Two Hops or More: On Hop-Limited Search in Opportunistic Networks

Suzan Bayhan
University of Helsinki
Finland
bayhan@hiit.fi

Esa Hyytiä
Aalto University
Finland
esa@netlab.tkk.fi

Jussi Kangasharju
University of Helsinki
Finland
jakangas@cs.helsinki.fi

Jörg Ott
Aalto University
Finland
jo@netlab.tkk.fi

ABSTRACT

While there is a drastic shift from host-centric networking to content-centric networking, how to locate and retrieve the relevant content efficiently, especially in a mobile network, is still an open question. Mobile devices host increasing volume of data which could be shared with the nearby nodes in a multi-hop fashion. However, searching for content in this resource-restricted setting is not trivial due to the lack of a content index, as well as, desire for keeping the search cost low. In this paper, we analyze a lightweight search scheme, *hop-limited search*, that forwards the search messages only till a maximum number of hops, and requires no prior knowledge about the network. We highlight the effect of the hop limit on both search performance (i.e., success ratio and delay) and associated cost along with the interplay between content availability, tolerated waiting time, network density, and mobility. Our analysis, using the real mobility traces, as well as synthetic models, shows that the most substantial benefit is achieved at the first few hops and that after several hops the extra gain diminishes as a function of content availability and tolerated delay. We also observe that the *return path* taken by a response is on average longer than the *forward path* of the query and that the search cost increases only marginally after several hops due to the small network diameter.

Categories and Subject Descriptors

C.2.1 [**Computer-Communication Networks**]: Network Architecture and Design–Store and forward networks.

Keywords

Mobile opportunistic networks, mobile search, hop neighborhood, temporal distance.

1. INTRODUCTION

Mobile devices can establish opportunistic networks—a flavour of Delay-tolerant Networks (DTNs) [8]—among each other in a self-organising manner and facilitate communication without the need for network infrastructure. The capabilities of today's smart mobile devices yield substantial computing and storage resources and means for creating, viewing, archiving, manipulating, and sharing content. In addition, content downloaded from Internet is recommended to be cached at the handset [23]. This yields a huge reserve of data stored in mobile devices, which users may be willing to share. While this is typically done using Internet-based services, opportunistic networks enable content sharing among users in close proximity of each other.

In this paper, we focus on a human-centric DTN, where nodes search for information stored in some of the nodes. The nodes lack a global view of the network, i.e., there is no service that could index the stored content and assist a searching node in finding content (or indicate that the sought information does not exist). This means that operations are decentralized and, as mobile nodes have resource constraints (e.g., energy), we need to control the spread of the messages when searching. One such control mechanism is imposing the maximum number of hops a message can travel. We call a search scheme that limits the message's path to maximum h hops as h-hop search. *Hop-limited search* [16, 26] is of our interest as it is a lightweight scheme that does not require intensive information collection about the nodes or the content items. However, determining the optimal h is not straightforward. Although a large h tends to increase replication, it also increases the chance of finding the desirable target (content or searching node), thereby decreasing replication.

While many works in the literature apply hop limitations [5, 26], mostly two-hop such as [2], to the designed routing protocols, the motivation behind setting a particular hop value is not clear. In [4], we analyze the effect of hop limit on the routing performance by modelling the optimal hop limited routing as *all hops optimal path* problem [14]. However, opportunistic search necessitates considering the availability of the sought content as well as routing of the response to the searching node. In our previous work [16], we analytically modelled the search utility and derived the optimal hop count for a linear network, e.g., search flows through a single path and response messages follow the same path. In this paper, we provide an elaborate analysis of hop-limited search in a mobile opportunistic network considering the search success, completion time, and the cost.

Search is a two-phase process. We refer to the first phase in which query is routed towards the content providers as *forward path* and the second phase in which response is

University of Helsinki, Department of Computer Science, Aug.2015.
This work is accepted for publication in ACM MsWiM 2015.

routed towards the searching node as *return path*. First, we present an analysis on the forward path via an analytical model. We show the interplay between tolerated waiting time (how long the searching node can wait for the forward path), content availability, and the hop count providing the maximum search success ratio for a specific setting. Next, we verify our analysis of the forward path and elaborate also on the return path via simulations.

Our results suggest the following:

- Generally speaking, search performance increases with increasing h especially for scarcely available content. However, the highest improvement is often achieved at the second hop. After that, the improvements become smaller and practically diminish when $h > 5$.
- We observe that return path requires on average longer hops/time compared to the forward path, which may imply that optimal settings for the forward path does not yield the optimal performance on the return path.
- We show that the search cost first increases and after several hops ($h \approx 4$) it tends to stabilise. This is due to the small diameter of the network; even if h is large letting the nodes replicate a message to other nodes, each node gets informed about the search status quickly so that replication of obsolete messages is stopped.

The rest of the paper is organised as follows. Section 2 introduces the considered system model followed by Section 3 introducing the hop-limited search. Section 4 numerically analyzes the forward path, while Section 5 focuses on the whole search process. Section 6 overviews the related work and highlights the points that distinguish our work from the others. Finally, Section 7 concludes the paper.

2. SYSTEM MODEL

We consider a mobile opportunistic network of N nodes as in Fig. 1. Nodes move according to a mobility model which results in i.i.d. meeting rates between any two nodes. We use the following terminology:

Searching node (n_s) is a node that initiates the search for a content item. We assume that content items are equally likely to be sought, i.e., uniform *content popularity*. In reality, content items have diverse popularities; e.g., YouTube video popularity follows a Zipf distribution [12]. We choose uniform distribution to avoid a bias toward the "easy" searches.

Tagged node is a node that holds a copy of the sought content. Only a fraction α of the nodes are tagged, where α is referred to as the *content availability*. Although it is expected that search dynamics such as caching changes α, we assume that it does not change over time. The sets of all and tagged nodes are denoted by $\mathcal{N}$ and $\mathcal{M}$, and their sizes N and M, respectively. The number of tagged nodes is $M = \alpha N$. Every node is equally likely to be a tagged node.

Forward path and return path: In the first step of the search, a query is disseminated in the network to find a tagged node. In case the first step is completed, a response generated by a content provider is routed back towards n_s in the second step. We refer to the former as the *forward* or *query path* and to the latter as the *return* or *response path*.

Tolerated waiting time (T) is the maximum duration to find the content (excluding the return path). Note that total tolerated waiting time is $2T$ if we assume the same delay restriction for the return path.

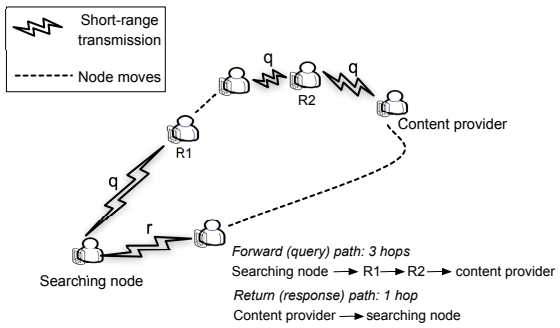


Figure 1: Network model, forward and return paths.

3. HOP-LIMITED SEARCH

To describe the basic operation of hop-limited search, assume that n_s creates a query that includes information about its search at time $t = 0$. When n_s encounters another node that does not have this query, n_s replicates the query to it to reach a tagged node faster. A node acquiring a copy of the message becomes a *discovered node*. The discovered node also starts to search and replicate the query to *undiscovered nodes*. Each message contains a header, referred to as *hop count*, representing the number of nodes that forwarded this message so far. Messages at n_s are 0-hop messages; those received directly from n_s are 1-hop messages. We refer to a search scheme as h-hop search if a search message can travel at most h hops. Hence, when a node has h-hop message, it does not forward it further. We also call a node with a i-hop message as the i-hop node for that particular message. In h-hop search, when an i-hop and j-hop node meet (without loss of generality we assume $i < j \leq h$), the state of j-hop node is updated to $(i+1)$-hop if $j > i+1$. The forward path completes when a copy of the query reaches a tagged node.

3.1 Search Success Ratio

Let us first consider a search scheme that does not put any hop limitation but instead limits the total number of replications on the forward path and the return path to K and K', respectively. For a content item with availability α, we can calculate the forward success ratio of this search scheme as:

$$\text{Forward success ratio} = 1 - (1-\alpha)^K. \tag{1}$$

Similarly, we calculate the search success ratio, i.e., both steps are completed, as:

$$P_s = \sum_{k=1}^{K} Pr\{\text{k content providers are discovered}\} \times Pr\{\text{at least one of k responses reaches } n_s\}.$$

We can expand the above formulation which leads to:

$$P_s = \sum_{k=1}^{K} \binom{K}{k} \alpha^k (1-\alpha)^{K-k} \left(1-(1-\frac{K'}{N-1})^k\right). \tag{2}$$

Let $\gamma = \frac{K'}{N-1}$, i.e., the probability that a response reaches n_s. After replacing γ into (2), we apply some manipulations by the help of binomial theorem: that is $(x+y)^n =$

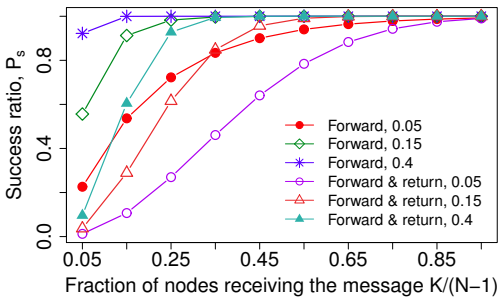


Figure 2: Search success with increasing degree of replication K for $\alpha = \{0.05, 0.15, 0.40\}$.

$\sum_{k=0}^{n} x^k y^{n-k}$. Then, we find the search success as:

$$P_s = 1 - (1 - \alpha\gamma)^K. \tag{3}$$

Please refer to Appendix A for the details of the above derivation.

Fig. 2 plots the success ratio with increasing fraction of nodes that receive the message. We plot the success of both the forward path and the total search under various content availability values. The results are for equal number of replications for the forward and return path, i.e., $K' = K$. The figure shows that to ensure a desirable level of success, search has to cover certain fraction of nodes, which depends on the content availability. In hop-limited search, there is no explicit restriction on number of replications K, but rather it is implicitly set by h and T. This result brings us to the question of how search coverage in the number of nodes changes with h and T.

Let $\mathcal{N}_h(t)$ be the set of discovered nodes excluding n_s at time t and $N_h(t) = |\mathcal{N}_h(t)|$ its size. Faloutsos *et al.* [9] define N_h for a static graph as node neighborhood within h hops. Similarly, we define $N_h(T)$ as the number of nodes that can be reached from a source node less than or equal to h hops under time limitation T. Let $P_h(T)$ denote the *forward path success ratio* defined as the probability that a query reaches one of the tagged nodes in a given time period T under h-hop limitation. For a search that seeks a content item with availability α, we approximate $P_h(T)$ as follows:

$$\begin{aligned} P_h(T) &= \sum_{n=1}^{N} P\{N_h(T) = n\} \cdot (1 - (1-\alpha)^n) \\ &= 1 - E[(1-\alpha)^{N_h(T)}] \\ &\approx 1 - (1-\alpha)^{E[N_h(T)]}. \end{aligned} \tag{4}$$

Note that (4) provides an upper bound for $P_h(T)$ and will only be used to understand the effect of h and T. In numerical evaluations, we relax all the simplifications (e.g., i.i.d meeting rates) and experiment using real mobility traces. Given (4), our problem reduces to discovering how $E[N_h(T)]$ changes with h and T. However, as observed in [11], each meeting may not lead to a new node being discovered, i.e., some nodes may meet again or a node may be met by several nodes. Therefore, total meetings in time period T results in a very optimistic estimate for $E[N_h(T)]$. In addition to the overlaps of opportunistic contacts, the hop restriction may result in lower $E[N_h(T)]$. In contrast to static networks [1], modelling $E[N_h(T)]$ for general time-evolving networks is not straightforward. Therefore, we derive $E[N_h(T)]$ from mobility traces and plug it into (4). See also [17] for an analysis of how mobility and node density affect the communication capacity of a mobile opportunistic network.

3.2 Benefit of One Additional Hop

We define the effect of increasing hop count from h to $h+1$ on the search performance as the *added benefit* and calculate it as the difference between $P_h(T)$ and $P_{h+1}(T)$: $\Delta P_{h,h+1} = P_{h+1}(T) - P_h(T)$, where $P_h(T)$ values are either approximated as above or obtained from experiments. Given that users can perceive only significant improvements rather than minimal changes in performance, we are more interested in larger values $\Delta P_{h,h+1}$. Let h_β denote the hop count beyond which the added benefit of one additional hop is lower than β. More formally, h_β is defined as:

$$h_\beta \triangleq \arg\max_h (\Delta P_{h,h+1} \geq \beta) + 1. \tag{5}$$

While $\beta \to 0$ yields the optimal hop count h^* achieving the highest performance, i.e., $h_\beta = h^*$, we search for h values that lead to higher β. Setting $\beta \to 0$, we assess until which hop there is still some gain, albeit minimal, whereas higher β values help discovering the last hop with substantial benefit. We refer to these β values as *any-benefit* and *fair-benefit* hop, respectively.

3.3 Time to Forward Path Completion

In this section, we derive the *forward path completion time* which is the time required for an initiated query to reach one of the content providers. Let us denote by $m_i(t)$ the total number of i-hop nodes at time t. These nodes have received the message via $(i-1)$ relays. Subsequently, we denote the state of a Continuous Time Markov Chain (CTMC) by:

$$S(t) = (m, m_0(t), m_1(t), \ldots, m_i(t), \ldots, m_{h-1}(t), m_h(t))$$

where

- $m = N_h(t) + 1$ is the number of nodes with a copy of the query (n_s and all discovered nodes) and
- $m_i(t)$ nodes are i-hop nodes and $\sum_{i=0}^{h} m_i(t) = m$.

States that have m nodes holding the search query are represented as S_m. The state space consists of all S_m transient states where $m \in \{1, \ldots, N-M\}$ and the absorbing state S_{tagged}, which represents discovery of a tagged node. Hereafter, we drop the time parameter for clarity. From state S_m, three types of events trigger state transitions:

- *Meeting a tagged node*: An arbitrary i-hop node in $\mathcal{N}_h$ where $i < h$ meets with one of the tagged nodes in $\mathcal{M}$. The resulting state is S_{tagged} and the search ends. There is only one such transition from every state.
- *Meeting a node that is undiscovered and not tagged*: An i-hop node meets an undiscovered untagged node. There are $\sum_{i=0}^{h-1} 1_{[m_i>0]}$ such transitions from this state where indicator function $1_{[f(\cdot)]}$ yields 1 if $f(\cdot)$ evaluates to *true*, and the resulting state is $S_{m+1} = (m+1, \ldots, m_{i+1}+1, \ldots)$.
- *Meeting among discovered nodes*: We call a meeting between i-hop and j-hop node an (i,j)-*meeting* where $i < j$. If $j > i+1$, N_h does not change but m_j and m_{i+1}

change. The new state is $S_m^{'} = (m, \ldots, m_{i+1}+1, \ldots, m_j - 1, \ldots)$. The number of such transitions from a state S_m is $\sum_{i=0}^{h-2} 1_{[m_i>0]} \left(\sum_{j=i+2}^{h} 1_{[m_j>0]} \right)$.

We call these three events *Type-tagged*, *Type-1*, and *Type-0* events; the corresponding transition rates are denoted by λ_{tagged}, λ_1, and λ_0, respectively. If the Type-1 event is a meeting with an i-hop node, we denote the respective rate as λ_1^i. Similarly, if a Type-0 event is due to an (i,j)-meeting, the rate is $\lambda_0^{i,j}$. For state $S_m = (m, m_0, m_1, \ldots, m_i, \ldots, m_h)$ the transitions leading to a state change are:

$$S_m \xrightarrow{\lambda_{\text{tagged}}} S_{\text{tagged}} \tag{6}$$

$$S_m \xrightarrow{\lambda_1^i} (m+1, \ldots, m_{i+1}+1, \ldots, m_h) \tag{7}$$

$$S_m \xrightarrow{\lambda_0^{i,j}} (\ldots, m_{i+1}+1, \ldots, m_j - 1, \ldots, m_h) \tag{8}$$

and the corresponding transition rates are:

$$\lambda_{\text{tagged}} = \lambda(m - m_h)M \tag{9}$$

$$\lambda_1^i = \lambda m_i (N - M - m) \tag{10}$$

$$\lambda_0^{i,j} = \lambda m_i m_j \tag{11}$$

where λ is the pairwise meeting rate.

Since solving the given Markov model may not be practical due to the state space explosion for large h and N, we approximate T_h and show its high accuracy in Section 4 by comparing it with the results of Markov model. Let $T(m,h)$ denote the remaining time to search completion under h-hop search and when m nodes hold the search query. Under h-hop search, only $m_{h^-} = m - m_h$ nodes are actively searching and can forward the query to their encounters. Assume that m_i are identical for all $i \geqslant 1$. Then, the number of searching nodes m_{h^-} is:

$$m_{h^-} = 1 + (m-1)(1 - \frac{1}{h}).$$

We calculate $T(m,h)$ as:

$$T(m,h) = \frac{1}{\lambda m_{h^-}} + \frac{\alpha}{\lambda} 0 + \frac{\lambda - \alpha}{\lambda} T(m+1,h). \tag{12}$$

We expand $T(m+1,h)$ similarly and substitute in (12). After a certain number of nodes are searching, the remaining time to search completion converges to zero, i.e., $T(m+k+1,h) \to 0$. Denote $q = (1 - \alpha/\lambda)$. Then, we find:

$$T(m,h) = \sum_{i=0}^{\infty} \frac{q^i}{\lambda(1 + (m+i-1)(1-h^{-1}))}$$

Solving for $m = 1$ gives an approximation for T_h:

$$\tilde{T_h} = \sum_{i=0}^{\infty} \frac{q^i}{\lambda(1 + i(1-h^{-1}))}. \tag{13}$$

4. FORWARD PATH

In this section, we evaluate the performance of hop-limited search (referred to as HOP) while varying (i) content availability, (ii) tolerated waiting time, (iii) network density, and (iv) mobility scenarios. We use $\alpha = \{0.40, 0.15, 0.05\}$ for *high*, *medium*, and *low* content availability, respectively.

4.1 Datasets

In our analysis, we use real traces of both humans and vehicles. For the former, we use the Infocom06 dataset [25] which represents the traces of human contacts during Infocom 2006 conference. For the latter, we use the Cabspotting dataset [21] that stores the GPS records of the cabs in San Francisco. To gain insights about more general network settings, we also analyze a synthetic mobility model that reflects realistic movement patterns in an urban scenario. Below, we overview the basic properties of each trace:

Infocom06: This data set records opportunistic Bluetooth contacts of 78 conference participants who were carrying iMotes and 20 static iMotes for the duration of the conference, i.e., approximately four days. In our analysis, we treated all devices as identical nodes which host content items and initiates search queries. This trace represents a small network in which people move in a closed region.

Cabspotting: This data set records the latitude and longitude information of a cab as well as its occupancy state and time stamp. The trace consists of updates of the 536 cabs moving in a large city area for a duration of 30 days. For our analysis, we focused only on the first three days and a small region of approximately 10 km×10 km area. 496 cabs appear in this region during the specific time period. The cab information is not updated at regular intervals. Hence, we interpolated the GPS data so as to have an update at every 10 s for each cab. Next, we set transmission range to 40 m to generate contacts among cabs. This trace represents an urban mobility scenario [15, 21].

Helsinki City Scenario (HCS): This setting represents an opportunistic human network in which the walking speed is uniformly distributed in [0.5,1.5] m/s. The nodes move in a closed area of 4.5 km×3.4 km. HCS [18] uses the downtown Helsinki map populated with *points-of-interests* (e.g., shops), between which pedestrians move along shortest path.

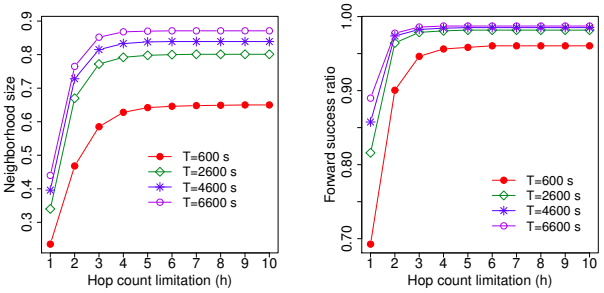
4.2 Forward Path Success Ratio

We derive the neighborhood size $N_h(T)$ as an average of 500 samples where each sample represents an independent observation of the network starting at some random time, from an arbitrary node and spanning an observation window equal to the tolerated waiting time. Next, we calculate $P_h(T)$ using (4). We use R [24] for these analysis. In the following, we mostly focus on the more challenging cases such as short T or low α, and report the representative results due to the space limitations.

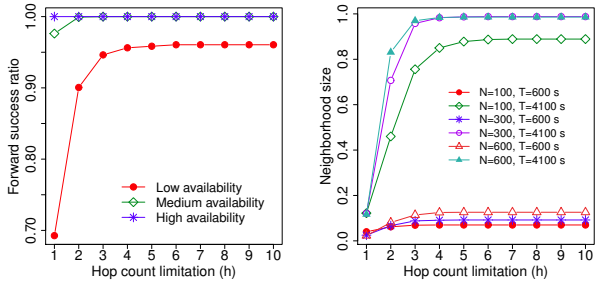
Fig. 3 illustrates the change in $N_h(T)$ represented as fraction of the network size for Infocom06 for various tolerated waiting time T and content availability α. For each T, we plot $N_h(T)$ and corresponding $P_h(T)$. From Fig. 3(a), we can see the significant growth of $N_h(T)$ at the second hop for all settings. While further hops introduce some improvements, we observe the existence of a *saturation point* [15]. After this point, the change in N_h is marginal either because all nodes are already covered in the neighborhood or higher h cannot help anymore without increasing T.

We present the resulting P_h for low content availability in Fig.3(b). As expected, the second hop provides the highest performance gain compared to the previous hop ($\Delta P_{h,h+1}$) as a reflection of highest $\Delta N_{h,h+1}$. As Infocom06 has good connectivity, almost all queries reach one of the tagged nodes after $h \approx 4$.

We present the effect of content availability for short T in Fig. 3(c). Regarding the effect of content availability, *we ob-*

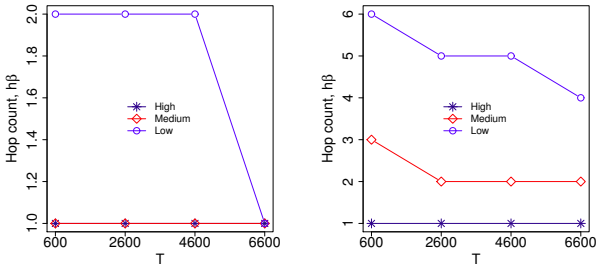


(a) Infocom06, $N_h(T)$. (b) Infocom06, $P_h(T)$.

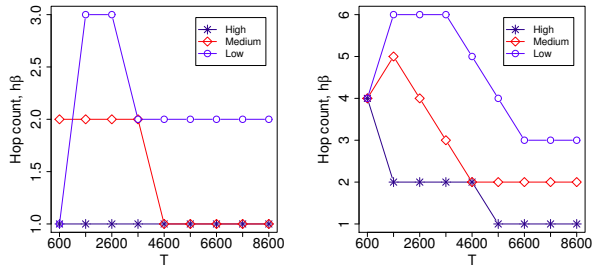


(c) Infocom06, $T = 600$ s. (d) HCS, $N_h(T)$.

Figure 3: (a) Growth of h-hop neighborhood for Infocom06 under various tolerated waiting time (T) and (b) corresponding forward success ratio for $\alpha = 0.05$. (c) Effect of content availability on $P_h(T)$ for Infocom06. (d) Growth of h-hop neighborhood for HCS under various N and T.



(a) Infocom06, fair-benefit. (b) Infocom06, any-benefit.



(c) HCS, fair-benefit. (d) HCS, any-benefit.

Figure 4: Fair- and any-benefit hops.

serve that search for a rare content item, i.e., low α, benefits more from increasing h compared to highly available content items. When many nodes are tagged, n_s meets one of these after some time. However, if the probability of meeting a tagged node is fairly low, using additional hops exploiting the mobility of the encountered nodes and spreading the query further is a better way of searching. A smart search algorithm can keep track of the content availability via message exchanges during encounters and can adjust the hop count depending on the observed availability of the content. For low and medium content availability, *the highest benefit is obtained at the second hop.* For a content item which is stored by a significant fraction of nodes, even a single hop search may retrieve the sought content.

Fig. 3(d) illustrates the growth of $N_h(T)$ under various network size N and T. We use our synthetic model HCS by setting $N = \{100, 300, 600\}$ nodes to observe the effect of network density on $N_h(T)$. For this particular setting, the clustering of lines based on T shows that time restriction is more dominant factor in determining $N_h(T)$ compared to N. As expected, $N_h(T)$ is higher for higher N. However, all settings exhibit the same growth trend with increasing h.

4.3 Fair-benefit and Any-benefit Hops

Fig. 4 illustrates the change in h_β with increasing T for $\beta \in \{0.0005, 0.1\}$ which represent any-benefit and fair-benefit hops, respectively. As stated before, Infocom06 has good connectivity among nodes as conference takes place in a closed area and nodes share the same schedule (e.g., coffee breaks and sessions). As a result of this good connectivity, first one or two hops provide almost all the benefits of multi-hop search (Fig. 4(a)). However, Fig.4(b) shows that the optimal hop in terms of the highest P_h is achieved at higher $h \leqslant 6$.

For HCS scenario that represents a network of urban scale, the effect of increasing T is a bit different. Fig. 4(c) shows that short and longer T may have the same operating point in terms of h_β. For low T, the benefit of increasing h diminishes due to the limited number of encounters and the resulting small set of discovered nodes, whereas for long T almost all nodes are discovered without requiring any further hops. The lower h_β with increasing T can also be explained by "*shrinking diameter*" phenomenon, which states that the average distance between two nodes decreases over time as networks grow [19]. Indeed, the diameter of the message search tree gets shorter over time and leads to a smaller hop distance between the searching node and a tagged node. Note the decreasing *any-benefit* hop in Fig. 4(d). For example, searching for a content item with medium availability achieves the highest performance at $h_\beta = 5$ for $T = 1600$ s, while it decreases gradually to $h_\beta = 2$ with increasing T.

4.4 Time to Forward Path Completion

Table 1 shows the search time T_h which is obtained by solving CTMC introduced in Section 3.3 and $\tilde{T}_h$ given by (13). We normalize every value by the maximum search time of each setting. In the table, we also list the approximation errors. First thing to note is the drastic decrease in search time at the second hop. In agreement with our conclusions from P_h, we see that second hop speeds the search significantly resulting in approximately 80% shorter search time for the low availability, 60% for the medium, and 50% decrease for the high availability scenario. As P_h, the most significant improvement in search time occurs for low con-

tent availability. Our approximation also exhibits exactly the same behaviour in terms of the change in the search time. As the error row shows, it deviates from the expected search time to some degree: 32% under-estimation to 8% overestimation.

Table 1: Search time and its approximation.

α	Time	$h=1$	$h=2$	$h=3$	$h=4$	$h=5$
Low	T_h	1	0.21	0.13	0.12	0.11
	$\tilde{T}_h$	1	0.14	0.12	0.11	0.11
	Error	0	-0.32	-0.08	-0.03	-0.03
Medium	T_h	1	0.40	0.33	0.31	0.31
	$\tilde{T}_h$	1	0.39	0.35	0.33	0.33
	Error	0	-0.01	0.07	0.06	0.05
High	T_h	1	0.51	0.46	0.45	0.45
	$\tilde{T}_h$	1	0.54	0.49	0.47	0.46
	Error	0	0.05	0.08	0.06	0.04

5. COMPLETE SEARCH

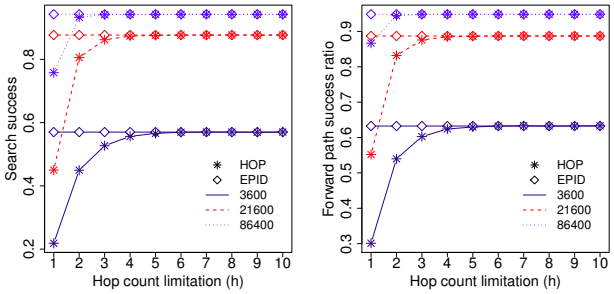
In Sec.4, we show that the first few hops yield the highest benefit in terms of forward path success ratio. In this section, we explore if this trend holds for the whole search, i.e., forward and return path. To this end, we carry out a wide range of simulations to gain insight to the following aspects: (i) the fair-benefit hops for various T and α settings, (ii) average temporal and hop distance to/from content, (iii) characteristics of the return path, and (iv) search cost.

Note that return path also applies hop-limited routing as well as the forward path. We first assume that an oracle stops the dissemination of a completed search immediately, and then relax this assumption in a scheme where nodes are informed about search state via control messages. Using the ONE simulator [18], we design an opportunistic network consisting of $N = 98$ nodes for Infocom06 and $N = 496$ for Cabspotting scenario. 5000 content items are distributed across nodes according to the availability ratios: for example, a content item with $\alpha = 0.4$ is randomly assigned to 40% of the nodes. This assignment is static during the course of the simulation: nodes do not generate new content items and nodes having received a copy during operation will not respond to requests. Every query interval, a random node initiates a query for a uniformly chosen content item. The query generation interval is uniformly distributed in [10,20] s. We simulate the complete system during which approximately 23000 queries are generated.

We use 20 m radio range for Infocom06 and 40 m for Cabspotting, infinite transmission capacity (i.e., all packets can be exchanged at an encounter), and 100 MB of storage capacity; message are 15 KB in size. To obtain an upper performance bound, we implement *epidemic search* (EPID) which does not impose any hop limitations. In the following, we refer to h-hop search as HOP.

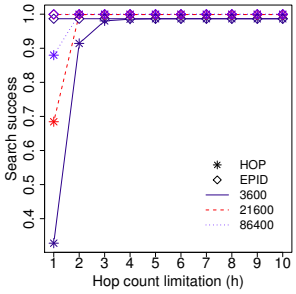
5.1 Fair-benefit and Any-benefit Hops

We analyse the effect of h on the search success ratio (P_s, the ratio of queries retrieving a response) and forward path success ratio (P_h). Fig. 5 depicts P_s and P_h for HOP and EPID under various settings: $T = \{3600, 21600, 86400\}$ s, all availabilities, and for Infocom06 and Cabspotting scenarios. Total tolerated search time (i.e., forward and the return path) is set to $2T$.

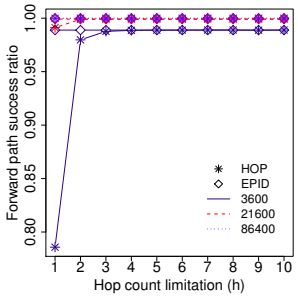


(a) P_s, Infocom06.

(b) P_h, Infocom06.



(c) P_s, Cabspotting.



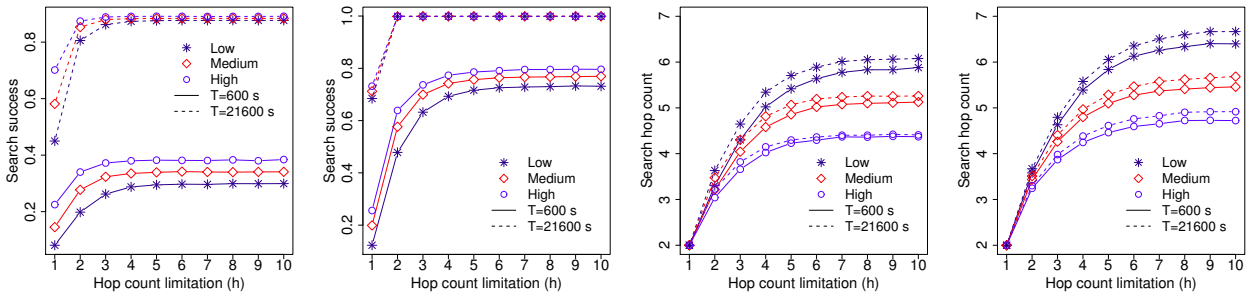
(d) P_h, Cabspotting.

Figure 5: Effect of tolerated waiting time, $\alpha = 0.05$.

As Fig. 5 shows, the highest $\Delta P_{h,h+1}$ is at the second hop for all T. However, it takes more hops until HOP achieves the same success ratio as EPID: $h = 5$ for $T = 3600$ s, $h = 4$ for $T = 21600$ s, and $h = 3$ for $T = 86400$ s. Cabspotting scenario achieves higher success ratio under the same time limitations compared with Infocom06. Although Infocom06 represents a small area and closed group scenario, we observe more contacts in Cabspotting setting. This may be attributed to higher transmission range as well as higher mobility of the nodes.

If we consider only the forward path as we did previously in Fig. 3(b), we observe in Fig 5(b) that the second hop provides almost all the benefits for longer waiting times. For more strict T, we see that the search success ratio is below 1 meaning that all benefits of multi-hop routing is exploited. Under this setting, the search performance cannot be improved without increasing T. Having visited P_s and P_h figures for Infocom06, we conclude that although the target content is discovered at the second hop with high probability, it does not necessarily ensure that the return path completes in two hops. As such, return path is similar to a search for the content with $\alpha = 1/N$, i.e., a scarce content item. Hence, it may require further hops. For Cabspotting, two to three hops yield the same performance as EPID.

Figs. 6(a) and 6(b) demonstrate the effect of content availability on the whole search success for $T = 600$ s and $T = 21600$ s. As we see in the figures, P_s values are clustered based on T for both Infocom06 and Cabspotting. This behaviour can be interpreted as the tolerated waiting time being more dominant factor in determining the search performance rather than the content availability under the considered settings. Regarding h_β, fair-benefit hops are around 2-3 hops whereas any-benefit hops vary from 2 to 6 hops. Finally, Figs. 6(c) and 6(d) demonstrate the total hop count



(a) P_s, Infocom06. (b) P_s, Cabspotting. (c) Hop count, Infocom06. (d) Hop count, Cabspotting.

Figure 6: Effect of content availability under $T = 600$ and $T = 21600$ s.

for the whole search route. In contrary to P_s, we observe a clustering according to content availability in the total search hop counts. While the tolerated waiting time has some effect on the average search path length, the effect of content availability is the dominating factor. Low availability content items are retrieved from more far-away nodes whereas items with higher availability are retrieved from closer nodes. Nevertheless, the total path lengths are much shorter than the imposed limit, which means that content is typically retrieved from "nearby" nodes.

5.2 Effective Distance to/from Content

In the previous section, we presented experiment results for different tolerated waiting time settings and referred to them as *short* or *long* T. In fact, temporal dynamics of the network (e.g, average inter-contact time) determine whether a T value is short or not. In this section, we aim to provide insights about the distance to content both in terms of number of hops and time. To this end, we focus on the forward paths under EPID and remove the time restriction (i.e., $T = \infty$). In general, the performance of an opportunistic network is governed by the routing protocol and the node mobility (among other factors). We eliminate the effect of the algorithm by using EPID. This allows us finding the average hop distance between a searching node and a tagged node. Similarly, as we avoid any routing effects, we obtain the shortest time it takes to reach a tagged node from a searching node. We refer to this time as *temporal distance* to content. Let us define *effective distance to content* as the maximum distance, be it hops or time, which ensures that 90% of the paths between a searching node and a tagged node is lower than this distance [19]. We define *effective distance from content* similarly for the return path.

In Fig. 7, we plot the effective hop and temporal distances for both the forward and return paths under different content availabilities for Infocom06 and Cabspotting. While the effective distance to content (i.e., forward path length) is inversely proportional to content availability, the effective distance from content (i.e., return path length) is only slightly affected by content availability. It is also noteworthy that the response hop distance is longer than the query distance for all settings. Regarding temporal distances in Fig. 7 (second row), Cabspotting has much shorter temporal distance compared to Infocom06. This difference explains higher success ratios achieved in Cabspotting compared with Infocom06 under the same T in Fig. 5(c) vs. Fig. 5(a). In Fig. 7 (bottom-left), effective temporal distance is shorter

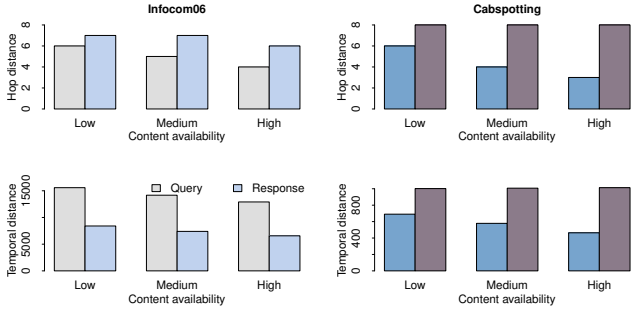


Figure 7: Effective hop and temporal distance for query and return path for Infocom06 and Cabspotting. Left bars: query, right bars: response.

for the responses as opposed to higher hop counts observed in Fig. 7 (top-left). However, as we do not preserve the coupling between the query and response paths of the search, this result does not necessarily contradict with our claim that response path takes longer and is more challenging. In fact, responses that are still looking for the searching node are not accounted for, which may in turn lead to the shorter effective distance. Effective temporal distance is paramount for unveiling the conditions of feasibility of search. For example, setting $T = 600$ s lead to very poor search performance for Infocom06 as shown in Fig. 6(a) because the network evolves slower than this time.

5.3 Characteristics of the Return Path

In the previous sections, we analyzed the query and response paths separately. One arising question is the relation between the forward and return paths: whether the latter depends on the former. In this section, we present the analysis of the query and response paths where query and response path coupling is preserved. We simulate EPID with a restriction on the search time. Using the completed queries, we calculate the Pearson correlation coefficient between the query and response hop (ρ_{hop}) as well as the query and response time (ρ_{temp}). We also find the ratio of the response hop to the query hop (γ_{hop}) as well as the ratio of response temporal path length to query temporal path length (γ_{temp}) to explore how challenging the return path is compared to the forward path.

Table 2: Correlation between forward and return paths, Infocom06.

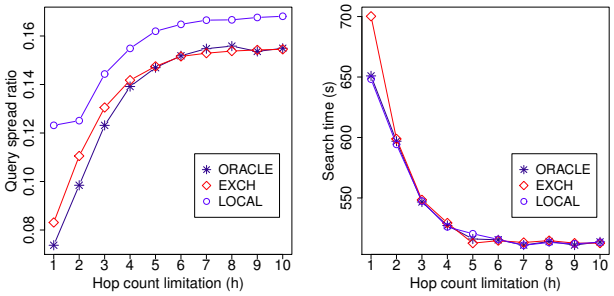
T	α	ρ_{hop}	ρ_{temp}	γ_{hop}	γ_{temp}	P_h	P_s
600	Low	0.30	0.36	1.47	2.23	0.35	0.30
	Med.	0.29	0.34	1.72	3.09	0.42	0.34
	High	0.27	0.32	1.97	4.02	0.48	0.38
3600	Low	0.35	0.43	1.4	2.18	0.63	0.57
	Med.	0.35	0.38	1.61	2.98	0.67	0.61
	High	0.33	0.32	1.85	4.13	0.70	0.65
86400	Low	0.33	0.13	1.39	2.60	0.95	0.94
	Med.	0.35	0.13	1.62	3.52	0.95	0.94
	High	0.35	0.12	1.86	4.50	0.95	0.95

Table 2 summarises this analysis for Infocom06, which also agrees with the analysis for Cabspotting scenario. First, we do not observe any strong correlation between return and forward path lengths for any of the settings (i.e., ρ_{hop} and ρ_{temp} are around 0.1–0.4). This result may be conflicting with the intuition that the information found in nearby/far-away will also be routed back quickly/slowly. However, search process is more complicated due to the intertwined effects of mobility and restrictions on the total search time. For example, consider a forward path with a very large number of hops. Due to the remaining short time before the tolerated waiting time expires, search can only go a few hops towards the searching node. This leads to a long forward path with a very short return path which challenges the above-mentioned intuition. With higher content availability, the required number of forward hops decreases whereas the return path length seems to be barely affected. Hence, the corresponding γ_{hop} and γ_{temp} increase. For all scenarios, return path is on the average longer than the forward path. Using this observation, a tagged node receiving a query can set the time-to-live field of its response message longer than the received query's forward path time.

5.4 Cost of Search

Obviously, increasing hop count increases the neighborhood which in turn increases the chance of finding the sought content. However, the larger neighborhood should also be interpreted as larger number of replication, i.e., higher search cost. In fact, the neighborhood size represents the upper bound of number of replications for a search message if no other search stopping algorithm is in effect. In this section, we evaluate the cost of search considering two simple mechanisms that aim to keep replication much below the upper bound set by $N_h(T)$.

We consider three cases: (i) ORACLE: there is a central entity from which a node retrieves the global state of a message in its buffer and can ignore it if outdated, e.g. a completed search or a query already reaching a tagged node, (ii) EXCH: upon encounter, nodes exchange their local knowledge about search activities in the network, (iii) LOCAL: nodes exchange their knowledge *only* on the shared messages. Note that an ORACLE can be an entity in the cellular network and nodes can access it via a control channel. In fact, this communication with the cellular network is more costly (e.g., energy) compared to the opportunistic communication. Nevertheless, this scenario serves as an optimal benchmark to assess the performance of the other scenarios. EXCH pro-actively spreads information about existing queries to all nodes opportunistically, which may be considered as leaking information to nodes not involved in search. LOCAL circumvents this by only using its local knowledge and sharing information with peers only about queries that the other node has already seen.[1]



(a) Query spread ratio. (b) Search time.

Figure 8: Hop count vs. search cost and time for Infocom06 (α=0.15, T =600 s).

Fig. 8 shows *query spread ratio* which is defined as i.e., the fraction of nodes that have seen this query, and search time. First, we should note that the resulting search success (not plotted) is almost the same under all schemes. Second, note the non-increasing query spread ratio for $h > 5$. This result confirms that the network is a small-world network where all nodes get informed quickly about the search status and drop the outdated messages timely. Hence, even for larger h, nodes detect the outdated messages via local and shared knowledge. In Fig. 8(a), we observe that EXCH maintains the same performance as ORACLE, whereas LOCAL results in more replication as fewer nodes are informed about the completed queries/responses. Nevertheless, because of the small network diameter, a higher h does not result in an explosion of query spread in the network. Regarding search time, we observe substantial decrease in search time also for $h > 2$ and $h < 5$ in contrast to the vanishing benefits after second hop derived from our analytical model (Table 1). Search time tends to stabilise after $h \geqslant 5$. Hence, although several hops are sufficient in terms of search success, further hops (e.g., $h \approx 4$) can be considered for faster search.

5.5 Discussion

Our analysis shows that the two factors affecting the optimal hop count are the content availability (α) and the product of meeting rate and tolerated waiting time (λT). The latter is the number of meetings before tolerated waiting time expires where λ depends on network density and mobility model. If both α and λT are low (scarce content and very few contacts due to network sparsity or short search time), search performance is expected to be low. However, it increases with increase in either of these factors. If one of these factors is large, it is sufficient to have a low hop count limit (e.g., two or three) to obtain good performance. That is, when $\alpha\lambda T$ is large relative to the expected number of tagged nodes met during the search time, limiting the search to a few hops still achieves good results. As $\alpha\lambda T$

[1]This is easy to implement even without cryptographic methods by using a two-stage protocol, in which nodes first forward the queries they are carrying and then share information about those queries they received from their peer.

decreases, the required hop limit to maintain good search performance is larger. This allows devising adaptation when issuing search queries. Nodes can monitor the request and response rate for certain (types of) content items and thus infer popularity and availability in their area.[2] They can also assess the regional node density and meeting rate λ and thus determine the required hop count h given T or vice versa. Moreover, nodes can monitor search performance and determine how well their (region in a) network operates. To improve performance, nodes can decide to increase the availability of selected contents via active replication of the scarce content, obviously trading off storage capacity and link capacity for availability. Such decision could be based entirely upon local observations, but could also consider limited information exchange with other nodes.

We believe that with the guidance of our analysis, a node can decide on the best hop count depending on the network density and the content availability that can both be derived from past observations by the node. Although we show that search cost stops increasing after a few hops, keeping h low may be desirable if we interpret it as a measure of the social relation between two nodes (e.g., one hop as friends; two hop as friend-of-friend). In other words, lower h can be interpreted as more *trustworthy* operation. Moreover, protocols involving lower number of relays are more scalable and energy-efficient [10].

6. RELATED WORK

While the DTN literature has many proposals for message dissemination, which exploit the information about the network such as (estimated) pairwise node contact rates, node centrality, communities, and social ties, efficient mechanisms for content search remain largely unexplored. In a sense, this is reasonable as search can be considered as a two-step message delivery: query routing on the forward path and the response routing on the return path. The forward path is less certain as the content providers are unknown. In this regard, the return path is less challenging as the target node (and a recent path to reach it) is already known. However, routing the response looks for a particular node, whereas the forward search is for a *subset of nodes* whose cardinality is proportional to the content availability. Thus, search requires special treatment rather than being an extension of the message dissemination.

Two questions for an efficient search are (i) which nodes may have the content [3] and (ii) when to stop a search [22]. The former question requires assessment of each node in terms of its potential of being a provider for the sought content. For example, *seeker assisted search* (SAS) [3] estimates the nodes in the same community to have higher likelihood of holding the content as people in the same community might have already retrieved the content. Given that people sharing a common interest come together at a certain "space" (e.g., office, gyms), [10] defines *geo-community* concept and matches each query with a particular geo-community. Hence, the first question boils down to selecting relays with high probability of visiting the target geo-community. As [10] aims to keep the search cost minimal, searching node employs two-hop routing and determines the relays, which thereby reduces the issue of when to stop the search. Deciding when to stop search is nontrivial as the search follows several paths and whether the searching node has already discovered a response is not known by the relaying nodes. Hyytiä *et al.* [16] model the expected search utility with increasing hop count and then finds the optimal hop, while [22] estimates the number of nodes having received a query and possible responses by using the node degrees. In our work, we showed that simple schemes via information sharing can stop search timely and maintain similar performance to that of an oracle due to the small world nature of the studied networks.

Different from all these above works, main focus of our paper is more on a fundamental question: *how does the hop limitation affect the search performance?* While this question has been explored in general networking context, content-centricity and the time constraints require a better understanding of flooding in the context of search. Therefore, we first provided insights on search on a simplified setting and next analyzed the effect of various parameters, e.g., time and real mobility traces, via extensive simulations.

In the literature, several works focus on two-hop forwarding in which the source node replicates the message to any relay and the relays can deliver the message only to the final destination [2, 6, 13]. Grossglauser *et al.* in their seminal work [13] show that two hops are sufficient to achieve the maximum throughput capacity in an ideal network with nodes moving randomly. The capacity increase is facilitated by the reduced interference on the links from source to relay and relay to the destination. Chaintreau *et al.* [6] assess this two-hop forwarding scheme in a DTN scenario with power-law inter-contact times and employ an *oblivious forwarding algorithm* (e.g., memoryless routers that do not use any context information such as contact history). Similarly, [5] focuses on a DTN with power-law distributed inter-contact times and derives the conditions (i.e. range of Pareto shape parameter) under which message delivery has a finite delay bound for both two-hop and multi-hop oblivious algorithms. The authors show that "*as long as the convergence of message delivery delay is the only concern, paths longer than two-hops do not help convergence*" as two hops are sufficient to explore the relaying diversity of the network [5]. Another work supporting two-hop schemes is [10] which shows that two-hop search is favourable for opportunistic networks – a resource-scarce setting, as "*one-hop neighbors are able to cover the most of the network in a reasonable time*" in a network with *sufficiently many* mobile nodes. In our work, we include those cases when longer paths still yield (some) performance improvement. Finally, our work supports the conclusion of [7], which theoretically proves the *small-world* in human mobile networks.

Our work is closely related to the k-hop flooding [26] which models the spread of flooded messages in a random graph. Unlike [26], we focus on content search in a realistic setting, and using real mobility traces (both a human contact network and a vehicular network) we provide insights on the effect of increasing hop count on the search success, delay, and cost under various content availability and tolerated waiting time settings. Despite the differences, it is worthwhile noting that our results agree with the basic conclusions of [26].

7. CONCLUSIONS

Given the volume of the content created, downloaded, and stored in the mobile devices, efficient opportunistic search is paramount to make the remote content accessible to a mo-

[2]If nodes cache response contents opportunistically [20], availability would grow with popularity.

bile user. Current schemes mostly rely on routing based on hop limitations. To provide insights about the basics of such a generic search scheme, we focused on a hop-limited search in mobile opportunistic networks. First, by modelling only the forward path, we showed that (i) the second hop and following few hops bring the highest gains in terms of forward path success ratio and (ii) compared to single-hop delivery, increase in hop count leads to shorter search time and after a few hops search time tends to stabilize. Next, we revisited these findings via simulations of the entire search process. While simulations validated our claim for the forward path, we observed that return path on average requires longer time and more hops. Moreover, our results do not indicate strong correlation between the return and forward paths. Finally, we showed that search completes in less than five hops in most cases. This is attributed to the small diameter of the human contact network which has also a positive impact on search cost; nodes are informed about search state quickly and stop propagation of obsolete messages. Our simulations validated that increasing hop count to several hops accelerates the search and later search completion time stabilizes.

As future work, we will design a search scheme that adapts the hop count of query and response paths based on the observed content availability and popularity. Moreover, we believe that content-centric approaches should be paid more attention to implement efficient search schemes in mobile opportunistic networks.

Acknowledgements

This work was supported by the Academy of Finland in the PDP project (grant no. 260014).

8. REFERENCES

[1] L. A. Adamic, R. M. Lukose, and B. A. Huberman. Local search in unstructured networks. *Handbook of graphs and networks*, 2006.

[2] A. Al Hanbali, P. Nain, and E. Altman. Performance of ad hoc networks with two-hop relay routing and limited packet lifetime. *Performance Evaluation*, 65(6):463–483, 2008.

[3] S. Bayhan, E. Hyytiä, J. Kangasharju, and J. Ott. Seeker-assisted information search in mobile clouds. In *ACM SIGCOMM workshop on Mobile cloud computing*, 2013.

[4] S. Bayhan, E. Hyytiä, J. Kangasharju, and J. Ott. Analysis of hop limit in opportunistic networks by static and time-aggregated graphs. In *IEEE ICC*, 2015.

[5] C. Boldrini, M. Conti, and A. Passarella. Less is more: long paths do not help the convergence of social-oblivious forwarding in opportunistic networks. In *ACM Int. workshop on Mobile Opportunistic Networks*, 2012.

[6] A. Chaintreau, P. Hui, J. Crowcroft, C. Diot, R. Gass, and J. Scott. Impact of human mobility on opportunistic forwarding algorithms. *IEEE Trans. on Mobile Comp.*, 6(6):606–620, 2007.

[7] A. Chaintreau, A. Mtibaa, L. Massoulie, and C. Diot. The diameter of opportunistic mobile networks. In *ACM CoNEXT*, 2007.

[8] K. Fall. A delay-tolerant network architecture for challenged internets. In *ACM SIGCOMM*, 2003.

[9] M. Faloutsos, P. Faloutsos, and C. Faloutsos. On power-law relationships of the Internet topology. In *ACM SIGCOMM Computer Communication Review*, volume 29, pages 251–262, 1999.

[10] J. Fan, J. Chen, Y. Du, P. Wang, and Y. Sun. Delque: A socially aware delegation query scheme in delay-tolerant networks. *IEEE Transactions on Vehicular Technology*, 60(5):2181–2193, Jun 2011.

[11] W. Gao, Q. Li, and G. Cao. Forwarding redundancy in opportunistic mobile networks: Investigation, elimination, and exploitation. *IEEE Transactions on Mobile Computing*, 2014.

[12] P. Gill, M. Arlitt, Z. Li, and A. Mahanti. Youtube traffic characterization: a view from the edge. In *ACM SIGCOMM Conf. on Internet measurement*, 2007.

[13] M. Grossglauser and D. Tse. Mobility increases the capacity of ad hoc wireless networks. *IEEE/ACM Trans. on Networking*, 10(4):477–486, Aug 2002.

[14] R. Guérin and A. Orda. Computing shortest paths for any number of hops. *IEEE/ACM Transactions on Networking (TON)*, 10(5):613–620, 2002.

[15] M. A. Hoque, X. Hong, and B. Dixon. Efficient multi-hop connectivity analysis in urban vehicular networks. *Vehicular Comm.*, 1(2):78–90, April 2014.

[16] E. Hyytiä, S. Bayhan, J. Ott, and J. Kangasharju. Searching a needle in (linear) opportunistic networks. In *ACM MSWiM*, 2014.

[17] E. Hyytia and J. Ott. Criticality of large delay tolerant networks via directed continuum percolation in space-time. In *IEEE INFOCOM*, 2013.

[18] A. Keränen, J. Ott, and T. Kärkkäinen. The ONE Simulator for DTN Protocol Evaluation. In *Proc. of Int. Conf. on Simulation Tools and Techniques*, 2009.

[19] J. Leskovec, J. Kleinberg, and C. Faloutsos. Graphs over time: densification laws, shrinking diameters and possible explanations. In *ACM KDD*, 2005.

[20] J. Ott and M. Pitkänen. DTN-based Content Storage and Retrieval. In *The First IEEE WoWMoM Workshop on Autonomic and Opportunistic Communications (AOC)*, June 2007.

[21] M. Piorkowski, N. Sarafijanovic-Djukic, and M. Grossglauser. CRAWDAD data set epfl/mobility (v. 2009-02-24), Feb. 2009. Downloaded from `http://crawdad.org/epfl/mobility/`.

[22] M. Pitkänen, T. Karkkainen, J. Greifenberg, and J. Ott. Searching for content in mobile DTNs. In *IEEE PerCom*, 2009.

[23] F. Qian, K. S. Quah, J. Huang, J. Erman, A. Gerber, Z. Mao, S. Sen, and O. Spatscheck. Web caching on smartphones: ideal vs. reality. In *ACM MobiSys*, 2012.

[24] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2013.

[25] J. Scott, R. Gass, J. Crowcroft, P. Hui, C. Diot, and A. Chaintreau. CRAWDAD data set cambridge/haggle (v. 2006-01-31). `http://crawdad.org/cambridge/haggle/`, Jan. 2006.

[26] M. Vojnovic and A. Proutiere. Hop limited flooding over dynamic networks. In *IEEE INFOCOM*, 2011.

APPENDIX

A. DERIVATION OF SEARCH SUCCESS

Given that the message is received by K nodes and each node has probability α of holding the requested content, the probability that an initiated query reaches one of the content provider(s) equals to the probability that at least one of the K nodes has the content. We denote then the forward success ratio as:

$$\text{Forward success ratio} = 1 - (1-\alpha)^K. \tag{14}$$

If we consider the whole search path with the assumption that K nodes at maximum holds the query and only $K^{'}$ copies are allowed for the response message, we calculate the search success ratio, i.e., both steps are completed, as:

$$P_s = \sum_{k=1}^{K} Pr\{\text{k content providers are discovered}\} \times Pr\{\text{at least one of k responses reaches } n_s\}. \tag{15}$$

The first factor in (15), corresponding to the event that k content providers are discovered, obeys Binomial distribution with parameters (K, α). The second factor is conditioned on k, the number of content providers reached by the query replicas. Out of these k responses, we calculate the probability that at least one of them reaches the destination, i.e., the searching node. With the assumption that $K^{'}$ copies of the response is allowed on the return path, to calculate the probability of finding n_s, we find the probability that none of the k responses reaches n_s. Since there are $(N-1)$ nodes a response message created by a particular content provider can reach and $K^{'}$ selection without replication, the probability that a response reaches n_s equals to:

$$\gamma = \frac{K'}{N-1}.$$

Substituting these formulations into (15), we find:

$$P_s = \sum_{k=1}^{K} \binom{K}{k} \alpha^k (1-\alpha)^{K-k} \left(1-(1-\gamma)^k\right). \tag{16}$$

Separating (16) into two parts, we find:

$$P_s = \sum_{k=1}^{K} \binom{K}{k} \alpha^k (1-\alpha)^{K-k} - \sum_{k=1}^{K} \binom{K}{k} \alpha^k (1-\alpha)^{K-k} (1-\gamma)^k. \tag{17}$$

Next, we notice that each summation term above can be compactly written by the help of the binomial theorem, which is:

$$(x+y)^n = \sum_{k=0}^{n} x^k y^{n-k}.$$

Taking into account that the summation in (17) starts from 1, we simplify the first term in (17) as follows:

$$\sum_{k=1}^{K} \binom{K}{k} \alpha^k (1-\alpha)^{K-k} = (\alpha + 1 - \alpha)^K - \binom{K}{0} \alpha^0 (1-\alpha)^{K-0} = 1 - (1-\alpha)^K. \tag{18}$$

Applying the same expansion for the second term in (17), we find:

$$\sum_{k=1}^{K} \binom{K}{k} \alpha^k (1-\alpha)^{K-k} (1-\gamma)^k = \sum_{k=1}^{K} \binom{K}{k} (\alpha - \alpha\gamma)^k (1-\alpha)^{K-k}$$
$$= (\alpha - \alpha\gamma + 1 - \alpha)^K - \binom{K}{0} (\alpha - \alpha\gamma)^0 (1-\alpha)^{K-0}$$
$$= (1-\alpha\gamma)^K - (1-\alpha)^K. \tag{19}$$

Substituting (18) and (19) into (17), we find:

$$P_s = 1 - (1-\alpha)^K - \left((1-\alpha\gamma)^K - (1-\alpha)^K\right),$$

which gives us the search success probability:

$$P_s = 1 - (1-\alpha\gamma)^K. \tag{20}$$