

Autonomous Gossiping: A self-organizing epidemic algorithm for selective information dissemination in wireless mobile ad-hoc networks^{*}

Anwitaman Datta, Silvia Quarteroni, Karl Aberer

School of Computer and Communication Sciences (I&C),
Swiss Federal Institute of Technology (EPFL),
1015 Lausanne, Switzerland
{anwitaman.datta, silvia.quarteroni, karl.aberer}@epfl.ch

Abstract. We introduce autonomous gossiping (A/G), a new genre epidemic algorithm for selective dissemination of information in contrast to previous usage of epidemic algorithms which flood the whole network. A/G is a paradigm which suits well in a mobile ad-hoc networking (MANET) environment because it does not require any infrastructure or middleware like multicast tree and (un)subscription maintenance for publish/subscribe, but uses ecological and economic principles in a self-organizing manner in order to achieve any arbitrary selectivity (*flexible casting*). The trade-off of using a stateless self-organizing mechanism like A/G is that it does not guarantee completeness deterministically as is one of the original objectives of alternate selective dissemination schemes like publish/subscribe. We argue that such incompleteness is not a problem in many non-critical real-life civilian application scenarios and realistic node mobility patterns, where the overhead of infrastructure maintenance may outweigh the benefits of completeness, more over, at present there exists no mechanism to realize publish/subscribe or other paradigms for selective dissemination in MANET environments.

Keywords: Mobile Ad-hoc Network (MANET), Selective Information Dissemination, Content Driven Communication, Epidemic Algorithm, Self-organization, Any-to-Many Communication, Flexible casting

1 Introduction

Mobility is one of the most distinguishing traits of the present age. Infrastructure based networks have come of age, with support for online gaming in mobile devices to multimedia services. At the same time, mobile ad-hoc networks (MANET) have, from being used exclusively in mission critical military and recovery operations, permeated into our daily life. In the recent years, mobile

^{*} The work presented in this paper was supported (in part) by the National Competence Center in Research on Mobile Information and Communication Systems (NCCR-MICS), a center supported by the Swiss National Science Foundation under grant number 5005-67322.

ad-hoc networking is gaining popularity for non critical civilian applications. This has come about with the proliferation of devices like mobile phones, PDAs, digital cameras, portable music (MP3) players which support ad-hoc networking enabled by technologies like IEEE 802.11b or Bluetooth.

Infrastructure based cellular networks and MANETs have contrasting design, one based on dedicated infrastructure and global coordination, while the other is decentralized and self-organizing. But from application and usage point of view, most likely they will play complementary roles. Thus, despite the fact that most work done until now in MANET community has focussed on point-to-point routing, most likely communication with another particular node will not be the defining application for MANETs, because of the difficulties of route maintenance in a dynamic setting, and simultaneous communication between several pairs, and communication among long distance nodes. MANETs will instead do what peer-to-peer has done to the wired world, bringing the resources at the edges of the wireless network to collaborate together for better resource usage. Always using the infrastructure can be expensive, so using the MANET wherever possible can bring immense cost benefit to the end users, and also reduce the load on the infrastructure, as well as extend its coverage. Such coexistence of MANET and cellular network can be used for more purposes than one.

- Multi-hop routing in cellular networks is used to extend the coverage of cellular networks. (e.g., [20])
- The MANET can be used as a cache, such that content can be discovered in a peer-to-peer manner whenever possible, instead of each user individually downloading it using the infrastructure. (e.g. [17, 25])
- Using the MANET for geographically confined information dissemination or group communication.

The focus of this paper is on information dissemination. There is an increasing opportunity for people to share resources (e.g., files) in a peer-to-peer manner, even in a wireless environment, using ad-hoc networking. One way to go about it is when a mobile node makes an explicit request for a resource, and the whole network is flooded with a query, as is the case with many mobile ad-hoc route discovery algorithms [27, 38], similar to file discovery by query flooding in P2P networks like Gnutella [12]. Flooding in a wireless network is in fact relatively efficient as compared to in wired networks because of wireless multicast advantage [42]. Improvements of the basic flooding approach using advertisements and geographic information have also been recently studied [41].

The dual problem to searching resources in the network, that of disseminating resources selectively in a wireless mobile ad-hoc network is yet relatively unexplored, and is the focus of this paper. So far there exist mechanisms to broadcast information in the complete network [39], or in a specific geographic area (Geocast) [35, 33], apart from to any one specific mobile node (unicast/mobile ad-hoc routing [27, 38]) or any one arbitrary node (anycast). But selective dissemination paradigms based on mobile nodes' content requirement is still a relatively unexplored area. Proactive multicast group establishment, maintenance and usage can be prohibitive, particularly given the fact that membership varies over time

(managing group membership is difficult), content requirement can be diverse (need for multiple multicast groups), and the network itself is very dynamic because of node mobility, capacity and energy constraints.

Recently on demand routing has been extended to flood the whole network to locate not any particular node but to rather locate all nodes with a particular content requirement [8], but it still relies on and has the overheads of broadcasting and does not scale well with respect to increasing number of content categories, nor adapts well if the population of interested parties vary over time. A similar approach where the source has global information of the destination nodes is proposed for differential destination multicast [30], but such approaches can work only in small groups.

We propose autonomous gossiping (A/G) for selective information dissemination on the lines of intentional multicast [1]. However intentional multicast uses Intentional Name Resolvers to form an overlay to support routing. In contrast, A/G is a candidate solution for an infrastructure-less mobile ad-hoc networking environment. A/G's dissemination mechanism is essentially stateless, where neither routing information is required nor global information like knowledge of the destination nodes is required, nor does A/G flood the whole network. It is important to note that even overlay maintenance using soft-states is very expensive, and sometimes simply impossible in a highly mobile environment. Thus, while there exists proposals like tree based multicasting for information dissemination which will work under the assumption of quasi-static environments (like sensor networks or overlays on top of internet infrastructure), we are proposing a best effort mechanism for any arbitrarily dynamic environment where construction and maintenance of infrastructure (overlays) is prohibitive, while retaining the selectivity instead of broadcasting in the whole network. A/G spreads in the network by following a greedy mechanism much like an epidemic does, such that it spreads to immediate neighbors that are interested in the particular content (susceptible to the epidemic), and yet avoiding the ones that are not interested in the content (resistant to the epidemic). Node mobility is exploited to enable interaction with newer nodes. Thus, while mobility is detrimental to mechanisms that use states (routing information), mobility is helpful to a stateless mechanism like A/G.

Some possible applications for A/G include sharing files, directed advertisement [34, 3] for mobile commerce and tourism supported with reduced infrastructure, or dissemination of meta-data about services being provided. Meta-data so disseminated is like advertisement, which can then be used by the receivers to request for and use some other service. It is likely that lightweight infrastructure (infostations [23]) is deployed to "inject" data in the MANET from time to time, for example, timetables for local transportation or menus of a restaurant. Or otherwise, some members in the MANET network may also download the resource using the cellular network, while the others avail it in a peer-to-peer manner. While at other instances the mobile nodes will already carry or generate data of their own, e.g., music files or pictures from digital camera. Or alternatively, the meta-data can be about the quality of a resource being provided by

a service provider, and an user can gather such feedback from fellow users before paying for the resource. So to say, A/G can play the role of word of mouth advertisement even for resources offered by the dedicated service providers.

So far, in a decentralized environment, all advertisement mechanisms propose that sources push the information while the receivers are supposed to pull the information, and the schemes vary in how to make the push/pull efficient and reliable. These mechanisms are variously known as match making, geographic hash tables (GHT) or content based multicasting (CBM) [34, 3, 37, 44]. From advertisement perspective, dissemination schemes in general should and A/G in particular takes this a step forward, and instead of both source and destinations participating actively (pushing and pulling), the information is pushed to targeted audience, while trying to avoid uninterested users. From dissemination perspective the novelty of A/G is that it disseminates without the need to maintain routing or other soft-states (e.g., an overlay) to support multicast, unlike [40, 5, 33].

Furthermore, in this paper we do not consider the mechanisms by which the data items enter the network or are annotated, nor the copyright issues of the disseminated information, but only the epidemic mechanism by which they are disseminated to other nodes that are potentially interested in the data, while scrupulously avoiding spam.

1.1 Application scenarios

Networking the automobiles on the roads: One of the niche areas where ad-hoc networking has a promising future is the automobiles on the road. Consider that some person may be interested in the opening hour of a ski resort and the quality of snow on a particular day, while another person may be interested in the traffic information after a particular exit on the highway, while yet another person may be interested in the cultural programmes in offing in the next city, while another person may be interested in architectures. Broadcasting in the MANET network would result in everyone being informed of all the things, causing information overload, but selectively propagating it on the basis of preferences will mean that if two users with similar profiles come in contact with each other, they will exchange their information. This also means that users may not always get the information they desire, but at the same time, they won't be spammed by multiple other unnecessary information. Thus the two metrics that can be used to measure the quality of dissemination are the standard information retrieval metrics of completeness and precision [4].

Specialized units in a disaster recovery mission: Consider another application related to another niche application of MANETs, that of disaster recovery. A/G can play a significant role in effective coordination of a team of mobile units, where different units have different specializations (and hence different preferences for receiving information). If the information is about a suspected alive person, most likely an unit with a sniffer dog and an unit for removing rabbles to

rescue the trapped person are needed, while, if the information is that a person has already been rescued, a medic will be needed instead. At the same time, in order to facilitate smooth working of each of the units independently, it is best if they are not distracted by information that is not of any immediate concern to them.

Apart from the completeness and precision [4], the latency of the information dissemination is also important, hence we'll measure the quality of selective dissemination based on the evolution of completeness and precision in the system with respect to time.

One of the benefits of A/G is that by tuning the dissemination criteria, users can achieve flexible casting, including broadcast, geocast, multicast and anycast. Other applications include environment monitoring and distributed aggregation, sharing content (picture, ringtones, music, gaming profiles, dating profiles) and advertisements.

1.2 Paper organization

We give an overview of A/G in Section 2, where we elaborate A/G's salient features. The system model studied in this paper is discussed in Section 3 followed by simulation results in Section 4. In Section 5 we summarize the related work. We conclude in Section 6, along with a glimpse of future possibilities.

2 Autonomous gossiping in mobile ad-hoc networks

The essential idea of any proactive dissemination scheme is to provide a mechanism so that mobile hosts do not have to search for resources when they are needed, but rather they automatically get the same (pushed). In autonomous gossiping, instead of mobile nodes making the decision to push data to each other, it is the data items themselves that try to identify other hosts which are suitable, based on the data item's own profile and host's advertised profile (similar to a subscription).

Hence, the dissemination aspect, which is the focus of this paper is a self-organizing mechanism whereby there is no need to explicitly build and maintain subscription lists, but dissemination is purely based on locally available information (mobile node and data profiles) and autonomous decisions.

Here we give an intuitive introduction to the concept of A/G and its distinguishing features. A more concrete description of a system model to realize autonomous gossiping for proof of concept is described in the next section.

When we travel around, we meet friends, strangers and sundry. We talk, exchange news and information, that is, we gossip. While the encounters are random, the gossip is often with only individuals with whom there is a shared interest. When we move around, similar to our social encounters, the mobile devices (mobile hosts in a MANET) we carry also come in contact with other mobile hosts as well as possibly fixed beacons. Just like the real world counterpart of "interesting information", such interacting nodes may have assets (any

electronic resource, from now on referred as data items), which might be of interest for each other. By enabling these nodes to autonomously gossip among themselves according to their mutual profiles, people (their mobile devices) may come across important or at-least interesting data items. This intuition motivated us to define the notion of autonomous gossiping. A/G may possibly be implemented using various strategies. Here we describe a novel strategy which uses ecological and economic paradigms. We consider that mobile nodes are like habitats, where the node profile (interests) defines the preferred data items. Data items compete among themselves for limited resource (e.g., memory) in these mobile nodes (hosts). In order to enforce selectivity and discourage spam, hosts reward (or punish) data items required (not required) locally, thus in the long run hosting relevant data items and purging unnecessary ones. Data items, in order to survive, should move to hosts where it is required. This enforces A/G’s selectivity and underlines the self-organizing aspect of the dissemination scheme. A/G has the following salient features.

Exploiting mobility: Unlike traditional approaches where node mobility is the source of difficulties, A/G utilizes the opportunity of coming in contact with new nodes arising from mobility to disseminate data. This idea of exploiting mobility is similar to Smart-Tag [6] and work being done at Media Lab Asia, like DakNet and RuralWiFi [2], where mobile objects (like buses) are equipped with Bluetooth/802.11b in order to “carry” data/e-mails back and forth between fixed points, for instance rural areas with no wired infrastructure and well connected urban localities. Other applications exploiting mobility include routing information dissemination based on encounters [24].

Selectivity: The information dissemination aspect of A/G is essentially an epidemic algorithm but unlike previous usage [31, 7, 15] for broadcast, A/G is selective in nature. Each unique data item may be considered as a different epidemic, and multiple epidemics spreading in the network simultaneously, each epidemic selectively infecting network members based on their vulnerability (advertised profile). Self-organization makes the selectivity (propagation of the epidemic to susceptible nodes) autonomous, thus contributing to the name “autonomous gossiping”.

Self-organization: What differentiates autonomous gossiping as a selforganizing rather than just a distributed decentralized algorithm as is the case for other rumor/epidemic algorithms [31, 7, 15], is its use of the economic and ecological paradigms that helps it adhere to Francis Heylighen’s characterization of self-organization [26]: “The basic mechanism underlying self-organization is the *noise* driven variation which explores different regions in a system’s *state space* until it enters an *attractor*”. The *state space* consists of any and all mobile nodes in the system being populated by any and all data items. The *noise* results from the motion of nodes which brings random nodes within radio contact of each other, random decisions at mobile hosts to use and reward (or otherwise punish) data items and autonomous local decisions by data items to migrate or replicate to neighboring nodes or do nothing. The *attractor* is the data distribution at different mobile nodes retaining the selectivity such that nodes host only data items

matching their own need (profile), and mobile nodes have high utility for storing the data items locally and acting as a data carrier, and data items replicating as often as possible, and still retaining high average utility, so that they have better utility and chance of staying in the network. This is in contrast with how rumor spreading has been dealt in other contexts [7, 31, 15] where there is no need of self-organization since the goal is to essentially broadcast the rumor/epidemic to a complete well-defined population, and different rumors are independent of each other, particularly, not competing among themselves for limited resources.

Flexible-casting: The self-organizing and middlewareless nature of A/G makes it suitable to support any of broadcasting, multicasting, geocasting and combinations therefrom without the need of routing information, and hence A/G is a suitable candidate for flexible casting independent of the endpoint-to-endpoint routing protocol being used. We call the resulting casting as *flexible-casting* since each of the above mentioned casting mechanisms can be seen as and achieved as special cases of flexible-casting based on the context description, that is the content and location information associated with individual data items being gossiped. This is later explained in Table 1.

3 System model

The concept of A/G based flexible casting is generic. Rest of the paper describes one way to realize such a selective dissemination. Hence the system model, and policies described next are not necessarily the unique way to realize A/G, but is one candidate solution. The results in the paper are also specific to the model described here.

A/G can be used not only in the context of MANET environment, but also others, for instance, in a sensor or peer-to-peer network. For P2P networks we will then have to emulate and model random interactions among peers.

In this paper, we concentrate solely on mobile ad-hoc networks, where nodes' mobility leads to their random meetings. The objective in doing so is to give a proof-of-concept of how information can be disseminated to interested parties in a decentralized network in a self-organizing manner using local interactions, with minimal spamming.

For this paper, we assume an isolated system comprising of a fixed number of mobile nodes confined in a predefined geographic region. These nodes are mobile, and communicate with each other in a wireless ad-hoc manner. To start with each of the nodes carries some unique data items. During the period for which the system is studied, no new node or data item is introduced, nor any existing node leaves the region. However, useless data items are potentially purged from the system. Though we study a simplified system with no new data items or nodes joining it, this assumption does not affect the dissemination mechanism.

Each of the hosts m_k has a memory limit of M , and individual profile m_k^p . The profile presently comprises of abstract categories C_j s. These categories may correspond to topics or more elaborate content description as in publish/subscribe systems. When using textual annotations, topics may correspond to discrim-

inative keywords or to abstract categories extracted from larger text collections, e.g., using latent semantic indexing techniques [4]. Thus for instance, $m_1^p = \{C_1, C4, C7\}$. The hosts are initialized with random number of ($\leq M$) unique data items d_j s, each of which has its own profile d_j^p , described in a manner similar to mobile host profile description. Data items also have a scalar called the associated utility, d_j^u . Moreover, data items are initialized with an attribute d_j^l , which defines the target geographical locations (zones) the item is supposed to belong to. d_j^l may have one or multiple entries. Thus, a data item d_j is represented as a tuple $\{d_j^p, d_j^u, d_j^l\}$. Every topic in a host profile m_k^p is associated with weighting according to the degree of interest (affinity) of the host m_k to the topic MW_k^i . Similarly profile d_j^p of data item d_j has affinity DW_j^i according to the relevance of the data item d_j to the topic i . Similarity/affinity is computed using standard techniques defined for information retrieval [4]. The use of affinity information is described later, in Section 4.2. Each host also has its next destination as its goal zone, which is used to determine whether the host is suitable for data items with nontrivial location attribute d_j^l .

It is also possible for a data item to be relevant to all topics. This is modelled in the host and data item profile by the presence of a unique category, named “*any*”. Similarly data items may be relevant to all locations, in which case its location context is represented as a special category “*any*”. Such a model for describing profiles gives the flexibility of dissemination as described in Table 1.

Table 1. Profile choice and corresponding flexible-casting scheme

Flexible-casting scheme for a data item	content profile d_j^p	location attribute d_j^l
Broadcast	any	any
Geocast	any	one destination
Multiple Geocast	any	multiple destinations
Content based dissemination	content description	any
Content and location context based dissemination	content description	one or multiple destinations

Communication model We assume a symmetric communication model, such that hosts can communicate with other hosts within a r_{comm} radius. Each host also advertises its own profile and its target destination (goal), such that data items resident at the other neighboring hosts can see the advertised profile, and thus decide to replicate, migrate or do nothing.

Host Mobility The mobility of hosts is a hybridized version of city walk and random walk [9]. At initialization, several anchor points (A_k) are defined in the area. Each host chooses an anchor randomly as its next target destination, and moves toward the anchor at a variable speed parameterized by average speed. Occasionally perturbation is added to the host motion, such that it deviates from a linear path. Once a host reaches within a radius r_{anchor} of target anchor point,

it chooses a new target destination and continues. This is intended to emulate motion in a populated area like a town or city, where most of the motion is on streets between distinct locations.

Present simulations assume that the set of possible targets for a given host coincides with the set of the possible destinations for a given data item. Further, we assume that hosts advertise their target location to neighboring nodes.

At every time step t , a host probabilistically uses some of its resident data items if their profiles have something in common (content-based match) or if the host and data item destinations coincide (location-based match).

Therefore, if d_j is residing in m_i and $\text{similarity}(d_j^p, m_i^p) \geq \text{threshold}$ where threshold is a predefined threshold, or $\text{destination}(m_i) \in d_j^l$, then d_j is probabilistically used and rewarded (for example, by incrementing its utility, $d_j^u(t) = d_j^u(t-1) + 1$). The probabilistic decision to use can possibly be weighted according to the degree of similarity, or can be a constant. If similarity is below threshold then data item is not used, and when any data item resides in a host for a period of time T_{latency} without being used, the data item is penalized (utility is decremented). These are implementation details, and in a very general case, the choice of $\text{similarity}(\cdot, \cdot)$ function [4], threshold , and the probabilistic usage policy as well as reward/punishment policies can be decided by individual hosts.

In the simulations, we assume all hosts use the same set of functions and policies, and is elaborated in Section 4.

3.1 Policies

Depending on the hospitality received at the present host, data items decide to either continue to reside, migrate or replicate to another host with a more suitable profile and/or goal zone, and the data item's associated utility is used in the decision process.

To enable such decisions, threshold values, th_l and th_h ($th_l \leq th_h$) for utility are used by the data items, as elaborated next. The data item's present host is denoted as m_h and the neighboring mobile host where the data item may choose to migrate or replicate is denoted as m_n . These policies are essentially heuristic, and other decision criteria too may be used for defining A/G policies.

Migration: A data item decides to migrate from mobile host m_h to any mobile host m_n if $\text{similarity}(d_j^p, m_i^p) < \text{threshold}$ and $d_j^u \leq th_l$. If $d_j^u > th_l$ then the data items can afford to wait for some better host instead of migrating desperately. These conditions are implicitly bio-inspired, where data items desperately migrate from a hostile environment in search of a better habitat, however if they can afford to wait for better hosts then they wait, since desperate migrations don't guarantee better environments.

The utility value d_j^u may either be incremented as a reward for making a prudent decision to leave an inhospitable host, or it may be left unchanged. In our simulations presented in this paper, we have chosen the former, that of rewarding the migrating data items.

Replication: If the data item has a high utility, it can afford to replicate and thus increase its population. This is again bio-inspired, where reproduction is possible in a favorable environment. In order to replicate, the data item is required to pay such that if a data item replicates, then both the resulting replicas have an utility value that is decremented from the original utility value before replication. This is the case if $d_j^u \geq th_h$ and $similarity(d_j^p, m_i^p) \geq threshold$. However if the utility is relatively low, $th_l \leq d_j^u \leq th_h$, the utility value is not decremented for replication, thus providing a mechanism for data items' survival. However, this case may be eliminated by choosing $th_l = th_h$ and otherwise, this should be a narrow band, so that it can not be abused by data items in order to replicate unnecessarily, since that will lead to spam.

Note that if the number of replication is restricted to a predefined number, it is also possible to realize multicast [10] (though not discussed in this paper or in Table 1).

Replica reconciliation: If a data item finds one of its own replicas being present in the new host to which it migrates or replicates, then only one replica continues to reside at the host, with the utility being one more than the maximum utility of the two replicas. Since two replicas chose the particular host, it is very likely that the host is appropriate for the data item, and though only one replica is retained (to make space for a greater diversity), incrementing the utility provides better chances for its survival as well as future replication.

Migration anyway: When location is a matching criterion and one data item is located on a host which does not match its location attribute, it tries to migrate to all neighboring hosts with that location as a goal zone, if any. The intuition is to reach the target location first. In case of Geocast that is good enough. Even for the case where both content and location are important, reaching the location first and then spreading locally to other and more significant hosts makes it easier to achieve completeness, at the cost of precision of content match.

The data items (replicas) compete with each other for the limited memory at each hosts, and the evolutionary paradigm of *extinction of the unfit* is used, so that data item with least utility value is discarded in the event of memory shortage, and in the event of a tie, it is randomly chosen. Other elimination policies like LRU (least recently used) are alternative tie-breaker.

The objective of the data items is to have as high a degree of replication as possible, and still have a high utility value. For this they must identify the suitable hosts (profiles). The hosts on the other hand need to maximize the aggregate utility value of the resident data items. This completes the ecological and economic model for data items competing for survival, and mobile nodes hosting and rewarding relevant data items while punishing and purging other data items. This ensures a symbiotic relationship rather than a parasitic one between mobile nodes and data items.

4 Simulation

A Java based discrete time simulator has been developed in order to simulate and study the effectiveness of autonomous gossiping as a information dissemination paradigm. Mathematica's J/Link extension is used to integrate display of statistical information. Since the initialization as well as the rest of the simulation is randomized, the results are not reproducible, and give only a qualitative understanding of how autonomous gossiping works. Multiple runs of the simulation though give similar qualitative results.

4.1 Initialization

A rectangular region of dimension 1000 units * 1000 units (a square) was simulated, with five randomly chosen anchor points in the region for node mobility. 20 mobile nodes with a memory for a maximum of 20 data items each, and initially storing 179 distinct data items in total were instantiated. Four anchor points were used.

Each host had a maximum of 3 content categories to be advertised, while each data item also had at most 3 content categories. There were 20 possible categories to choose from in order to describe content. For geocast, the content description of all data items were "any".

Data items were penalized if not used in 10 consequent time units ($T_{latency}$). Data items with some matching profile with the host were used in every time step with a probability of 0.05. Mobile hosts could communicate with other hosts within a range of 20 units (r_{comm}). The lower and upper thresholds determining the replication/migration policies were chosen as $3(th_l)$ and $5(th_h)$. The implicit simplifying assumption for the initial study is that all data items have equal size. The random instantiation of data items, their placements in the hosts and location of anchor points and hosts had been stored in a log file, such that all simulations had the same initial condition (wherever valid). The content and location context of data items were of course in accordance to which casting technique was being simulated. All simulations were run for 500 time steps.

Note that we do not use any real life units for measuring distances or time, but use artificial ones instead (conveniently called just units), because the simulations were not meant to quantify A/G's performance if used in any particular setting or technology (IEEE 820.11b, Bluetooth or any other), but only to make a case for A/G as a selective dissemination mechanism in MANET environments in general. Consequently, it is possible that some of the assumptions made in the simulation are not yet supported by the existing technologies. Even in that case, it is our belief that the assumptions are not a fry cry from the reality, and will be realized in near future.

4.2 Results

Given space constraints and the fact that broadcast has been exhaustively investigated by many researchers, we concentrate and provide results for the other cases.

The results demonstrate the quality of information dissemination achieved using A/G. We measure the quality of information dissemination on the basis of two standard metrics used in information retrieval [4], namely *recall* and *precision*, and the latency in achieving the corresponding completeness and precision.

Recall (now onwards called *completeness* in this paper) signifies the percentage of mobile nodes reached as compared to the total target set of nodes. In this paper we use the term completeness instead of recall since it is more intuitive from the broadcasting/multicasting perspective.

Precision is determined by the percentage of the reached mobile nodes that are actually interested in the data item. Thus precision is essentially the reverse of the degree of spam caused because of the dissemination mechanism.

The target set of interested nodes comprises of all the nodes which share a commonality of interest with the data item's content (determined using similarity metric) and/or location context.

For the simulations, if data item d_j has content C_i with affinity DW_j^i and mobile node m_k has an interest in C_i with affinity MW_k^i (it is assumed that these affinities are precalculated using standard information retrieval techniques [4]), then if $|(MW_k^i - DW_j^i)| \leq \textit{threshold}$ then we say that the mobile node m_k is interested in the data item d_j because of the C_i content. Thus $|(MW_k^i - DW_j^i)|$ determines the dissimilarity¹. A smaller *threshold* means a stricter matching criteria and vice versa.

So if $\exists C_i \in d_j^p, m_k^p$ s.t. $|(MW_k^i - DW_j^i)| \leq \textit{threshold}$ then we can say that m_k is interested in d_j , otherwise m_k treats d_j as spam.

For commonality of interest in location context, the intuition is to reach the target location(s) first, as has been elaborated in Section 3.1.

Note that if d_j^p or d_j^l is "any", then the respective matching condition is evaluated as true. The *x-axis* in all the figures represent time.

Content based dissemination: In Figure 1 and Figure 2 we show the completeness and precision of location based A/G for various *threshold* values, where the dissemination is content based. We observe that as far as completeness is concerned, A/G is effective as a content based dissemination scheme, with completeness value getting close to one. However precision of A/G decreases (spam increases) with stricter matching criterion (lower threshold). In order to fulfill the completeness property, a data item needs carriers, more of which consider the data items as mismatch if a stricter matching criterion is used. While it gets harder to find matching nodes, the data items need to find relatively more carriers. This conflict of interests leads to lesser precision when the matching criterion is stricter.

Geocast: Figure 3 shows the completeness and precision of A/G for Geocast. The precision is high since the choice of carrier nodes is easier to make for the

¹ Note the subtle difference of usage of *threshold* for dissimilarity instead of similarity (in contrast to the usage in Section3.1).

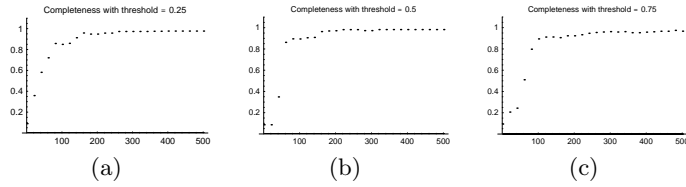


Fig. 1. Completeness of content based dissemination

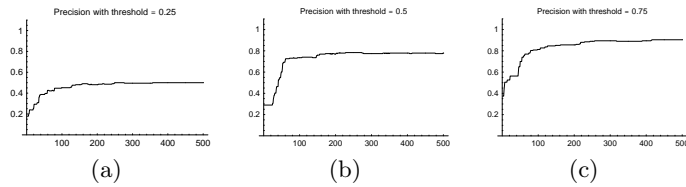


Fig. 2. Precision of content based dissemination

data items. However, since the nodes are assumed to be continuously moving it is difficult to achieve completeness because the membership of the target population itself keeps varying over time. Hence, in Figure 3(b) we measure completeness in a different way. If a data item belonging to a particular location reaches that location once, we consider the purpose of A/G to be achieved.

Dissemination of the data item within the zone is essentially beyond the scope of A/G, however simple extrinsic mechanism like geographically constrained broadcast may be used to that end.

The difficult part is to ensure persistence of the data item within the geographic region after reaching the region, since the carrier nodes are free to move in and out of the region. However, this will be a problem for any mechanism, since the fluctuation is purely because of the mobile nodes' movement. On the other hand, as a best effort solution to the problem, the data item migrates from the current host that is moving away from the target zone to new nodes which move towards the target zone.

Content and location context based dissemination: Next we show the precision based on both content and target location matches and completeness achieved. We show in Figure 4 and Figure 5 the cases where the content matching threshold of 0.75 and 0.5 have been used respectively. With target location and content both being matching criteria, it is even harder to find suitable mobile nodes, and as a consequence both the precision and completeness are inferior in comparison to when the dissemination is based on either one of the criteria. Noteworthy is the fact that a relatively relaxed content matching (higher *threshold*) leads to an acceptable performance.

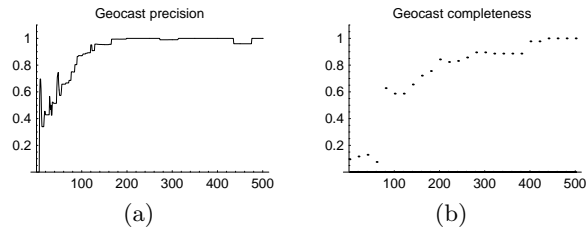


Fig. 3. Geocast - Dissemination in a targeted location

Note that in these simulations there were only 3 content descriptors for each of host and data item profiles out of a possibility of 20, and hence the threshold of 0.75 still means that there is at least one common content in the two profiles (host and data item), even if the matching is not very tight.

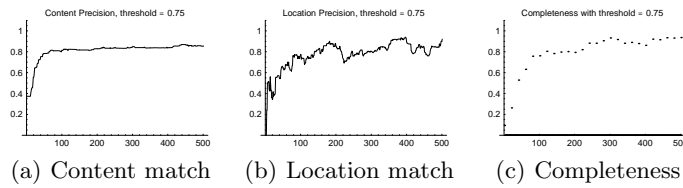


Fig. 4. Content and location context based dissemination (threshold = 0.75)

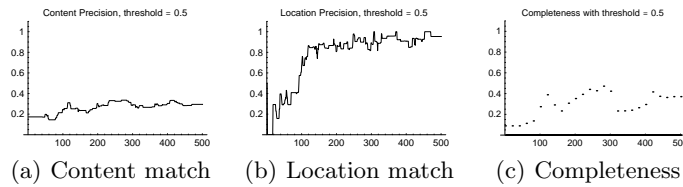


Fig. 5. Content and location context based dissemination (threshold = 0.5)

The initial results described above show that A/G is a robust and effective epidemic mechanism to selectively disseminate information, where the selectivity is based on content and/or the location context of the information (data item). The precision metric is to measure the effectiveness of selectivity, and the completeness metric for the effectiveness of dissemination.

5 Related work

The work presented in this paper has similarities and differences compared with several domains, as mentioned below:

- Location dependent service (LDS) [21] perspective
- Manet routing/casting perspective
- Caching [17], pre-fetching(7DS [36]) and collaborative replication [25]
- Spatio-temporally restricted dissemination [43]
- Peer-to-Peer systems like Freenet [11] and work done in the group communication and publish subscribe domain.

Because of space constraint, we only mention them above, and a more exhaustive related work can be found in the accompanying technical report [16], where we elaborate the differences in A/G and how A/G mostly complements these existing techniques.

6 Conclusion

Though research in mobile ad-hoc network is predominantly networking centric, the idea of content driven communication instead of IP driven point-to-point communication has been of interest for a while [28, 1, 11]. This paper extends and explores the ideas of content centric communication, particularly studying dissemination based on epidemic mechanisms, and still preserving selectivity in a self-organizing manner. Most epidemic algorithms that have so far been investigated, either in the context of wired or wireless and ad-hoc networks are for communication in the whole network, or a predefined subset (broadcast/multicast) [19, 31, 39, 7, 15, 22, 18, 32].

Thus autonomous gossiping is a new genre of epidemic algorithm resembling more closely actual epidemic spreading, and it brings the two domains of content-based communication and epidemic algorithms together, specifically in the context of MANET. A/G spreads data items selectively based on vulnerability of other nodes, instead of treating all nodes homogeneously and flooding the network.

Self-organizing economic and ecological mechanisms of rewards and punishments and competition among the data items in order for data items to survive and propagate in the network help enforcing A/G's selectivity unlike previous mechanisms that rely on goodwill rather than enforcement of selectivity. A/G has certain characteristics that qualify it as a proactive mechanism similar to pre-fetching schemes, and at the same time has on-demand characteristic, if seen from the view-point of intentional multicast [1].

We also introduce the notion of flexible casting, such that the same mechanism can be used for either of broadcasting, multicasting, geocasting or combination of multicast and geocast. A/G's stateless nature of communication helps achieving such a flexible casting without the need to establish and maintain separate routing tables.

The system model (Section 3) including the assumption of a mobile ad-hoc network, node mobility, communication and policies are implementation details used for proof of concept, but the idea of A/G itself is more generic and applicable in environments other than MANETs.

The results presented in the paper are initial and meant to provide a qualitative study of the feasibility and effectiveness of A/G rather than comparing or quantifying performance issues like expected latency, bandwidth and energy usage of the mechanism in more realistic environment [29] models.

A/G in MANET is similar to and thus can also be used to simulate and study how diseases spread in nature, since the motion model and epidemic propagation maps to the real world unlike some of the previously studied epidemic algorithms [7, 31, 15]. Theoretical biologists and statistical physicists too study epidemic spreading with immunization and mutations [14], and try to discover effective immunization mechanisms [13] to prevent spread of epidemic. Our goal of information dissemination is dual to that of theirs, and hence provides complementary perspective to the same problem, and understanding either (spread/immunization) will add to understanding of and achieving the complementary goals.

Acknowledgement

The authors will like to thank Ms. Emma Frejinger for her help with writing an initial version of the simulator for autonomous gossiping during her stint as a female undergraduate summer intern (2002) sponsored by NCCR-MICS and hosted at Distributed Information Systems Laboratory (LSIR) at EPFL.

References

1. W. Adjie-Winoto, E. Schwartz, H. Balakrishnan, and J. Lilley. The design and implementation of an intentional naming system. In *Symposium on Operating Systems Principles*, pages 186–201, 1999.
2. Media Lab Asia. DakNet, Rural WiFi. <http://www.medialabasia.org/>.
3. I. Aydin and C. C. Shen. Facilitating match-making service in ad hoc sensor networks using pseudo quorum. In *ICCCN*, 2002.
4. R. A. Baeza-Yates and B. A. Ribeiro-Neto. *Modern Information Retrieval*. ACM Press / Addison-Wesley, 1999.
5. S. Banerjee, C. Kommareddy, K. Kar, S. Bhattacharjee, and S. Khuller. Construction of an efficient overlay multicast infrastructure for real-time applications. In *INFOCOM*, 2003.
6. A. Beaufour, M. Leopold, and P. Bonnet. Smart-tag based data dissemination. In *First ACM International Workshop on Wireless Sensor Networks and Applications (WSNA02)*, June 2002.
7. K.P. Birman, M. Hayden, O. Ozkasap, Z. Xiao, M. Budiu, and Y. Minsky. Bimodal multicast. *ACM Transactions on Computer Systems*, 17(2):41–88, 1999.
8. C. Borcea, C. Intanagonwiwat, A. Saxena, and L. Iftode. Self-routing in pervasive computing environments using smart messages. In *Proceedings of the First IEEE International Conference on Pervasive Computing and Communications (PerCom'03)*. IEEE Computer Society, March 2003.

9. T. Camp, J. Boleng, and V. Davies. A survey of mobility models for ad hoc network research. *Wireless Communications & Mobile Computing (WCMC): Special issue on Mobile Ad Hoc Networking: Research, Trends and Applications*, 2(5):483–502, 2002.
10. C. Carter, S. Yi, P. Ratanchandani, and R. Kravets. Manycast: Exploring the space between anycast and multicast in ad hoc networks. In *ACM MobiCom*, 2003.
11. I. Clarke, T.W. Hong, S.G. Miller, O. Sandberg, and B. Wile. Protecting free expression online with freenet. *IEEE Internet Computing*, 6(1):40–49, Jan./Feb. 2002.
12. Clip2. The Gnutella Protocol Specification v0.4, Jun. 2001.
13. R. Cohen, D. Ben-Avraham, and S. Havlin. Efficient immunization strategies, 2003 (Submitted at Physical Review Letters).
14. S.M. Dammer and H. Hinrichsen. Epidemic spreading with immunization and mutations. *Physics Review E*, March 2002.
15. A. Datta, M. Hauswirth, and K. Aberer. Updates in highly unreliable, replicated peer-to-peer systems. In *Proceedings of the 23rd International Conference on Distributed Computing Systems, ICDCS*, 2003.
16. A. Datta, S. Quarteroni, and K. Aberer. Autonomous Gossiping: A self-organizing epidemic algorithm for selective information dissemination in wireless mobile ad-hoc networks. Technical Report IC/2004/07, Swiss Federal Institute of Technology, Lausanne (EPFL), 2004. <http://lsirwww.epfl.ch/AutoGoss>.
17. N. Davies, K. Cheverst, K. Mitchell, and A. Friday. Caches in the air: Disseminating tourist information in the guide system. In *In Proceedings of Second IEEE Workshop on Mobile Computing Systems and Applications*, 1999.
18. C. de M. Cordeiro, H. Gossain, and D.P. Agrawal. Multicast over wireless mobile ad hoc networks: present and future directions. *IEEE Network Magazine*, 17(1):52–59, Jan-Feb 2003.
19. A. Demers, D. Greene, C. Hauser, W. Irish, J. Larson, S. Shenker, H. Sturgis, D. Swinehart, and D. Terry. Epidemic algorithms for replicated database maintenance. In *Proceedings of the Sixth Symposium on Principles of Distributed Computing*, pages 1–12, 1987.
20. O. Dousse, P. Thiran, and M. Hasler. Connectivity in ad-hoc and hybrid networks. In *Proc. IEEE Infocom*, June 2002.
21. M.H. Dunham and V. Kumar. Location dependent data and its management in mobile databases. In *DEXA Workshop*, pages 414–419, 1998.
22. P. Th. Eugster and R. Guerraoui. Probabilistic multicast. In *3rd IEEE International Conference on Dependable Systems and Networks (DSN 2002)*, pages 313–322, June 2002.
23. R. Frenkiel, B.R. Badrinath, J. Borras, and R. Yates. The Infostations challenge: Balancing cost and ubiquity in delivering wireless data. *IEEE Personal Communications Magazine*, Feb. 2000.
24. M. Grossglauser and M. Vetterli. Locating nodes with ease: Mobility diffusion of last encounters in ad hoc networks. In *The Proceedings of IEEE Infocom*, 2003.
25. T. Hara. Effective replica allocation in ad hoc networks for improving data accessibility. In *Proceedings of IEEE INFOCOM*, pages 1568–1576. IEEE Computer Society, 2001.
26. F. Heylighen. The science of self-organization and adaptivity. *The Encyclopedia of Life Support Systems*, 2002.
27. X. Hong, K. Xu, and M. Gerla. Scalable routing protocols for mobile ad hoc networks. *IEEE Network Magazine*, 16(4):11–21, Jul-Aug 2002.

28. C. Intanagonwiwat, R. Govindan, and D. Estrin. Directed diffusion: A scalable and robust communication paradigm for sensor networks. In *Mobile Computing and Networking*, pages 56–67, 2000.
29. A. Jardosh, E. M. Belding-Royer, K. C. Almeroth, and S. Suri. Towards realistic mobility models for mobile ad hoc networks. In *Proceedings of Mobicom*, 2003.
30. L. Ji and M. S. Corson. Differential destination multicast - a manet multicast routing protocol for small groups. In *Proceedings of Infocom*, 2001.
31. R. M. Karp, C. Schindelhauer, S. Shenker, and B. Vöcking. Randomized rumor spreading. In *FOCS*, 2000.
32. A. Khelil, C. Becker, J. Tian, and K. Rothermel. An epidemic model for information diffusion in manets. In *Proceedings of the 5th ACM international workshop on Modeling analysis and simulation of wireless and mobile systems*, pages 54–60. ACM Press, 2002.
33. Y. B. Ko and N. H. Vaidya. Flooding-based geocasting protocols for mobile ad hoc networks. *Mobile Networks and Applications*, 7(6):471–480, 2002.
34. B. Nath and D. Niculescu. Routing on a curve. In *HOTNETS-I*, 2002.
35. J.C. Navas and T. Imielinski. Geocast - geographic addressing and routing. In *Mobile Computing and Networking*, pages 66–76, 1997.
36. M. Papadopouli and H. Schulzrinne. A Performance Analysis of 7DS, A Peer-to-Peer Data Dissemination and Prefetching Tool for Mobile Users. In *Advances in wired and wireless communications, IEEE Sarnoff Symposium Digest*, March 2001.
37. S. Ratnasamy, B. Karp, Y. Li, F. Yu, R. Govindan, S. Shenker, and D. Estrin. GHT: A Geographic Hash Table for Data-Centric Storage. In *Proceedings of the First ACM International Workshop on Wireless Sensor Networks and Applications (WSNA 2002)*, Oct. 2002.
38. E. Royer and C. Toh. A review of current routing protocols for ad-hoc mobile wireless networks. *IEEE Personal Communications*, Apr. 1999.
39. Y. Sasson, D. Cavin, and A. Schiper. Probabilistic broadcast for flooding in wireless mobile ad hoc networks. In *IEEE Wireless Comm. and Networking Conference (WCNC 2003)*, 2003.
40. M.A. Spohn and J.J. Garcia-Luna-Aceves. Exploiting relative addressing and virtual overlays in ad hoc networks with bandwidth and processing constraints. In *ICWN*, 2003.
41. J. Tchakarov and N. Vaidya. Efficient content location in wireless ad hoc networks. In *IEEE International Conference on Mobile Data Management (MDM)*, 2004.
42. J.E. Wieselthier, G.D. Nguyen, and A. Ephremides. On the construction of energy-efficient broadcast and multicast trees in wireless networks. In *Infocom*, 2000.
43. B. Xu, A. Ouksel, and O. Wolfson. Opportunistic resource exchange in inter-vehicle ad-hoc networks. In *Mobile Data Management (MDM)*, 2004.
44. H. Zhou and S. Singh. Content based multicast (cbm) for ad hoc networks. In *Mobihoc*, 2000.