, April 2001.

[3] A. Heinemann, J. Kangasharju, F. Lyardet, and M. Mühlhäuser. iClouds - peer-to-peer information sharing in mobile environments. *European Conference on Parallel Processing*, 2003.

[4] X. Hong, M. Gerla, G. Pei, and C.-C. Chiang. A group mobility model for ad hoc wireless networks. In *ACM International Workshop on Modeling and Simulation of Wireless and Mobile Systems*, 1999.

[5] A. Khelil, C. Becker, J. Tian, and K. Rothermel. An epidemic model for information diffusion in manets. In *MSWiM*, 2002.

[6] Y.-B. Ko and N. Vaidya. Geotora: a protocol for geocasting in mobile ad hoc networks. In *ICNP*, 2000.

[7] Y.-B. Ko and N. Vaidya. Geocast enhancements of aodv for vehicular networks. *ACM SIGMOBILE Mobile Computing and Communications Review*, 2002.

[8] G. Kortuem, J. Schneider, and D. Preuitt. When peer-to-peer comes faces-to-faces: Collaborative peer-to-peer computing in mobile ad hoc networks. In *International Conf. on P2P Computing*, 2001.

[9] H. Lim, L.-C. Kung, J. C. Hou, and H. Luo. Zero-configuration, robust indoor localization: Theory and experimentation. In *INFOCOM*, 2006.

[10] W. Lou and J. Wu. Double-covered broadcast (dcb): a simple reliable broadcast algorithm in manets. In *INFOCOM*, 2004.

[11] L. McNamara, C. Mascolo, and L. Capra. Content source selection in bluetooth networks. In *Mobiquitous*, 2007.

[12] M. Motani, V. Srinivasan, and P. S. Nuggehalli. Peoplenet: Engineering a wireless virtual social network. In *MOBICOM*, 2005.

[13] M. Musolesi and C. Mascolo. Designing mobility models based on social networks theory. *Mobile Computing and Communications Review*, 11, 2007.

[14] S.-Y. Ni, Y.-C. Tseng, Y.-S. Chen, and J.-P. Sheu. The broadcast storm problem in a mobile ad hoc network. In *MOBICOM*, 1999.

[15] M. Papadopouli and H. Schulzrinne. Effects of power conservation, wireless coverage anc cooperation on data dissemination among mobile devices. In *MOBICOM*, 2001.

[16] T. Repantis and V. Kalogeraki. Data dissemination in mobile peer-to-peer networks. In *MDM*, 2005.

[17] T. C. Schelling. Models of segregation. *American Economic Review*, 59, 1969.

[18] L. Vu and K. Nahrstedt. Adaptive mobility-assisted data dissemination in mobile disaster/recovery environments. In *MILCOM*, 2007.

[19] K. H. Wang and B. Li. Group mobility and partition prediction in ad hoc wireless networks. In *IEEE International Conference on Communications*, 2002.

[20] O. Wolfson, B. Xu, and R. M. Tanner. Mobile peer-to-peer data dissemination with resource constraints. In *MDM*, 2007.

[21] O. Wolfson, B. Xu, H. Yin, and H. Cao. Search-and-discover in mobile p2p network databases. In *ICDCS*, 2006.

[22] X. Yang and A. Bouguettaya. Using a hybrid method for accessing broadcast data. In *MDM*, 2005.

[23] W. H. Yuen, R. D. Yates, and S.-C. Mau. Exploiting data diversity and multiuser diversity in noncooperative mobile infostation networks. In *INFOCOM*, 2003.

Digital Object Identifier: 10.4108/ICST.MOBIQUITOUS2008.3870
http://dx.doi.org/10.4108/ICST.MOBIQUITOUS2008.3870