Floating Content: Information Sharing in Urban Areas

Jörg Ott, Esa Hyytiä, Pasi Lassila
Aalto University, Comnet

Tobias Vaegs
RWTH Aachen University, ComSys

Jussi Kangasharju
University of Helsinki

***Abstract*—Content sharing using personal web pages, blogs, or online social networks is a common means for people to maintain contact with their friends, colleagues, and acquaintances. While such means are essential to overcome distances, using infrastructure services for location-based services may not be desirable. In this paper, we analyze a fully distributed variant of an ephemeral content sharing service, solely dependent on the mobile devices in the vicinity using principles of opportunistic networking.**

The net result is a *best effort* service for *floating content* in which: 1) information dissemination is geographically limited; 2) the lifetime and spreading of information depends on interested nodes being available; 3) content can only be created and distributed locally; and 4) content can only be added, but not explicitly deleted. First we present our system design and summarize its analytical modeling. Then we perform extensive evaluation for a map-based mobility model in downtown Helsinki to assess the operational range for floating content, which, at the same time also validate the analytical results obtained for a more abstract model of the system.

I. Introduction

Online social network applications for sharing content, opinions, and experiences (Facebook, Flickr, YouTube, Twitter, etc.) are widespread means for people to foster their relationships irrespective of physical distance. Increasing mobile Internet use has made sharing experiences from mobile devices popular, as the authors witness daily from their acquaintances, e.g., on Facebook. Context- and location-aware services, such as digital graffiti [1], and (to some extent) Google Maps and Google Earth, have been constructed around mobile users, however, relying on services in the network infrastructure: to store and maintain data as well as to determine geographic proximity. While network-based social applications are essential to overcome distances and connect people around the world, relying on infrastructure services for location-aware services may often not be desirable:

- *Location privacy* concerns arise since the location typically needs to be communicated at least to some level of accuracy to obtain the right context information.
- *Content privacy* issues occur since shared information will be available at some "central" location and thus easily subject to censorship (a mixed blessing).
- *Connectivity* to the infrastructure becomes a prerequisite that is often limiting, especially for traveling users who may face high roaming charges, unavailability of data services, or simply no network coverage.
- *Geographic validity:* locally relevant content may be of little concern to the rest of the world; storing it in some well-accessible location may not be of much use.
- *Temporal validity:* Information may only be valid for a limited amount of time; yet, centrally stored content rarely is tagged using expiry information, leading to the content never being deleted—yet, quite frequently never being read either (WORN, write-once, read never).
- *User identification* of some kind is usually applied to limit the amount of data posted and creates some sense of responsibility towards the service provider.

In this paper, we propose and analyze a fully distributed variant of an ephemeral content sharing service, solely dependent on the mobile devices in the vicinity using principles of opportunistic networking, built upon our earlier reported work-in-progress [2]. Any user may create content and define its geographic *origin*, and *validity radius*—that together define the *anchor zone* in which content is meaningful—plus an *expiration time*.[1] The creator's device starts disseminating the content to its neighbors within the validity radius, as do other nodes, reducing the replication and increasing the deletion probability as a function of the Euclidean distance from the origin. As long as there are enough supportive mobile devices around in the anchor zone to replicate and store a piece of content, it *floats*. Other mobile nodes interested in this piece of content—because of its location or because its metadata matches what they are looking for—will be able to obtain a copy when they get "in range", i.e., enter its anchor zone, and have a copy disseminated to them.[2] When the node density becomes too low (even temporarily), the content will disappear (it *sinks*).

The net result is a *best effort* service for *floating content* in which: 1) information dissemination is geographically limited 2) the lifetime and spreading of information depends on interested nodes being available in the anchor zone of a particular content item; 3) traffic can only be created and caused locally, thus limiting DoS effects; and 4) content can

[1]Moreover, content may be tagged with metadata (e.g., content channels [3]) to support filtering and selective forwarding and replication.

[2]They could also obtain one by means of one-hop queries to their neighbors, but we defer this to future study.

only be added, but not deleted, so that security issues are kept outside the system.[3] We expect that such as simple mechanism will allow creating more sophisticated services on top.

II. Related Work

Concepts similar to our floating content have been introduced in research already a few years ago, such as in [1], [3], [5]–[13]. Some, e.g., [3], [7], [11], focus on distributing content over an ad-hoc or DTN-like network, using the wireless network only as a kind of a cache for Internet content. They do not consider the case of managing content *purely in the wireless domain*. Others, e.g., [8], [9], consider a purely wireless case, but focus on directed communications between users (unicast or multicast). In contrast, our work is not about directed communications, but about content floating in a particular location, i.e., the applications are different.

In [5], [10] the authors present models for hovering information, which both bear similarities with our floating content model. However, our work differs from theirs in an important way. Their focus is on defining the general model, whereas our focus is on *evaluating the general feasibility* of floating content systems. By feasibility, we understand aspects such as node density, required communication range, etc. Our work explores the parameter space of floating content systems to identify under which conditions such systems are feasible.

Work in a recent publication [12] presents a system concept very similar to our work in [2]. Their focus is mainly on implementation aspects and protocols and hence complementary to the work in this paper with a focus on evaluating the operating conditions for floating content systems. Moreover, in [14], a similar scheme has been considered in the context of VANETs. The objective has been to study the storage capability of such a system in terms of mean time to information loss. Several different scenarios have been considered by means of simulation experiments ranging from a one-way traffic setting to a city-wide scenario based on real vehicle traffic traces in San Francisco (SF Cab).

Another very recent paper [13] presents a system called Locus. Conceptually, the work only re-iterates already published concepts on floating content [2], [5], [10], [12], [15], [16]. Locus allows remote nodes to query for information placed in a given anchor zone, which makes it different from related work in this area. However, we believe this property is actually a severe weakness, since it allows for remote nodes to perform denial-of-service attacks anywhere in the network. In pure floating content work [2], [5], [10], [12], [15], [16] such attacks require physical presence in the targeted anchor zone, making them far less vulnerable, if not even impervious, to these kinds of attacks. Our work pre-dates that of Locus and already our published paper [2] contains a more extensive feasibility study. This paper goes far beyond that, performing an extensive evaluation of floating content in diverse urban settings and taking more key parameters into account.

The floating content concept is an example of delay tolerant networks. The fundamental properties of routing in such networks have been recently analyzed, see [17] and the references therein. However, the floating content service requires modeling the notion of information availability in the anchor zone. We have developed analytical models for this in [16]. In this paper, we validate the result for the so-called non-spatial model in [16] in a more complex mobility scenario.

III. Applications for Floating Content

The properties of floating content described above show interesting opportunities but also exhibit notable challenges. The major opportunity is to enable infrastructure-less localized information sharing without central oversight or data collection and without remote access, thus offering some degree of privacy: users have to be present to see something, a concept intuitively familiar from daily life. The major challenge is the best-effort operation: users cannot be guaranteed that information will stay around until its intended lifetime expires. Especially, content is expected to disappear from public places over night. Again, however, users can make intuitive judgments (to some extent assisted by technology as we will briefly discuss later): if a user is on a crowded market square, a busy street, or in a lively bar, chances are that content will float for a while, even if not all other people have floating content-enabled mobile devices. We also note that best-effort communication may be better than no communication at all.

Obviously, predicting that enough people will be around in an area is somewhat limited to the immediate future. Given density fluctuations over the day due to people's activities, we would expect floating content to remain likely available for no more than a few hours. Even availability for just one hour would be sufficient for various applications. Moreover, users may stay around or return and could re-issue updated content or serve as a permanent seed.

One class of applications for infrastructure-less best-effort local data availability in the order of one to a few hours is advertising and brokering for goods. While this also includes products and services in the legally gray area, announcing goods available on flea markets or bazaars is one example. As the composition of offerers and the available merchandise change dynamically, an instant catalog of (an area of) such a market can be composed and updated on

[3]We expect that the validity of local information be checked at low cost, e.g., in the real world by validating that there is really free beer in a particular bar and posting annotations in case a "story" is not true. In addition, within groups content postings may be signed leveraging the users' identities in other domains (e.g., of SIM cards [4]).

the fly. The information is by its very nature ephemeral and mostly of local relevance; moreover, there is no strict reliability requirement and content may be re-seeded. A related example is last-minute matching of people offering and seeking tickets for theater events as usually occur in flocks in front of the entrance.

Another class includes localized infrastructure-less content sharing among users. One example is bypassing censorship when spreading news (and keeping information localized and time-bounded as a protection against prosecution). Another one is information sharing among tourists and locals about attractions, instant comments on good offers (which may be most relevant within the immediate future) for which best-effort operation is perfectly fine. We also expect floating content to be useful for leaving notes to people when moving from one place to the next, e.g., to inform latecomers of the next meeting point. Note that a special case of this is instant content sharing in a co-located group (photos, music, videos); here, no best-effort issue arises as the content origins stay around.

Overall, the floating content concept can be carried further when rethinking elementary communication paradigms: on the one hand, floating content is (geo- and) information-centric by design, thus suggesting further considerations towards data-oriented or publish-subscribe styles of networking. On the other hand, the concept is inherently best effort; in contrast to this notion in the Internet, where close-to-instant repair mechanisms can recover lost packets, data that *sunk* is irrecoverable in an area (unless the originator returns). Taking these into account for future application and protocol design will likely yield more interesting uses; this is a topic for future study.

IV. Floating Content Design

We assume that all users are mobile nodes and that there is no supporting infrastructure for the system.[4] They use mobile phones or similar devices to communicate, so that ample storage capacity is available of which a sufficient fraction is set aside for floating content. The devices use wireless interfaces (e.g., Bluetooth or WLAN) for ad-hoc communication. Real-world traces have shown that median contact durations between mobile users, e.g., range from 15 s or less as in RollerNet [18] to some 120 s in conference environments [19].[5] Experiments using Bluetooth for inter-device communication yielded some 10–15 s channel setup delay and a net data rate of some 50 KB/s [19]. Thus, exchanging data up to several megabytes per contact appears feasible so that floating modestly sized pieces of contents could be supported (from text messages to photos); it also suggests that the communication capacity during a contact is likely to become the system bottleneck and mobile nodes may not need to reserve a lot of storage capacity for floating content.

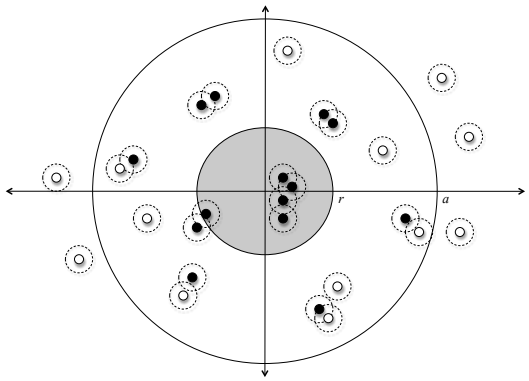

Figure 1. An anchor zone of an item, mobile nodes and their communication ranges: the content item gets replicated across and deleted from nodes as a function of the distance from the anchor point. The probability of a node carrying an item (black nodes) tends to 1 inside the anchor zone r and decreases until, after an availability threshold a, no more copies are found.

The devices also need to be location-aware, e.g., by using GPS, triangulation-based methods using WLAN access points or cellular base stations, or any other method offering reasonable accuracy; suitable location APIs are commonly available in modern smart phones. Since the floating content system is probabilistic, there are no strict requirements on the accuracy of positioning techniques; nodes are only required to agree on basic measurement parameters and the overall operation to determine the extent of anchor zones. Finally, nodes need roughly synchronized clocks to time out content items; both GPS and cellular networks may provide local time.

Mobile users originate ("post") content items with a defined anchor zone and TTL and must reside inside the anchor zone when doing so. Other users are interested in these items[6] and will accept copies to store and (probabilistically) further replicate within the anchor zone. We explicitly allow information items to disappear from the system and provide no guarantees about their availability. If, e.g., no (or too few) nodes are around to replicate a content item and the creator leaves the anchor zone, the corresponding items may disappear. Once the lifetime of an item expires, it will be deleted by all nodes.

A. System Operation

A node generates an information item I of size $s(I)$ with a certain lifetime (TTL) and assigns an anchor zone defined by its center P and two radii, r and a, as shown in figure 1: r defines the *replication range* within which nodes always try to replicate the item to other nodes they encounter; a defines the *availability range* within which the content item

[4]Content dissemination from fixed access points would lead us towards the PodNet model [3] paired with geographically limited distribution.

[5]See, e.g., the CRAWDAD archive for a collection of contact and mobility traces at http://crawdad.cs.dartmouth.edu/.

[6]The content can be organized into groups or channels similar to [3], but we restrict our considerations in this paper to a single channel.

is still kept around with limited probability, while outside a no copies of the item are to be found.

If two nodes A and B meet in the anchor zone of an item I, and A has I while B does not, then A replicates item I to B. Since replication is based purely on the location of nodes, in a simple case, every node in the anchor zone should have a copy of the item. Nodes leaving the anchor zone are free to delete their copy of the item.

In practice, the replication and deletion works as follows. Consider a node A having an item I, with an anchor zone defined by center point P and radii r and a. Let h denote the distance of node A from point P. When node A meets another node B, A will replicate item I to B with probability $p_r(h)$:

$$p_r(h) = \begin{cases} 1 & \text{if } h \leq r \\ R(h) & \text{if } r < h \leq a \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

where $R(h) \in [0, 1]$ is some (decreasing) function that gives the probability of replication outside the replication range but within the availability range. We define the deletion probability $p_d(h)$ in a similar way, $D(h) \in [0, 1]$:

$$p_d(h) = \begin{cases} 0 & \text{if } h \leq r \\ D(h) & \text{if } r < h \leq a \\ 1 & \text{otherwise} \end{cases} \quad (2)$$

The deletion function essentially serves the purpose of early prioritization to prevent buffers from filling up; it is evaluated upon each encounter with another node. If there is still a need to free buffer space the oldest messages are discarded.

These definitions yield the area outside the replication but within the availability range of the anchor zone as a buffer zone that offers smooth degradation of availability over distance. This definition of an anchor zone with two ranges is beneficial, because it provides protection against items disappearing when nodes move outside the replication range of the anchor zone for a brief moment and then return. Beyond the availability distance a (see figure 1) copies are deleted (immediately or upon encountering the next node, depending on the implementation).

In our evaluation presented below, we use $R(h) = 0$ and $D(h) = 0$, i.e., neither replication nor active deletion take place for $r < h \leq a$, rendering the buffer zone entirely passive. Note that if $r = a$, $R(h)$ and $D(h)$ become immaterial.

B. Floating Content Protocol

A floating content message m is identified by a unique message id Id_m and carries its anchor zone spec, i.e., anchor point (P_m, r_m, a_m); and its lifetime T_m in its headers; the content item I_m of size $s(I_m)$ is in the message body. We use a 4-phase protocol for exchanging floating content messages:

1) Nodes continuously send neighbor discovery beacons to discover peers.

2) After discovering a peer, i.e. receiving a beacon, a node sends a summary message that includes a vector of the content items available for replication, i.e., all items for which $p_r(h) > 0$ at the node's present location. Per item, the vector contains Id_m, s(I_m), (P_m, r_m, a_m), and T_m. The summary may be limited to what fits into one MTU size packet (or otherwise to a message sufficiently small to be exchanged efficiently so that most of the per-contact communication capacity is left to exchange content); if a node has more content to share, the list of content is spread across multiple summary messages in a round-robin fashion.

3) As soon as a device is aware of what a neighbor has to offer, it requests a subset of the content items: the receiver requests all those items for which evaluating $p_r(h)$ so suggests. It may prioritize the order in which elements are replicated according to a replication policy. The replication policy determines in which order messages are replicated when two nodes meet. We define five such policies: FIFO preserves the order in which messages were created/received at a node, RND randomizes the order, and three further algorithms use ascending order by anchor zone ($A_m = a_m$)[7], floating volume ($V_m = A_m \times s(I_m)$), and the product of volume and TTL ($V_m \times T_m$); we abbreviate these algorithms as *Smallest Area First (SAF)*, *Smallest Volume First (SVF)*, and *Smallest Total resource consumption First (STF)*. As mentioned above, $R(h) = 0$ so that replication only takes place within r_m.

4) Messages containing the requested items are then exchanged until the nodes lose contact or the batch is completed. In the former case, the receiver discards all incomplete messages; in the latter, the protocol returns to phase 2).

All phases are fully bidirectional so that message exchanges take place in both ways simultaneously. The protocol does not restrict message exchange to two nodes at a time (even though some link layer technologies may). Beaconing is continued throughout the message exchange process so that new nodes may be discovered while a node is already exchanging messages with another. The protocol phases and the incremental summary messages allow nodes to append messages they just received to the list of those offered to other nodes while already exchanging messages.

Nodes delete messages following one of two deletion policies when $p_d(h)$ suggests deleting a message m. Recall that, with $D(h) = 0$, no messages are deleted within a. Outside a, *immediate* deletion leads to m being discarded once a node leaves a, whereas *upon-encounter* deletion performs this action only when the next node is met, allowing a node

[7]At this point, we consider just the anchor zone radius and do not use the anchor zone area $A_m = a_m^2 \times \pi$.

to wander out of a and back in while keeping m. From an implementation perspective, the *upon-encounter* deletion policy is more sensible as it is triggered by an external event. Deletion takes place before messages are offered to other nodes.

C. Operational Considerations

We allow the user generating a content item to define the extent of the anchor zone. We only require the user to be in the anchor zone at the time of creation, but do not impose any limits on the anchor zone size. This may naturally cause problems because there is no incentive for users to limit the anchor zone. It would thus be easy to spam the system by inserting items with "infinite" anchor zones, limited in practice only by numeric representation of the radius in the protocol messages. But flooding items throughout the network is problematic, since it is known that flooding quickly exhausts the system resources (both channel and buffer capacity).

Without infrastructure and with potentially sparse and frequently disconnected node populations, accounting or reputation mechanisms like [20] do not appear feasible, e.g., due to their vulnerability against the Sybil attack [21] because of the lack of central authentication of identities. We therefore define a simple mechanism for resource management to discourage unlimited content distribution: at any given point, we prioritize items inversely with respect to a) their expected resource consumption and b) the distance from their anchor. We realize a) by means of the three resource-aware replication policies SAF, SVF, and STF that take into account the anchor zone size A_m, $A_m \times s(I_m)$, and $A_m \times s(I_m) \times T_m$, respectively, that we evaluate below. We realize b) by means of $p_r(h)$ and $p_d(h)$.

When a node needs to replicate or store more items than it is able to, it gives preference to items with the smallest anchor zones and evicts the items with the largest anchor zones. It is possible to post items with very large anchor zones, but their availability is likely to be quite low.

It is obviously possible for a spammer to move and create items with small anchor zones everywhere. We provide no mechanisms to guard against this. Instead, we consider the effort of having to move around to be a sufficient deterrent to most spammers. Moreover, making replication priority a function of the content lifetime (as in the STF policy) ensures that content will not stay around infinitely so that a spammer would have to revisit each place regularly.

Our approach has the advantage that it does not require security infrastructure or any degree of mutual trust. Of course, nodes could spam their immediate neighborhood or perform DoS attacks, but they could achieve the latter using physical layer interference anyway.

V. Analytical Model

The floating content service has been studied analytically in [16]. A fundamental objective of this work has been to derive the so-called *criticality condition* guaranteeing the availability of the information within the anchor zone with a high probability. This depends heavily on the mobility patterns of the users, especially inside the anchor zone. Next we present the so-called non-spatial model from [16].

A. Non-spatial Black-box Model

Consider the anchor zone just as a locale where the nodes enter and then spend some time and finally exit. Inside the locale, the nodes are exchanging information between each other. We assume that the nodes move randomly within the locale and the time spent there is relatively long so that the exact points of entrance and exit do not have any bearing. The whole population is assumed "well mixed" so that the proportion of information-carrying nodes is everywhere constant.

During the sojourn time, a node encounters randomly other nodes. Consider a system consisting solely of two nodes staying permanently in the locale and denote by ν the frequency at which they come in contact with each other, i.e., within each others' transmission ranges. Now, if the total population of nodes in the locale is N, then there are $\frac{1}{2}N(N-1) \approx \frac{1}{2}N^2$ pairs and the total rate of encounters is $\frac{1}{2}N^2\nu$. The fraction $2p(1-p)$ of these encounters are such that a node without the information gets it, and the total rate of such events is $p(1-p)N^2\nu$. This is the rate at which the size of the population of information-carrying nodes tends to increase. Let $1/\mu$ denote the node's mean sojourn time in the anchor zone. Then the total exit rate of nodes is $N\mu$, and in particular, the exit rate of information-carrying nodes is $Np\mu$, so that their net growth rate is

$$N\frac{d}{dt}p = N^2p(1-p)\nu - Np\mu. \tag{3}$$

In equilibrium, the two terms on the right hand side are equal leading to the stationary value $p^* = 1 - \mu/(\nu N)$. Existence of a positive solution $p^* > 0$ requires that,

$$N\nu/\mu > 1. \tag{4}$$

This is the *criticality condition*. Note that the quantity on the left hand side is the *average number of encounters a randomly chosen node experiences during its sojourn time*. By considering the sign of the derivative (3) one easily sees that the solution is stable. Below the criticality, the derivative is everywhere negative and the solution is driven to $p = 0$, i.e., the information cannot be sustained even in the black-box model.

B. Applicability of the Model

The black-box model for information exchange is a highly abstract model capturing only the essential elements. The crucial assumptions are 1) the fluid limit approximation and 2) a well-mixed mobility pattern. The fluid limit approximation means that the anchor zone is populated with a

large number of nodes. In this case, the criticality condition ensures that information remains available. In practice, the number of information-carrying nodes, Np, can be small and the information will be eventually lost due to stochastic fluctuations in the system. The well mixed mobility pattern assumption implies that all the nodes inside the anchor zone, with and without the information, are equally likely to encounter each other at any given point of time and independently of any past encounters.

In the black-box model, we have chosen to ignore the spatial aspects of the information exchange process. For example, in practice one can expect that the probability of a node carrying the information is smaller near the boundary than at the center of the anchor zone. We also note that that several key parameters are not explicitly present in (3). For example, the transmission range d affects directly the encounter rate ν, i.e., $\nu = \nu(d)$. Analysis of a more detailed model capturing the spatial aspects of the information exchange is challenging, and can only be done for some elementary models. For an example, see [16].

In this work, we apply the above black-box model, as it already captures the fundamental elements of the system, and characterizes the criticality condition, above which the information remains available in the anchor zone with a high probability. In particular, we estimate the three key parameters in (4) from the simulation experiments, and then compare how the criticality quantity $N\nu/\mu$ correlates with the observed average information life times. The similarity turns out to be very good, thus confirming the applicability of the theoretical model in these more complex mobility scenarios.

VI. Simulation Environment

While our initial evaluations [2], [16] focused on basic feasibility in a static and two idealized mobility settings (using random-waypoint and Manhattan mobility), for this paper, we choose the more sophisticated Helsinki City Scenario (HCS) [22] based upon a city map ($4500\,\mathrm{m} \times 3400\,\mathrm{m}$) of downtown Helsinki for evaluation (see figure 2). HCS models two types of nodes: most roam the city area following streets and walkways when moving to randomly chosen points on the map following a shortest path using pedestrian ($0.5 - 1.5\,$m/s) or car ($10 - 50\,$km/h) speeds; some follow a set of three predefined routes as trams with their own characteristic speed ($25 - 35\,$km/h).

We consider this kind of tourist-style mobility of restless users as we are only interested in their local interaction within different anchor zones and do not try to achieve city-wide routing functionality. As a consequence, nodes can be treated as "new" once they left an anchor zone and, therefore, daily routines and social relations would not add to the model (beyond possible density fluctuations over the day) so that we deliberately focus on simpler modeling.

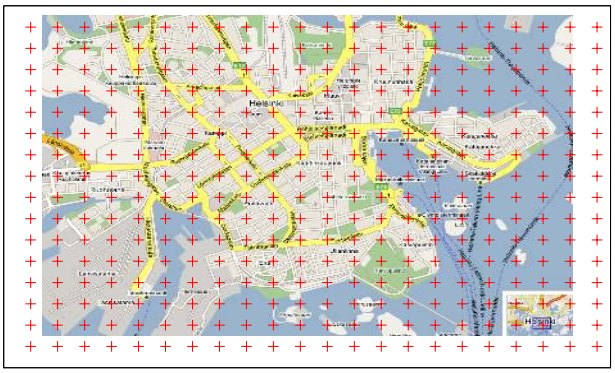

Figure 2. Simulation area and reference anchor locations every $200\,\mathrm{m}$.

We use the ONE simulator [22] for which we implemented: 1) a *FloatingApplication* class initiating content postings at different locations with configurable parameters (a, r, TTL, size); 2) a *FloatingRouter* that implements the five replication and two deletion policies described in section IV-B above; and 3) two classes *FloatingAppReporter* and *ContactConditionReport* that provide information about the node density and encounter frequencies as input parameters to our analytical model. The class *FloatingMessageReport* delivers statistics about creation, replication, and deletion of floating messages.

We create six different mobility scenarios as shown in table I: We use 2 Mbit/s data rate for the wireless links for 10 m (Bluetooth) and 50 m (WLAN ad-hoc mode) radio ranges and vary the number of nodes from 126 (small) to 252 (medium) to 504 (large scenario). Unless stated otherwise, we use 5 MB buffer space per node, except for trams that use always 50 MB.

Reference	radio range	#nodes	ped's	cars	trams
$S10(r,a)$	10 m	126	80	40	6
$S50(r,a)$	50 m	126	80	40	6
$M10(r,a)$	10 m	252	160	80	12
$M50(r,a)$	50 m	252	160	80	12
$L10(r,a)$	10 m	504	320	160	24
$L50(r,a)$	50 m	504	320	160	24

Table I
Overview of simulation scenarios

We conduct most of our simulations with two different anchor zone sizes: $a = r \in \{200\,\mathrm{m}, 500\,\mathrm{m}\}$. To assess the feasibility of floating content in downtown Helsinki, we choose fixed anchor locations every $200\,\mathrm{m}$ horizontally and vertically across the entire simulation area. The resulting 352 anchor points are shown as crosses in figure 2.

We run all simulations over a period of 24 hours, use a message TTL (T_{ttl}) of one hour and correspondingly a cooldown period of 1 hour. We report the mean values for a given scenario a) across all anchor zones in the entire simulation area, using an average weighted by the number of messages generated per anchor zone and b) restricted

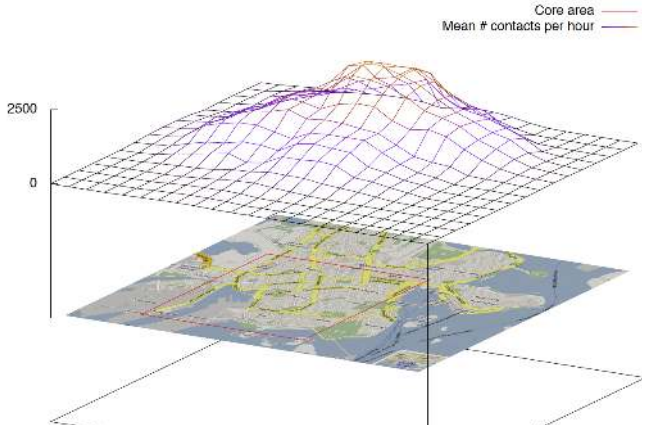

Figure 3. $S50(500, 500)$: Mean number of contacts per hour per anchor zone

to a *core area* of $1.8\times1.8\text{km}^2$ that includes diverse terrain (including water front) but where we do not expect any boundary effects due to simulation area limits.

Figure 3 shows the mean number of node contacts per anchor zone per hour for the scenario $S50(500, 500)$ (that we will also use in section VII-B). The core area is indicated as a square on the map. Expectedly, across all scenarios, the contact density increases with radio range, the number of nodes, and the anchor zone size from some 230 contacts/h for the densest spot in $S10(200, 200)$ to 17,500 contacts/h in $L50(500, 500)$.

VII. Evaluation

We carry out two series of simulations: an initial set of simulations uses minimal message sizes and sufficiently large buffers to establish under which of the above mobility models floating content is feasible and how it matches to the predictions of our analytical model (VII-A). We then evaluate the floating content concepts and different algorithms for more diverse parameter settings (VII-B).

A. Feasibility Considerations

We run simulations individually for each of the 352 anchor points with two different radii for the six mobility scenarios introduced above. Individual runs are used to isolate anchor zones from the impact of neighboring ones. In these scenarios, we use tiny messages so that message transfers are never aborted because nodes move out of range. Each node generates a message when it comes within $\frac{1}{2}r$ of the anchor point, but message generation per node is limited to one message per 30 min. This yields from 1350 messages for $Sx(200, 200)$ to 9350 messages for $Lx(500, 500)$ per 24 hours simulation time in the busiest spots. Next we consider the spatial distribution of the probability that content floats for up to one hour for all $S10$ scenarios. We additionally validate criticality condition (4) that defines a condition under which a message are will stay available with a high probability. To this end, the parameters for the mean number of customers N, the mean sojourn time $1/\mu$ and the mean contact rate ν for each anchor zone were also estimated from the simulation data (cf. section V).

Figure 4 depicts the value of the criticality factor $N\nu/\mu$ as well as the spatial probability that the content is available up to one hour in each anchor zone. The left column corresponds to the case with $r = a = 200\,\text{m}$, and the right column to $r = a = 500\,\text{m}$. In both columns, the left figure shows the observed floating probability (fraction of messages that is available at the end of $T_{\text{ttl}} = 1\text{h}$), and the right figure the corresponding criticality factor. From top to bottom, the number of nodes increases: 126 nodes, 252 nodes and 504 nodes.

When the criticality factor is less than 1, or close to one, the expectation is that the messages will not stay alive (i.e., loosely speaking, they will not float). As can be seen, the white/light blue areas in figure 4 match nicely (floating probability vs. criticality). Also, by looking in more detail at the case with 126 nodes and $r = a = 200\,\text{m}$, we observe that the small light green areas, reflecting a somewhat better chance of floating, nicely correlate with the small islands in the corresponding floating probability figure. Increasing the number of nodes to 504 raises the value of the criticality factor substantially above 1 (i.e., the criticality threshold) in some areas and it is reflected in a corresponding increase in the floating probability. Finally, when $r = a = 500\,\text{m}$ (right column) the criticality condition is well above the threshold almost everywhere except in the areas corresponding to the Baltic Sea, which also closely resembles the pattern of the corresponding very high floating probability estimates.

Using a larger radio range yields further improvement as shown in figures 5 and 6, especially for larger r and a. These latter two figures compare the mean floating time as a fraction of T_{ttl} for a piece of content across all scenarios considering the entire simulation area (except for places in the sea where there are no messages generated) and only the core area, respectively. The mean values are weighted by the number of messages (running from a few tens to some 17,000 per hour). We find many scenarios for which floating is basically feasible, i.e., content availability $p(t = T_{\text{ttl}}) >> 0$. Especially the large anchor zones show a trend towards slowly decaying $p(t)$. This is an indication that the system operation point is well above criticality which ensures that messages even in the simulated stochastic system will float with a high probability for a long time, say tens of hours. We also see that in those cases when messages do not float (but *sink*), the majority sinks during the first 10–20% of their lifetime. This is an important finding as a system might be able to warn a user still close by that her message is likely to disappear soon.

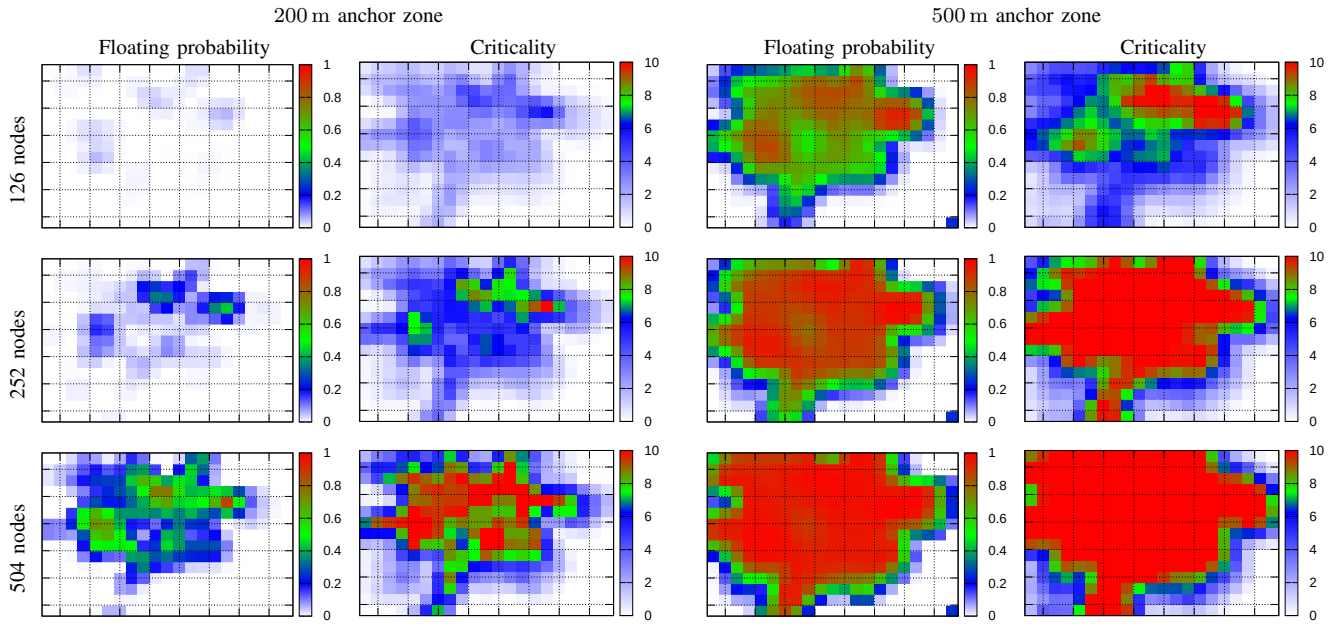

Figure 4. The observed floating probability and the corresponding criticality condition in the whole area per anchor location for $r = a = 200\,\mathrm{m}$ and $r = a = 500\,\mathrm{m}$ when using 10m range Bluetooth radio. From top to bottom, the number of nodes is 126 ($S10$), 252 ($M10$) and 504 ($L10$). Similarity between the estimate (criticality condition) and the observed floating probability is striking in each case.

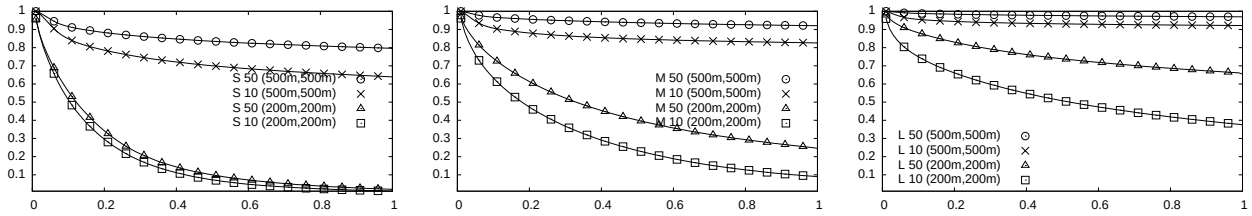

Figure 5. Floating duration as fraction of TTL: from left to right scenarios S, M, and L.

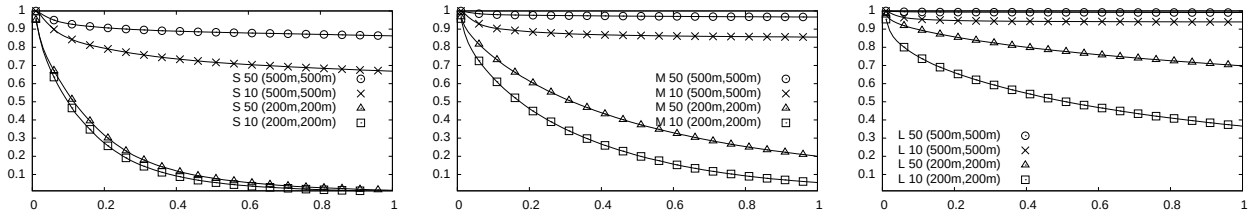

Figure 6. Floating duration for the $1.8 \times 1.8 km^2$ core area excluding borders and the Sea as fraction of TTL: from left to right scenarios S, M, and L.

B. Characterizing Floating Content Performance

We choose scenario $S50(r, a)$ (with $r = a = 500\,\mathrm{m}$ in most cases) for further evaluation as this provides on the one hand sufficient floating success rates for messages to realize best-effort content sharing, yet leaves room for improvement on the other.

One important element is understanding where limitations of floating performance come from. Given an encounter pattern, opportunistic communications can be limited by buffer size or by communication capacity when nodes encounter each other (determined by link rate and contact duration). We vary the buffer sizes available to the 120 mobile nodes and plot the floating success rate over the message size in figure 7a. We find that for messages smaller than 100 KB the system is limited by the encounter capacity, for large messages buffer space becomes an issue. We also find that 50 MB message buffer is sufficient for the system not to be buffer-limited.

In a separate set of simulations run (not shown), we chose message and buffer sizes and the link rate so that the entire buffers of two nodes could be exchanged even in the shortest possible contact (0.1s). We found that a buffer

capacity of 100 messages was sufficient to maximize floating performance.

The floating content algorithm provides for two distance parameters, the *replication* radius r and the *availability* radius a. While all simulations so far assumed $r = a$, we now evaluate the impact of $a > r$, i.e., the impact of a buffer zone $a - r$ in which messages are neither replicated nor deleted. In two cases, we keep r constant at $r \in \{200\,\mathrm{m}, 500\,\mathrm{m}\}$ as above and vary a in $[r, r + 500\,\mathrm{m}]$ in steps of $100\,\mathrm{m}$. In the third case, we keep $a = 500\,\mathrm{m}$ and vary r in $[100\,\mathrm{m}, 500\,\mathrm{m}]$. Figure 7b shows the resulting floating performance plotted over the buffer zone width, the difference $a - r$.

We find that the replication radius r is, for a given node density, of crucial importance for the floating performance. Looking at the first curve, the fraction of messages floating for their TTL degrades quickly as r falls below 400m (i.e., $a - r = 100$m). The second curve shows that, with a small replication range $r = 200$m, increasing the availability radius a improves data availability notably, but does not quickly reach a satisfactory level, only achieving some 33% at $a = 700$m (i.e., $a - r = 500$m). While the trend points upward, recall that increasing the availability range increases the resource consumption as data is spread more widely. The third curve shows that a sufficiently large replication range $r = 500$m does not require much of a buffer zone: the floating performance increases (up to some 7% in the figure when $a = 1000$m, i.e., $a - r = 500$m), but not significantly and with diminishing returns. With appropriate r, a small buffer size is useful, but not much more is needed, supporting the idea of limited geographic coverage and spatial re-use of buffer and communication capacity.

We also experimented (not shown) with two different deletion policies, immediate and upon-encounter, using the same choices of r and a as shown in figure 7b. We find that the floating performance shrinks by up to 0.3–0.4 when applying immediate deletion compared to upon-encounter deletion; the resulting curves for varying r and a are qualitatively similar to those in figure 7b. This shows the importance of keeping content even when outside the anchor zone and supports the idea of a buffer zone. This is in line with practical implementation considerations: a node would rather act upon an external trigger (such as encountering another node) than permanently scan its buffers for data items to expire.

The expected gain from a passive buffer zone with $R(h) = 0$ (and from upon-encounter deletion) depends, obviously, on the mobility pattern. That is, a buffer zone outside the replication radius is only useful if at least some of the nodes return to the replication zone directly from the buffer area. For pedestrians, this may happen quite often in practice, while perhaps less often for vehicles.

Finally, we simulate a mixed scenario based on $S50$ with 50 MB buffers: the nodes move as before and generate content items when they are inside the *core area* at mean rates of 1 (single), 2 (double), and 4 (quad load) messages per node per hour. Each generated message is anchored at the generating node's present location. Anchor zone sizes $r = a$ are chosen from $[500\,\mathrm{m}, 2000\,\mathrm{m}]$ in $100\,\mathrm{m}$ steps, message sizes from $[100\,\mathrm{KB}, 1000\,\mathrm{KB}]$ in $100\,\mathrm{KB}$ steps, and TTLs from $[30\,\mathrm{min}, 180\mathrm{min}]$ in $30\,\mathrm{min}$ steps, all uniformly distributed. We use upon-encounter deletion and simulate the five different replication policies introduced in section IV-B: FIFO, RND, SAF, SVF, and STF.

Figure 7c shows the overall floating performance across all content items. We observe that the single load already appears to saturate the floating capacity of the network and that increased load leads to performance degradation. We also note that with less saturation, the impact of the replication algorithm is less pronounced. We dissect the impact of the replication algorithm further in figure 8, plotting the floating performance as a function of message size, TTL, and anchor zone. Across all plots, we can see that the two simple replication algorithms, FIFO and RND, perform worse than those considering at least one property of the content items. The difference becomes more pronounced with increasing load and their respective prioritization become visible: we can influence that less resource consuming messages (smaller size or TTL) shall be preferred. The impact of conscious replication is most visible for the different anchor zone sizes. FIFO and RND do not support small anchor zones well. But also the algorithms aware of r and a do not manage to obtain similar performance across all anchor zone sizes because their respective radius (paired with the mobility model) determines the number of contacts available for replication.

Overall, SAF (taking into account only a) performs best across all anchor zones while not adding bias towards smaller or shorter-lived messages. Adding conscious preference for smaller or short-lived content pieces comes at the expense of messages with small anchor zones, simply because we presently use a non-weighted scalar product of the message parameters for prioritization: a large (or high TTL) message with a small anchor zone will thus be treated similarly to a small (or low TTL) message with a large anchor zone. Experimenting with different weights would be subject to future investigations; but before attempting to determine "optimal" weights or exploring more complex algorithms, we believe that broader studies on different mobility models and loads are necessary to avoid prematurely optimizing for specific cases.

VIII. Conclusion

We have presented the concept of *Floating Content*, a fully distributed variant of an ephemeral content sharing service, solely dependent on the mobile devices in the vicinity that uses principles of opportunistic networking. Our evaluation using the Helsinki City Scenario suggests that floating content in urban environments can be feasible even

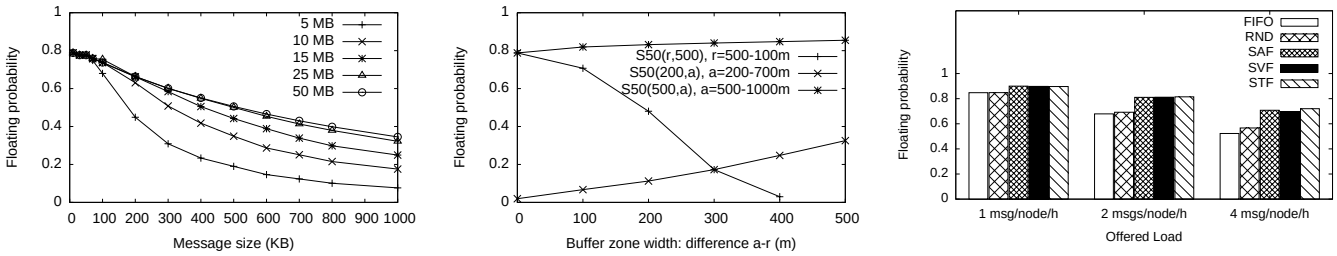

Figure 7. Floating performance in scenario $S50(500, 500)$ as a function of (a: left) the node buffer size for different message sizes, (b: middle) when varying a and r, and (c: right) for different offered loads and replication algorithms.

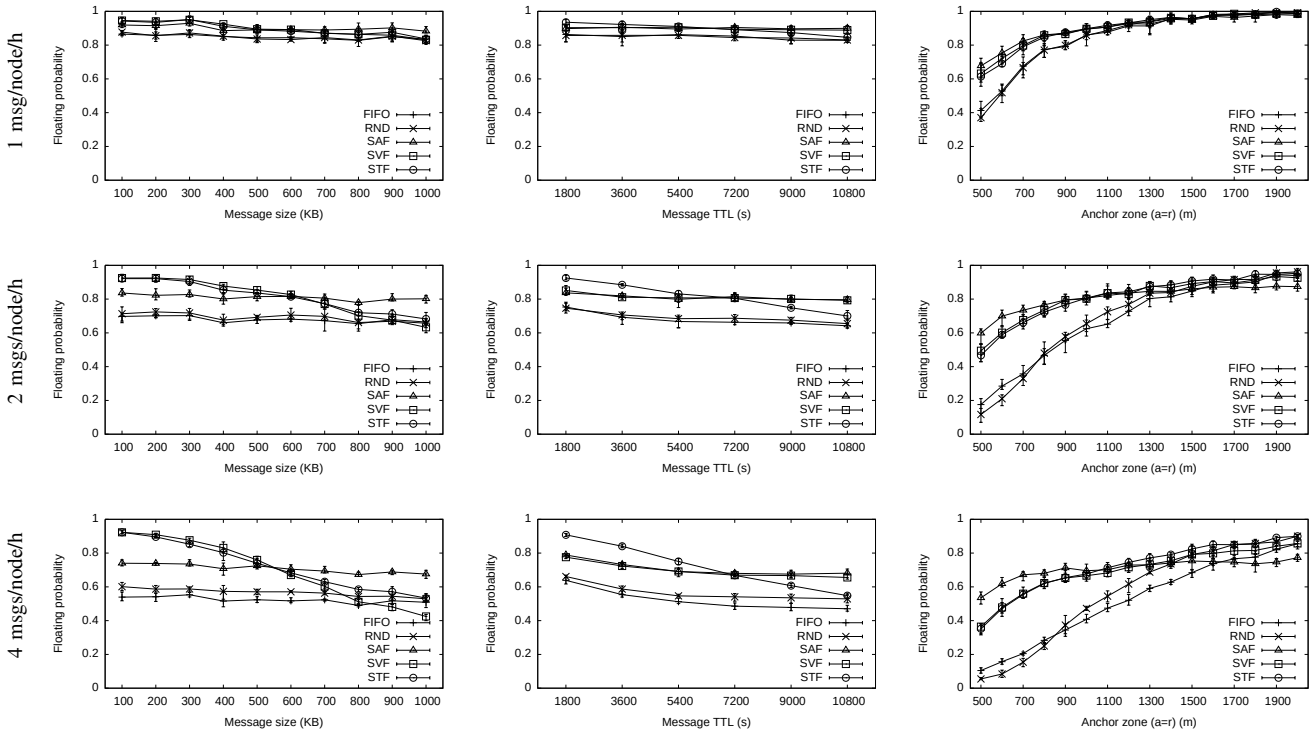

Figure 8. Floating performance as a function message size, TTL, and anchor zone for five different replication algorithms

with modest numbers of nodes supporting this application. Sufficiently large anchor zones are required for the node densities we investigated so that, while geographically limited sharing works, tightly constraining the availability may not be feasible unless the node density is very high. Indicatively, it seems that anchor zones should span multiple blocks in a city for sufficient replication, but a systematic exploration is for further study. We also find that—expectedly—floating content has limitations in border areas across all node densities when the movement of nodes into and out of an area is restricted, e.g., at the waterfront.

We could confirm the predictions of our analytical model for the probability that content would float in a given location. This is particularly important as the input parameters to our model can be estimated locally (possibly in cooperation with other nodes) so that a predictor could be calculated that can indicate the floating expectation to the user or an application. This is one direction of further study.

While we are aware of the limitations of synthetic mobility models (and of HCS), we believe that macroscopic mobility and social context and interaction patterns may be of lesser significance for localized content sharing—unless selective support based upon (closed) groups or content channels come in—so that we expect our findings to hold for other scenarios as well.

Our current and future work includes, on one hand, a validation of our analytical models and algorithms for a broader set of mobility models and mobility traces with diverse node densities and system loads. On the other hand, we have performed an implementation for the Nexus One Android

mobile phones, using which we target experimentation in lab scenarios and in various places of Helsinki to gain an initial understanding of floating content performance in different real environments. In the mid-term, it will be interesting to understand which more sophisticated applications could make use of such a best-effort service and how to evolve communication paradigms in support of those.

ACKNOWLEDGMENTS

The authors would like to thank Eamonn O'Neill and the anonymous reviewers for their suggestions.

This work was supported by TEKES as part of the Future Internet program of TIVIT (Finnish Strategic Centre for Science, Technology and Innovation in the field of ICT).

REFERENCES

[1] S. Carter, E. Churchill, L. Denoue, and J. Helfman, "Digital graffiti: public annotation of multimedia content," in *Conference on Human Factors in Computing Systems*, 2004, pp. 1207–1210.

[2] J. Kangasharju, J. Ott, and O. Karkilahti, "Floating Content: Information Availability in Urban Environments," in *Proc. of IEEE Percom 2010, Work in Progress session*, March 2010.

[3] V. Lenders, M. May, G. Karlsson, and C. Wacha, "Wireless ad hoc podcasting," *ACM/SIGMOBILE Mobile Comp. and Comm. Rev.*, 2008.

[4] N. Asokan, K. Kostiainen, P. Ginzboorg, J. Ott, and C. Luo, "Applicability of identity-based cryptography for disruption-tolerant networking," in *Proc. 1st Intl. ACM MobiSys Workshop MobiOpp*, 2007.

[5] A. Villalba Castro, G. Di Marzo Serugendo, and D. Konstantas, "Hovering information: Self-organizing information that finds its own storage," School of Computer Science and Information Systems, Birkbeck College, London, UK, Tech. Rep. BBKCS707, Nov. 2007.

[6] D. Corbet and D. Cutting, "Ad loc: Location-based infrastructure-free annotation," *ICMU 2006*, 2006.

[7] I. Leontiadis and C. Mascolo, "Opportunistic Spatio-Temporal Dissemination System for Vehicular Networks," in *Proc. 1st Int. ACM MobiSys Workshop MobiOpp*, 2007.

[8] W. Gao, Q. Li, B. Zhao, and G. Cao, "Multicasting in delay tolerant networks: A social network perspective," in *Proceedings of the tenth ACM international symposium on Mobile ad hoc networking and computing*. ACM New York, NY, USA, 2009, pp. 299–308.

[9] P. Hui, J. Crowcroft, and E. Yoneki, "Bubble rap: Social-based forwarding in delay tolerant networks," in *Proc. ACM MobiHoc*. ACM, 2008, pp. 241–250.

[10] A. Wegener, H. Hellbruck, S. Fischer, C. Schmidt, and S. Fekete, "AutoCast: An adaptive data dissemination protocol for traffic information systems," in *2007 IEEE 66th Vehicular Technology Conference, 2007. VTC-2007 Fall*, 2007, pp. 1947–1951.

[11] G. Karlsson, V. Lenders, and M. May, "Delay-tolerant broadcasting," *IEEE Transactions on Broadcasting*, 2007.

[12] O. R. Helgason, E. A. Yavuz, S. T. Kouyoumdjieva, L. Pajevic, and G. Karlsson, "A mobile peer-to-peer system for opportunistic content-centric networking," in *ACM SIGCOMM MobiHeld Workshop*, 2010.

[13] N. Thompson, R. Crepaldi, and R. Kravets, "Locus: A location-based data overlay for disruption-tolerant networks," in *Workshop on Challenged Networks*, Chicago, IL, Sep. 2010.

[14] B. Liu, B. Khorashadi, D. Ghosal, C.-N. Chuah, and M. Zhang, "Assessing the VANET's local information storage capability under different traffic mobility," in *Proc. of IEEE INFOCOM*, 2010.

[15] O. Karkulahti, "Distributed location-aware hovering information systems," Master's thesis, University of Helsinki, Oct. 2009.

[16] E. Hyytiä, J. Virtamo, P. Lassila, J. Kangasharju, and J. Ott, "When does content float? characterizing availability of anchored information in opportunistic content sharing," in *IEEE INFOCOM*, Shanghai, China, Apr. 2011.

[17] P. Jacquet, B. Mans, and G. Rodolakis, "Information propagation speed in mobile and delay tolerant networks," *CoRR*, vol. abs/0903.1157, 2009.

[18] P. U. Tournoux, J. Leguay, F. Benbadis, V. Conan, M. D. de Amorim, and J. Whitbeck, "The Accordion Phenomenon: Analysis, Characterization, and Impact on DTN Routing," in *Proc. IEEE INFOCOM*, April 2009.

[19] A. K. Pietiläinen, "Experimenting with Opportunistic Networking," in *Proc. of the ACM MobiArch Workshop*, June 2009.

[20] S. Buchegger and J.-Y. Le Boudec, "A robust reputation system for p2p and mobile ad-hoc networks," in *Proceedings of Workshop on Economics of Peer-to-Peer Systems*, Cambridge, MA, Jun. 2004.

[21] J. R. Douceur, "The Sybil attack," in *Proceedings of International Workshop on Peer-to-Peer Systems*, Cambridge, MA, Mar. 2002.

[22] A. Keränen, J. Ott, and T. Kärkkäinen, "The ONE Simulator for DTN Protocol Evaluation," in *SIMUTools '09: Proceedings of the 2nd International Conference on Simulation Tools and Techniques*. New York, NY, USA: ICST, 2009.