

Floating Content: Information Availability in Urban Environments

Jussi Kangasharju
University of Helsinki

Jörg Ott
Aalto University, Comnet

Ossi Karkulahti
University of Helsinki

Abstract—Content sharing via the Internet is a widespread means for people to foster their relationships irrespective of physical distance. 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. 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.

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) traffic can only be created and caused locally; and 4) content can only be added, but not deleted. We present the general model for floating content and results from our on-going evaluation. Our results show that floating content systems are feasible in scenarios which correspond to typical urban settings, in terms of number of nodes and their density, node mobility, and device capabilities.

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. Any user may create content and define its geographic *origin*, *validity radius*, and (optional) *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. Other mobile nodes interested in some type of content will be able to obtain a copy when they get “in range” of a particular piece of content and have either a copy disseminated to them or obtain one by means of a one-hop query.

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 target zone of a particular content item; 3) traffic can only be created and caused locally, thus limiting DoS effects; and 4) content can only be added, but not deleted, so that security issues are kept outside the system.² We expect that such a simple mechanism will allow creating more sophisticated services on top.

This paper is structured as follows. We review related work in section II. We formally describe the floating content model and its tuning parameters in sections III and IV where we also provide some basic analysis. We describe our evaluation of the model in section V and present the results for static and dynamic cases in section VI. We conclude our paper with a

¹Furthermore, content may be tagged with metadata according to some scheme (e.g., content channels [2]) to support filtering and selective forwarding and replication.

²We 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.

discussion of various issues and next steps in section VII, and conclude in section VIII.

II. RELATED WORK

Concepts similar to our floating content have been introduced in research already a few years ago, such as in [1]–[9]. Some, e.g., [2], [5], [9], 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., [6], [7], consider a purely wireless case, but focus on directed communications between users (unicast or multicast). Our work differs from these in that our work is not about directed communications, but about content floating in a particular location, i.e., the applications are different.

In [3], [8] 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, speed of movement, etc. Our work explores the parameter space of floating content systems to identify under which conditions such systems are feasible to deploy in the real world.

III. FLOATING CONTENT MODEL

We assume that all users are mobile nodes and that there is no supporting infrastructure for the system.³ The users are interested in information items “posted” by other users. They use mobile phones or similar devices to communicate, so that ample storage capacity is available of which, say, 1 GB is used for floating content. The devices have wireless interfaces (e.g., Bluetooth or WLAN) for ad-hoc communication within a certain range. The nodes cooperate by replicating content among interested parties as we describe below.

The content can be organized into groups or channels similar to [2], but for this initial study, we restrict our considerations to a single channel. We also assume that content items are rather small so that encounters between mobile nodes are sufficient for successful content exchange. Contact durations vary widely depending on node mobility.⁴ Sample real-world traces have shown that median contact durations between mobile users, e.g., range from 15 s or less as in RollerNet [11] to some 120 s in conference environments [12] and experiments using Bluetooth for inter-device communication yielded some 10–15 s channel setup delay and then a net data rate of some 50 KB/s [12]. Overall, this allows exchanging some 50 KB–5 MB per contact, clearly hinting at the communication capacity between mobile nodes as the system bottleneck and showing that even our simplified assumptions support reasonable contents, from text messages to JPEG photos to MP3 files. There is nothing inherent in our system design

³Content dissemination from fixed access points would lead us towards the PodNet model [2] paired with geographically limited distribution.

⁴See, e.g., the CRAWAD archive for a variety of contact traces [10].

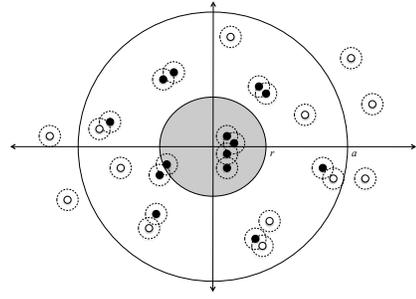


Fig. 1. An anchor zone of an item, mobile nodes and their communication ranges: the content item gets replicated across and deleted from nodes as a function of the distance from the anchor point. The probability of a node carrying an item (black nodes) tends to 1 inside the anchor zone r and decreases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

that would preclude dealing with larger information items if contact times are long enough. Investigations on information availability as a function of its size is left for further study.

Each item has an anchor zone, which is a real world area in which the items should be made available. We assume circular anchor zones defined by a center point and a radius.⁵ Figure 1 shows an example of an anchor zone and nodes.

As noted above, interested nodes keep copies of information items floating around in the anchor zone by probabilistically replicating the items when they meet. 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 an information item and the creator leaves the anchor zone, the corresponding information items will disappear (over time). Content items may be tagged with a lifetime and are discarded thereafter.

Anchor zones require nodes to be able to determine their position, e.g., by using GPS receivers or triangulation-based methods using WLAN access points, cellular base stations, or any other method offering reasonable accuracy. Since the 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.

IV. SYSTEM OPERATION

A node generates an information item I of size $s(I)$ and assigns an anchor zone (defined by its center and its radius) as well as a time to live $T(x)$ for this item. We require that the generating node be within the anchor zone at the moment of item creation. If two nodes A and B meet in the anchor zone of an item I , and A has I and B does not have it, 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

⁵Any other shapes are also possible, provided that they can be expressed in a relatively succinct manner since they need to be communicated along with the information items.

defined by center point c and radius r . Let d denote the distance of node A from point c . When node A meets another node B , A will replicate item I to B with probability p , where p is given by:

$$p = \begin{cases} 1 & \text{if } d \leq r \\ R(d) & \text{otherwise} \end{cases} \quad (1)$$

where $R(d) \in [0, 1]$ is some (decreasing) function that gives the probability of replication outside the anchor zone.

For deletion of items, we can define a similar function $D(d)$ which is 0 if the node A is in the anchor zone and some (increasing) function when A leaves the anchor zone. Allowing the item to survive outside the anchor zone is beneficial, because it provides additional protection against items disappearing because of nodes moving outside the anchor zone for a brief moment and then returning. We define an availability distance a (see figure 1) beyond which copies are deleted. The deletion function essentially serves the purpose of prioritization when the buffer gets full. It is evaluated upon each encounter with another node or whenever there is a need to discard content items to free buffer space. In our evaluation presented below, we use a step function: $D(d) = 0$ for $d \leq r$ and $D(d) = 1$ for $d > r$.

We allow the user generating the item to define the extent of the anchor zone. The anchor zone is characterized by its center and radius and it is sufficient for the user to be in the anchor zone at the time of creation. We do not impose any limits on how the user defines the anchor zone. 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.

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) the area of their anchor zone and b) the distance from their anchor. 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 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.

V. EVALUATION

We performed our initial evaluation with the ONE simulator [13]. We modify the simulator to account for the floating content distribution model: we create a special routing module based upon the *ActiveRouter* class and provide dedicated message generation and additional logging functions.

We have a world of 2000 by 2000 m in size where nodes are placed initially at random positions. At the center of the world is an anchor zone of a given radius. One node is always placed at the center of the anchor zone and at the beginning of the simulation, this node creates one information item, which then

gets spread according to the model in Section III. We assume that all nodes are involved in spreading this item⁶ and, as a metric, we use the availability of the item in the anchor zone. We define availability as the ratio of nodes in the anchor zone with the item to the total number of nodes in the anchor zone.

As our goal is to evaluate the general feasibility of floating content systems, we cover a very wide range of parameter choices. Below we describe the main parameters.

Number of nodes: Together with the world size, this defines the *node densities*. Node density is a crucial parameter for the feasibility of the system. If getting acceptable availability requires extremely high node densities, it is unlikely that we would encounter such situations in real life.

Communication range: This refers to the range at which nodes can send messages to each other. For simplicity, we assume a perfect circle of a given radius.

Anchor zone radius: This is the radius of the anchor zone, which is circular. Varying the communication range and anchor zone radius independently allows us to explore many different scenarios. However, we also discovered that scaling them with the same factor keeps availability at the same level.

Node mobility: We consider both static and mobile cases. In the mobility case, we used random waypoint model, and varied the speed of nodes.

We assume that when two nodes meet, they will be able to transfer the item from one node to another. In reality, this would depend on the actual contact time, but as discussed in Section III, generally, this should not be an issue.

VI. RESULTS

We briefly summarize the main findings from our evaluation. For a more thorough coverage including more graphs of the results, please see [14]. Below we present two of our main lines of evaluation so far.

How does mobility affect availability?

Figure 2a) shows the availability as a function of the number of nodes in mobile and stationary settings. The anchor zone radius was 500 m and the communication range 100 m. Another observation was that mixing mobile and stationary nodes gives the best availability, although in many cases having only mobile nodes comes quite close. Figure 2b) shows the availability as a function of the fraction of mobile nodes for 30, 50, and 100 nodes in the system. As we can see, mobility helps most in particular in cases with a small number of nodes, and that having only stationary nodes yields poor availability. As the number of nodes increases, the fraction of mobile nodes becomes less important; even a small fraction of mobile nodes is sufficient for good availability.

How does the ratio between anchor zone radius and communication range affect availability?

Figure 2c) shows the availability as a function of density (i.e., number of nodes). We plot several curves for different

⁶In some real scenarios, possibly only nodes interested in the item would spread it. Our results apply equally to this case, with the exception that the node density in such a scenario is calculated only over the interested nodes.

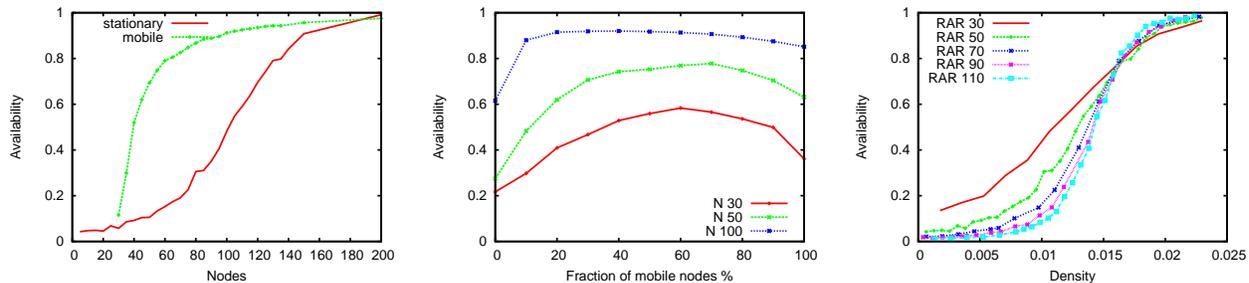


Fig. 2. a) Effect of mobility on availability (left), b) Mixing mobile and stationary nodes (center), c) Availability as function of density for different communication ranges. RAR denotes the ratio between anchor zone radius and communication range (right)

values of the ratio between anchor zone radius and communication range (RAR). Value 30 in this case means that the anchor zone radius was 3 times the communication range.

As seen in Figure 2c), cases with different anchor zone radii start to behave similarly beyond the *same* density. We call this the *critical density*, and it hints that there may be several fundamental properties inherent in floating content systems which depend on the node density, with availability being one example of them. The critical density was typically around 1 node per 50 m², i.e., relatively easily achieved in most urban settings. It exists both in stationary and mobile scenarios.

Environments with the same ratio for anchor zone radius and communication range behave similarly. This “scaling property” allows us to investigate many real-world scenarios with one simulation. For example, a communication range of 10 m and anchor zone radius of 50 m is the same as range of 100 m and zone radius of 500 m. Mobile environments with the same ratio for anchor zone radius and communication range do not follow the above scaling property as clearly as in the stationary case. Cases with smaller anchor zones obtain slightly inferior availability levels with lower node numbers, but fare better with bigger node numbers.

VII. DISCUSSION AND FUTURE WORK

Our results, both in this paper and in [14], show that floating content systems are feasible to build in realistic settings. The density required for 80% availability typically is around 1 node per 50 m², something easily achieved in most urban settings. Naturally, night time or other quiet periods pose a challenge. We can foresee two possible solutions. One, we simply ignore the issue and then applications of floating content are limited to cases which can live with information disappearing at nights. Two, we change the deletion functions such that information is not discarded according to the function $D(d)$, but can be kept until the node runs out of storage space. Given the amount of storage on mobile devices and the typically small items in floating content, this solutions does appear feasible.

Concerning mobility, the random waypoint model used in our preliminary study needs to be complemented with more appropriate models [15] and real-world traces [10]. Both of the above issues are part of our on-going evaluation of floating content systems. We also plan on implementing a prototype on the Maemo platform.

VIII. CONCLUSION

We have presented our work on evaluating the feasibility of floating content systems. Floating content is a fully distributed variant of an ephemeral content sharing service, solely dependent on the mobile devices in the vicinity using principles of opportunistic networking. Our results so far have shown that such systems are indeed feasible, and their requirements match the expected conditions in typical urban settings, e.g., in terms of node density, mobility, and communication ranges. Our future work revolves around continuing our evaluation and developing a prototype for floating content.

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] V. Lenders, M. May, G. Karlsson, and C. Wacha, “Wireless ad hoc podcasting,” *ACM/SIGMOBILE Mobile Computing and Communications Review*, 2008.
- [3] 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. BKKCS707, Nov. 2007.
- [4] D. Corbet and D. Cutting, “Ad loc: Location-based infrastructure-free annotation,” *ICMU 2006*, 2006.
- [5] I. Leontiadis and C. Mascolo, “Opportunistic Spatio-Temporal Dissemination System for Vehicular Networks,” in *MobiOpp Workshop*, Puerto Rico, USA, June 2007.
- [6] W. Gao, Q. Li, B. Zhao, and G. Cao, “Multicasting in delay tolerant networks: A social network perspective,” in *ACM MobiHoc*. ACM New York, NY, USA, 2009, pp. 299–308.
- [7] P. Hui, J. Crowcroft, and E. Yoneki, “Bubble rap: Social-based forwarding in delay tolerant networks,” in *ACM MobiHoc*. ACM, 2008, pp. 241–250.
- [8] 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.
- [9] G. Karlsson, V. Lenders, and M. May, “Delay-tolerant broadcasting,” *IEEE Transactions on Broadcasting*, 2007.
- [10] “Homepage of the Community Resource for Archiving Wireless Data At Dartmouth (CRAWDAD),” <http://crawdada.cs.dartmouth.edu/>, 2007.
- [11] 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.
- [12] A. K. Pietiläinen, “Experimenting with Opportunistic Networking,” in *Proc. of the MobiArch Workshop*, June 2009.
- [13] A. Keränen, J. Ott, and T. Kärkkäinen, “The ONE Simulator for DTN Protocol Evaluation,” in *SIMUTools*, 2009.
- [14] O. Karkulahti, “Distributed location-aware hovering information systems,” MS thesis, University of Helsinki, Oct. 2009.
- [15] F. Ekman, A. Keränen, J. Karvo, and J. Ott, “Working day movement model,” in *Proceeding of the 1st ACM SIGMOBILE workshop on Mobility models*. ACM, 2008, pp. 33–40.