Available online at www.sciencedirect.com
SCIENCE DIRECT
Theoretical Computer Science 355 (2006) 25–36
Theoretical Computer Science
www.elsevier.com/locate/tcs

# Broadcasting in unstructured peer-to-peer overlay networks

Fred S. Annexstein*, Kenneth A. Berman, Mijhalo A. Jovanović

*Department of ECECS, University of Cincinnati, Cincinnati, OH 45221, USA*

**Abstract**

Peer-to-peer overlay networks present new opportunities and challenges for achieving enhanced network functionality at the application level. In this paper we study the impact of point-to-point network latency on flooding broadcast operations in peer-to-peer overlay networks. We show that two standard protocol mechanisms, used to control the amount of network resources used during flooding, can in combination, significantly reduce the reach of broadcast messages. We prove that these standard mechanisms, known as "time-to-live bounds" and "unique message identification", can result in broadcast operations that only reach a vanishing fraction of the nodes. In addition, we provide empirical evidence that the trend suggested by our formal results are found in data obtained from the Gnutella network and through network simulations.


*Keywords:* Peer-to-peer overlay networks; Network broadcasting and flooding; Distributed network algorithms; Network reliability and reachability; Network simulation

## 1. Introduction

Internet application-level overlay networks present opportunities and challenges for achieving enhanced network functionality. In this paper we consider the problem of analyzing broadcast operations over unstructured peer-to-peer (P2P) overlay networks. In recent years a number of popular P2P applications (e.g., Gnutella, Kazaa, Bittorrent) have been built in an unstructured way in which connectivity is achieved principally through random connections. One fundamental operation required of many unstructured networks is to enable simple broadcast operations in which each peer can communicate messages that can reach large numbers of other peer nodes. The popular Gnutella network [13,12] uses message flooding as the principal search mechanism in an effort to maximize the coverage or reachability of search query messages. A body of recent research work suggests that hybrid approaches which combine flooding with some hierarchical structure and random walks may be optimal for the design of large-scale P2P search applications [9,10,3,4].

Nearly all dynamic networks must at some level implement flooding operations, since there is generally no other way to discover network linkages and update topology state information. To control the impact of flooding messages across the network, standard mechanisms are generally used to locally terminate the flood. Examples of such mechanisms include age fields, hop bounds, sequence numbers, and periodic updates. The reader is referred to the text [1] for classical results concerning tradeoffs and theoretical issues regarding these mechanisms. In more recent work flooding control mechanisms have been studied in the context of mobile ad hoc networks [15,8].

* Corresponding author. Tel.: +1 513 556 1807; fax: +1 513 556 7326.
*E-mail address:* fred.annexstein@uc.edu (F.S. Annexstein).

0304-3975/$ - see front matter 
doi:10.1016/j.tcs.2005.12.013

In this paper we focus on the effects of point-to-point Internet-network latencies on flooding broadcast operations. Such latencies can be highly variable and play a significant role in the performance of both structured and unstructured P2P networks. Latencies can be caused by the configuration and congestion of the underlying network transport layer (e.g., the Internet TCP/IP) or by the fact that peers are generally not bound to quality of service guarantees. We show that the flooding control mechanisms of time-to-live (TTL hop bounds) and unique message identification (UID sequence numbers) can, in combination, produce effects that significantly reduce the reach of broadcast messages. We prove that there exists an infinite class of networks for which broadcast messages reach only a vanishing fraction of the nodes of the network. In addition, in this paper we present efficient, exact and heuristic, algorithms that can be used (offline) to maximize the reach of message broadcasts. Finally, we provide empirical evidence that suggests that the trend indicated by our formal results are found in data obtained from the Gnutella network and through network simulations.

### 1.1. Control of broadcast flooding

We focus our attention on two specific governing mechanisms which control a flooding operation. These mechanisms work by executing a conditional test to determine whether or not a node should continue the flood locally. A node that continues the flood will forward the message to each of its neighbors, except the one from which the message was received.

*TTL-Mechanism*: The 'Time-to-Live' mechanism prevents messages from traveling farther than a specified number of hops, defined by an initial TTL value. TTL bounds are implemented by including in each message header a TTL value field. When a node receives a message it first checks to see if its TTL value is greater than zero. If it is greater than zero, the node continues the flood with a decremented TTL. Otherwise the message is dropped.

*UID-Mechanism*: The 'Unique Message Identification' mechanism prevents messages from being repeatedly transmitted from any node. This mechanism is implemented by including in each message header a UID (a unique ID label, or unique sequence number). When a node receives a message it checks to see if it has previously seen that message. If it has, the message is dropped and not forwarded. Otherwise, the node stores the new UID in a local table, and then continues the flood.

For the remainder of the paper we will consider broadcasts implemented as flooding operations working in a network environment of heterogeneous latencies, and operating under the constraints of the combination of the TTL- and UID-Mechanisms. When these pair of mechanisms are implemented together we show that they can impact network reachability via a phenomenon we call *short-circuiting*.

### 1.2. The short-circuiting phenomenon

This phenomenon characterizes a situation where a substantial number of nodes that are within the TTL bound of the broadcast source fail to receive the message due to the impact network latencies. We define latency of a network link as the time it takes a (unit size) message to traverse this link in the network. We interpret the latency of a message path as the sum of the latencies on the links involved in that path.

Consider a message broadcast from a source node $v_0$, and consider a path $P$ (i.e., a sequence of adjacent nodes) joining nodes $v_0$ and $v_p$, $P = \langle v_0, v_1, \ldots, v_p \rangle$. It is possible that there may be no throughput of the broadcast messages from $v_0$ to $v_p$ along $P$, even if the hop-length $p$ of the path $P$ is less than or equal to the initial TTL value $t$. This can result from heterogeneous latencies in the network, as the following scenario shows. Suppose there exists a message path $Q$ of length $q$, where $t \geqslant q > p$, from $v_0$ to some intermediate node $u = v_i$ of $P$, having a strictly smaller latency. Then a broadcast message originating from $v_0$, and following path $P$ will be killed (by UID-Mechanism) when it reaches $u$, since it is the duplicate of the earlier arriving message originating from $v_0$, but following path $Q$. Notice that there may also be no throughput along path $R$ consisting of the path $Q$ together with the subpath of $P$ from $u$ to $v_p$. This effect results from the fact that $R$ may have a hop-length strictly greater than $t$, and hence, by TTL-Mechanism there is no throughput of the broadcast message originating at $v_0$ along path $R$. And, indeed, there may be no throughput of the broadcast message along *any path* from $v_0$ to $v_p$.

The effects of short-circuiting are determined by the cardinality of the set of nodes that are reached by a particular message broadcast. We refer to this reached set of nodes as the *message horizon* of the broadcast. We measure the impact of short-circuiting by considering bounds on the *short-circuit broadcast ratio* which is defined as the cardinality of the broadcast message horizon divided by the cardinality of the set of nodes within the TTL-bound number of hops from the broadcast source—this is simply the number of nodes that would receive the broadcast message had the UID-mechanism not been used. To simplify the discussion we consider only connected, undirected, latency-weighted networks $G = (V, E, \ell)$; here we use $\ell$ to denote the latency function on the edge set $E$. For esthetic reasons, we are often interested in considering broadcast operations in which the TTL-bound is sufficient to reach the entire network. In other words, we want to consider networks in which the diameter (that is, the length of the longest minimum path length between node pairs) is no larger than the TTL bound used. In such a case, the bounds on the message horizon sizes directly yield bounds on the short-circuit broadcast ratio.

It is straightforward to verify that in any weighted, connected network any broadcast with a TTL-value of $t$ will reach at least $t$ nodes, even when both the TTL- and UID-Mechanisms are used. Somewhat surprisingly, this reachable set size is a tight bound, within a constant factor, due to the extreme effects of short-circuiting. That is, there are networks in which each and every broadcast with TTL-value of $t$ will reach only $O(t)$ nodes. In terms of the short-circuit-broadcast ratio, we show that there are networks in which broadcasts reach only a vanishingly small fraction of nodes. We note that in our proofs we construct an infinite class of small-diameter, highly clustered, "small-world" networks which have these extreme short-circuiting effects (see [7,14,2] for recent work on small-world networks). We summarize our results in the following theorem which is proved in Section 2.

**Theorem 1.** *Given integer parameters $t, n$ such that $1 \leqslant t \leqslant n$, there exists an n-node network $G = (V, E, \ell)$ with logarithmic diameter ($\Theta(\log n)$), such that any broadcast from any source node with an initial TTL value of t will only reach $O(t)$ nodes of G. Furthermore, there exists an infinite class of n-node networks for which the short-circuit broadcast ratio is $\Theta(\log n/n)$.*

In Section 3 we consider generalizations of the broadcasting problem, including versions in which multiple broadcasting sources are selected with a goal of maximizing the reach of broadcast messages. We also study generalizations in which the diameter or radius of the network is a parameter. We provide general, and nearly tight upper and lower bounds which account for these multiple parameters.

In Section 4 we consider algorithms that can be applied to maximize the message horizon by selecting one or more broadcast source node. We present an efficient algorithm which will precisely compute the message horizon set for any broadcast message. Furthermore, we present a heuristic algorithm that can efficiently find a set of $k$ broadcast source nodes which will maximize the message horizon to within a constant factor of optimal.

Our theoretical results suggest substantial negative impact in network environments that rely on both the TTL- and UID-Mechanisms to govern flooding. We therefore consider empirical evidence to test the expected significance of our results in real applications. In Section 5 we report a series of experimental results on instances of broadcasting in a variety of network settings. We provide measurements obtained through experimental studies with both simulated networks and the Gnutella P2P file sharing application. Our empirical results, reported in Section 5, provide supporting data that shows that the trends suggested by the theoretical results are likely present in the real world.

### 1.3. Formalizing the model

Given a latency-weighted network $G = (V, E, \ell)$, the flooding operation we study is defined by the following protocol regimen. Messages in the network we will denote $msg(uid, t, h)$, with unique message identifier $uid$, initial TTL-value $t$, and current hop-value $h$. Messages have payloads that are left unspecified for simplicity. The hop-value $h$ records the number of hops from the message's source node. The values of $uid$ and $t$ remain constant over the life of a message broadcast operation. The hop-value $h$ is increased with each hop, and is compared with $t$ at each node to test if the TTL limit has been reached. We will denote a message (ready for broadcast) originating at node $v_0$, with initial TTL-value $t$, by $msg(uid_{v_0}, t, 1)$. The broadcast regimen operates as follows, and defines the valid message paths associated with the transmission of the broadcast message.

(1) Source $v_0$ sends $msg(uid_{v_0}, t, 1)$ to all the neighbors of $v_0$, injecting the message on all links connected to $v_0$ at the same time.
(2) Nodes process messages on first-come-first-served basis as follows: when a node $v$ receives message $msg(uid_{v_0}, t, h)$ it checks whether $uid_{v_0}$ has been seen previously. If it has, then the message is dropped with no further processing.

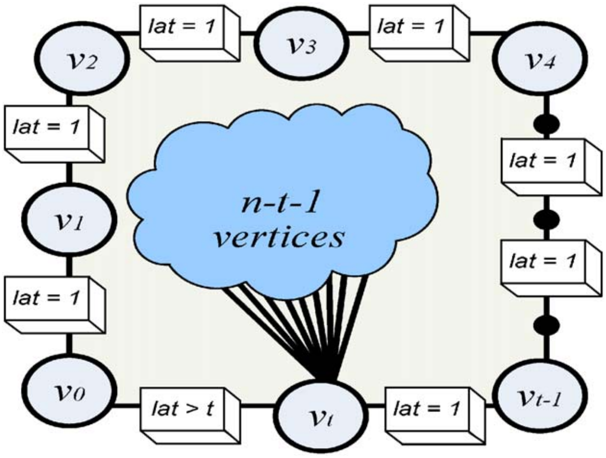


Fig. 1. The $n$-node network constructed in the proof of Proposition 1. A message broadcast from $v_0$ with a TTL-bound of $t$ will never reach the cloud of $n-t-1$ nodes, due to the large latency on edge $(v_0, v_t)$. However, a broadcast from $v_0$ with a TTL-bound of $t-1$ will reach all $n$ nodes in the network.

(3) If not, then $v$ records $uid_{v_0}$ in its local table, and checks whether or not $t = h$. If $t > h$, then $v$ replicates the payload and forwards the message $msg(uid_{v_0}, t, h+1)$ (with incremented hop count) to all its neighbors (except the node from which it received the message). If $t = h$ then the message is dropped and not forwarded.

For a given latency-weighting $\ell$ of the network, we use $\mathrm{H}_\ell(v_0, t)$ to denote the message horizon (or simply, $t$-horizon) of a message originating from $v_0$ with a TTL-bound of $t$. In other words, $\mathrm{H}_\ell(v_0, t)$ denotes the set of all nodes $v$ which receive a message $msg(uid_{v_0}, t, *)$ originating from $v_0$, where $*$ is any value.

## 2. The extremal effects of short-circuiting

Before we prove the main result on short-circuiting we prove a simpler result that shows, for a given source node, that increasing the TTL-bound can have significant negative impact on the size of the message horizon.

**Proposition 2.** *For all integers $t, n$, such that $2 < t < n$, there exists n-node network $G = (V, E, \ell)$, and a source vertex $v_0 \in V$, with the property that any broadcast from $v_0$ with an initial TTL value of t will reach only t nodes of G; However, any broadcast from $v_0$ with an initial TTL value of $t-1$ will reach all n nodes of G.*

**Proof.** We construct $G$ as follows. First, start with a $(t+1)$-node cycle $\langle v_0, v_1, \ldots, v_t, v_0\rangle$. Now add $n-t-1$ new nodes $U$, such that each new node is joined by a single edge to $v_t$. Assign a high latency to the edge $(v_0, v_t)$, so that it dominates the latency of the $t$-path $\langle v_0, v_1, \cdots, v_t\rangle$, see Fig. 1. A message broadcast from $v_0$ with a TTL-bound of $t$ will propagate through this $t$-node path $\langle v_0, v_1, \cdots, v_t\rangle$, however, the message will be killed at $v_t$ by the TTL-mechanism. Hence no node in the set $U$ will ever receive a copy of the broadcast message.

Finally, note that a message broadcast from $v_0$ with a TTL-bound of $t'$, with $2 \leqslant t' < t$, will reach all the $n-t$ nodes of $U$, in addition to at least $t'$ nodes of the $t$-path. $\square$

Our main theorem (Theorem 1) generalizes the proposition above by showing that the essence of the result is not dependent on the location of the source—that is, there are networks in which every node will experience extreme short-circuiting effects. Furthermore, the proof shows that an infinite class of highly-clustered, logarithmic diameter networks has this property. We now proceed with the proof of Theorem 1.

**Proof of Theorem 1.** To prove the upper bound on the message horizon we first construct a small network (cluster) that has the property that messages that enter the cluster always fail to propagate out of the cluster. We will use this cluster to build an $n$-node network promised in the statement of the theorem. For a given TTL-bound of $t$, we construct

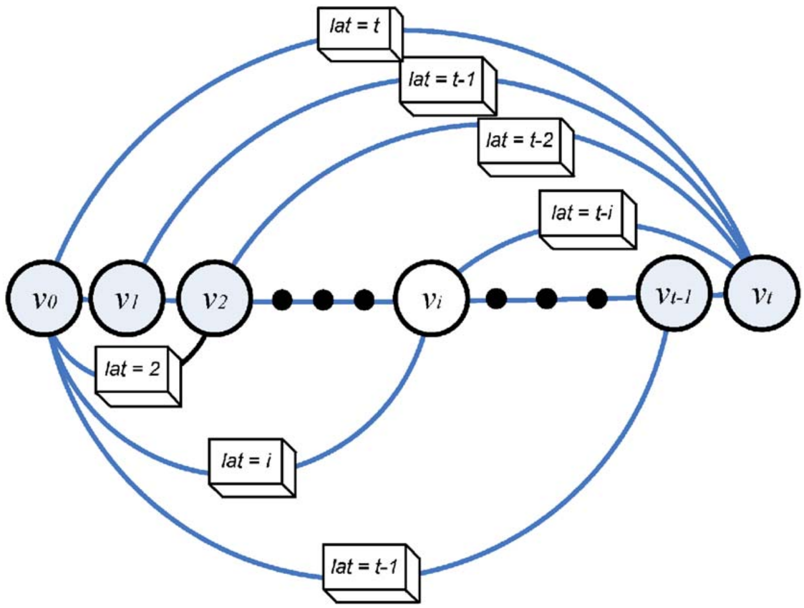


Fig. 2. The black-hole network cluster $BH_t$ defined in the construction of Theorem 1. Any message broadcast in $BH_t$ from end node $v_0$ with an initial hop count of $h < t$ will reach other end node $v_t$ with a hop count of precisely $t$, and thus cannot propagate out of the cluster.

a cluster called a *t-black-hole* graph $BH_t$. We use this name since, as we show, messages that enter the cluster never escape. To construct $BH_t$ we begin with a path $P = \langle v_0, v_1, \ldots, v_t \rangle$ of length $t + 1$ joining two end nodes $v_0$ and $v_t$. Assign a latency of weight zero to each edge $(v_i, v_{i+1})$ on this path. Now join each node $v_i$ (for $1 \leqslant i \leqslant t - 1$) to both end nodes $v_0$ and $v_t$; and assign latency weights for these edges equal to the distance between each pair in the original path, i.e., the edge $(v_i, v_0)$ is assigned weight $i$ and the edge $(v_i, v_t)$ is assigned weight $t - i$ (see Fig. 2).

The following claim states that messages that are broadcast from one end of a $t$-black-hole graph $BH_t$, can reach the other end only by using up all the TTL value in the message. This claim can be easily verified by the reader.

**Claim 3.** *For any integer h, such that* $0 \leqslant h < t$, *a message broadcast in* $BH_t$ *from end node* $v_0$ *with an initial hop count of h will reach other end node* $v_t$ *with a hop count of precisely t. The same result holds if we reverse roles of* $v_0$ *and* $v_t$.

It follows that by linking together black-hole graphs end-to-end, we can produce large graphs with large short-circuiting effects, independent of the broadcast source location. For example, consider constructing an $n$-node graph by linking together in a circle $n/t$ copies of the $BH_t$ graph. For purposes of exposition we will assume that $t$ divides $n$. This graph has the property that from any node location a message horizon has size at most $3t + 1 = \mathrm{O}(t)$. To prove the upper bound for graphs of logarithmic diameter, consider the following construction.

Consider a complete binary tree $T = (V_T, E_T)$ having $|V_T| = n/t$ vertices. We define the *BH-extension of T*, denoted by $T[BH_t]$ to be the $n + 1$ node graph obtained from $T$ by replacing each edge of $T$ with a black-hole graph $BH_t$, such that the end nodes of each black-hole are placed at the parent and child nodes of each tree edge. It is clear that the diameter of this graph is $diam(T[BH_t]) < 2 \log n$, for any $t \leqslant n$.

We now analyze the size of the maximum $t$-horizon set in the graph $T[BH_t]$. Suppose node $v$ is an end node of a some component black-hole in $T[BH_t]$. Then, since the $t$-horizon $\mathrm{H}(v, t)$ of $v$ is simply the union of the nodes in all the $t$ black-holes for which $v$ is an end node, we have that $|\mathrm{H}(v, t)| \leqslant 3t + 1$. On the other hand, assume that $v$ is not an end node of a black-hole. We can assume that $v$ lies in the interior of some component black-hole having end nodes $v_0$ and $v_t$. In this case the $t$-horizon of $v$ is the union of the nodes in all the component black-holes for which either $v_0$ or $v_t$ is an end node. Hence, in this case, and in the general case, we have that $\mathrm{H}(v, t)$ is of size at most $5t + 1 = \mathrm{O}(t)$.

We note that essentially the same argument can be used for any network that is a BH-extention of a bounded degree, logarithmic diameter network. □

## 3. Generalized broadcasting models

In this section we consider several generalizations of Theorem 1. We consider generalizations where the radius of the network is a parameter of the problem, and generalizations in which multiple message source nodes are optimally selected to increase the message horizon.

The *radius* of a network is the smallest radius for any node in the network, where the radius of a node is the least integer that upper bounds any shortest path distance from that node. It is possible to generalize Theorem 1 for graphs of arbitrary radius, as the following corollary states.

**Corollary 4.** *For all integers $r, t, n \geqslant 1$, there exists n-node network $G = (V, E, \ell)$ with radius r such that any broadcast with an initial TTL value of t will reach only $\mathrm{O}(n^{1/r}t)$ nodes of G.*

**Proof.** We modify the construction from the proof of Theorem 1 by considering a $n/t$ node rooted tree $T$ of depth $r$, whose interior nodes have $n^{1/r}$ children. Now consider the network $G = T[BH_t]$ which is the $BH_t$-extension of the tree $T$. Since the degree of each node of the original tree is bounded above by $n^{1/r} + 1$, we have that any broadcast with an initial TTL value of $t$ in the network $G$ will reach only at most $2n^{1/r} + 1$ of the $BH_t$ components, and thus the message horizon reaches at most $\mathrm{O}(n^{1/r}t)$ nodes of the network $G$. □

Now consider the impact of selecting $k \geqslant 1$ locations for broadcasting sources with the goal of increasing the message horizon. Such a generalization is natural in P2P networks, such as Gnutella, since peers often freely select neighbors to improve network connections. We consider the problem of determining bounds on the size of the message horizon for an optimal selection of $k \geqslant 1$ locations. We consider the case where the same message (or more precisely, the same message body) is broadcast from different locations, and each broadcast is per-source independent. That is, each broadcast source uses a unique $uid_s$ (for each $s$ in the set) so that messages from different sources will not interfere with each other. Hence, a node may receive the same message from different sources, however, the UID-mechanism will still insure that message originating from a single source will never circulate in loops. We define the $t$-horizon $\mathrm{H}_\ell(S, t)$ for a set of source nodes $S$, as the union of the nodes sets $\mathrm{H}_\ell(s, t)$ over each $s \in S$.

**Proposition 5.** *Given integers $k, t, n$, such that $kt \leqslant n$, there exist n-node networks such that for any set S of k broadcast sources, the size of the t-horizon $\mathrm{H}_\ell(S, t)$ is $\mathrm{O}(kt)$; furthermore, for any connected n-node network there is a set S of k broadcast sources for which the t-horizon $\mathrm{H}_\ell(S, t)$ is $\Omega(kt)$.*

**Proof.** The upper bound follows easily from the proof of Theorem 1. Since, from the construction given in that proof, each one of any $k$ broadcasting sources will reach only $\mathrm{O}(t)$ nodes.

We argue the lower bound as follows. For any connected network, we can determine a set of nodes $S = \{v_0, v_1, \ldots, v_k\}$ that can reach at least $kt/2$ nodes. Choose $v_0$ arbitrarily, and let $\mathrm{H}(v_0, t)$ be the $t$-horizon of $v_0$; and, let $H_0 \subset \mathrm{H}(v_0, t)$ be the set of nodes reachable by shortest-latency paths from $v_0$ within $t/2$ hops. Now let $v_1$ be any node that lies outside of $\mathrm{H}(v_0, t)$. Let $\mathrm{H}(v_1, t)$ be the $t$-horizon of $v_1$ and let $H_1 \subset \mathrm{H}(v_1, t)$ be the set of nodes reachable by shortest-latency paths from $v_1$ within $t/2$ hops. It follows that $H_0 \cap H_1 = \emptyset$ since otherwise there would exist a shortest-latency path of $t$ hops from $v_0$ to $v_1$, which would imply that $v_1 \in \mathrm{H}(v_0, t)$, a contradiction. Continuing, iteratively choose $v_i$ for each $i = 2, 3, \ldots, k-1$ so that each $v_i$ is not a node in any previously generated message horizon set. It follows from the selection that each set $H_i$ is pairwise disjoint from all $H_j$ for all $i \neq j$. Since each set $H_i$ has cardinality at least $t/2$ the result follows. □

We now consider a generalization in which all the parameters we have considered are taken together. Given integers $n, r, k, t \geqslant 1$, let $\mathrm{H}(n, r, k, t)$ denote the minimum size message horizon, for an optimal placement of $k$ sources broadcasting a message in an $n$-node network of radius-$r$ with initial TTL value of $t$. We prove the following theorem that gives nearly tight upper and lower bounds on the cardinality of this set.

**Theorem 6.**

$$c_1 n^{\frac{1}{r}} k^{1-\frac{1}{r}} \leqslant |\mathrm{H}(n, r, k, t)| \leqslant c_2 n^{\frac{1}{r}} (tk)^{1-\frac{1}{r}},$$

*where $c_1, c_2$ are constants independent of $n, r, k, t$.*

We prove the theorem with two lemmas giving the upper and lower bounds.

**Lemma 7.** *Given integers $n, r, k, t \geqslant 1$,*

$$|\mathrm{H}(n, r, k, t)| \leqslant c n^{\frac{1}{r}} (tk)^{1-\frac{1}{r}},$$

*for a constant $c \leqslant 8$.*

**Proof.** Consider a tree $T$ of depth $r$, whose root node has $y$ children and whose vertices at level $i$ all have $x$ children, $i = 1, \ldots, t-1$. Then, the number $n_T$ of nodes of $T$ satisfies:

$$n_T = 1 + y + yx + \cdots + yx^{r-1} > yx^{r-1} + 1. \tag{1}$$

Consider the $n$-node network $G = T[BH_t]$, that is, the BH-extension of $T$. We have that $n = t(n_T - 1) + 1 > tyx^{r-1}$. Therefore,

$$1 \leqslant y < nt^{-1} x^{1-r}. \tag{2}$$

Now consider locating $k$ sources at a subset $S$ of vertices in $G$. It follows that the size of the $t$-horizon $\mathrm{H}(S, t)$ of $S$ is bounded by the following.

$$|\mathrm{H}(S, t)| < ty + (2tx + t + 1)k \leqslant ty + 4tkx. \tag{3}$$

Applying inequality (2) we obtain

$$|\mathrm{H}(S, t)| < ty + 4tkx \leqslant nx^{1-r} + 4tkx. \tag{4}$$

We now choose a value of $x$ to minimize the size of $\mathrm{H}(S, t)$, e.g., let $x = (n/4tk)^{\frac{1}{r}}$. Substituting this choice in (4) we obtain

$$|\mathrm{H}(S, t)| < 2(n^{\frac{1}{r}} (4tk)^{1-\frac{1}{r}}). \tag{5}$$

Note that so long as $n > 4kt$, then $x \geqslant 1$ and $y \geqslant 1$. However, for $n \leqslant 4kt$ then the upper bound follows trivially since the right side is at least $n$. So Lemma 7 follows. □

We now prove the lower bound of Theorem 6 through the following lemma.

**Lemma 8.** *Given integers $n, r, k, t \geqslant 1$, such that $k \leqslant n$,*

$$|\mathrm{H}(n, r, k, t)| \geqslant c n^{\frac{1}{r}} k^{1-\frac{1}{r}},$$

*for a constant $c \geqslant 1/4$.*

**Proof.** We prove this result by showing that in any weighted $n$-node network $G$ of radius $r$, there exists a set $S \subset V(G)$ of $k$ sources that have $t$-horizon that matches the bound. To show this we need only to apply the fact that each source can reach all the nodes that are adjacent, i.e., in its immediate neighborhood, which of course, is independent of $t$. All nodes in the immediate neighborhood of a source must be reachable in a broadcast, since they cannot be short-circuited.

Now assume that $G$ is an $n$-node graph of radius $r$, and let $T_G$ be any radius-$r$ spanning tree of $G$. We choose a set of $k$ sources $v_0, v_1, \ldots, v_{k-1}$ as follows. First select one node $v_0$ of minimum radius in $T_G$. Now greedily select the remaining $k-1$ sources by choosing the nodes of highest degree in $T_G$. Let $S_0$ denote this set of $k$ nodes. For each $1 \leqslant i \leqslant t$, let $S_i$ denote those nodes of $T_G$ at hop distance exactly $i$ from $S_0$, where the hop distance to a set is defined as the smallest possible distance to any node in the set.

Note that the sum of degrees of nodes in $S_0$ in $T_G$ can be bounded as follows:

$$\sum_{v \in S_0} deg_{T_G}(v) \leqslant 2(|S_1| + |S_0| - 1).$$

This follows since the sum of degrees is at most twice the number of edges in the induced subgraph on $S_0 \cup S_1$, which follows from Euler's formula. Since the induced graph is a forest, we have that the number of edges in the induced subgraph is at most $|S_0| + |S_1| - 1$. Hence, the average degree in $T_G$ of nodes in $S_0$ is at most $2(|S_1|/|S_0| + 1)/|S_0|$. Let $m$ be the lower integer of this average value.

Now by the greedy selection of $S_0$, we have that each $v \in V - S_0$ has degree $deg(v) \leqslant m$. By this degree bound, we have that for each $i \geqslant 1$,

$$|S_i| \leqslant m|S_{i-1}| \leqslant m^2|S_{i-2}| \leqslant \cdots \leqslant m^{i-1}|S_1|.$$

Since the radius of $S_0$ in $T$ is $r$, we have that

$$n \leqslant |S_0| + |S_1| + |S_1|m + |S_1|m^2 + \cdots + |S_1|m^{r-1}.$$

So,

$$n \leqslant |S_0| + |S_1|(rm^{r-1}) \leqslant (|S_0| + |S_1|)(r2^{r-1}(|S_0| + |S_1|/|S_0|)^{r-1})$$

and so

$$n \leqslant 2^{r-1}(r+1)(|S_0| + |S_1|)^r/|S_0|^{r-1}$$

and hence,

$$|\mathrm{H}(n, r, k, t)| \geqslant |S_0| + |S_1| \geqslant (n^{\frac{1}{r}}|S_0|^{1-\frac{1}{r}})/(2^{1-\frac{1}{r}}(r+1)^{\frac{1}{r}}) \geqslant \frac{1}{4}n^{\frac{1}{r}}k^{1-\frac{1}{r}}.$$

The proof of the lower bound follows. □

## 4. Algorithms for maximizing the message horizon

In the previous section we considered extremal results which provided upper and lower bounds on the size of minimum message horizon given an optimal selection of message broadcast sources. In this section we provide algorithms that can be used to maximize the message horizon for given network instances.

### 4.1. An algorithm to compute the message horizon

We now show that there is an efficient algorithm which will precisely compute the message horizon set $\mathrm{H}_\ell(v_0, t)$, for a given source node $v_0$. We show more generally that we can compute the short-circuit distance (or simply, SC-distance), denoted by $\mathrm{d}_t(v_0, v)$. This distance is defined as the number of hops used by a broadcast message to reach a destination node $v$ from the broadcast source node $v_0$.

The following is an adaptation of a standard shortest-path algorithm that can be used to efficiently compute these short-circuit distances. This is done by maintaining a priority queue of latency values $L(v)$, representing an upper bound on the elapsed time from the beginning of the broadcast to the time the message first reaches node $v$. The algorithm proceeds similar to a discrete-event simulation where the minimum $L$-value is used to determine the next event, and thus we determine the delivery order of messages. By an inductive agreement, the following procedure correctly computes the delivery order and the short-circuit distance for each node.

*Procedure* **SC-distance** $(G, v_0, t)$

*Input*: An weighted graph $G = (V, E, \ell)$, source node $v_0 \in V$, and TTL-bound $t$

*Output*: The short-circuit distances $\mathrm{d}_t(v_0, v)$ from $v_0$ to each node $v$.

*Step* 1. Initialize $L(v_0) = \mathrm{d}_t(v_0) = 0$, Mark $v_0$.
For all $v \neq v_0$,
if $(v_0, v)$ is an edge, set $L(v) = \ell(v_0, v)$ and $\mathrm{d}_t(v) = 1$
otherwise set $L(v) = \mathrm{d}_t(v) = \infty$.

*Step* 2. Greedily choose an unmarked node $v$,
so that $L(v)$ is minimized; mark $v$

*Step* 3. If $\mathrm{d}_t(v) < t$ then update estimated latencies and SC-distances as follows:
for all unmarked $u$ adjacent to $v$, if $L(u) > L(v) + \ell(v, u)$, then
set $L(u) = L(v) + \ell(u, v)$ and $\mathrm{d}_t(u) = \mathrm{d}_t(v) + 1$

*Step* 4. If any unmarked nodes remain, go to Step 2.

The algorithm above can clearly be used to find the message horizon of a single node, simply by returning the set of all vertices with finite SC-distance. Obviously, the node with the largest message horizon can be found by determining the cardinality of the message horizon for each node.

### 4.2. Heuristic algorithm to maximize the message horizon

To increase the reach of messages we consider the case of using multiple, independent broadcast sources. In Section 3 we presented upper and lower bounds on the message horizon when using $k$ sources for a message broadcast. We consider now a greedy method as a heuristic for selecting $k$ sources to maximize the message horizon.

*Procedure* **Greedy k-Maximum t-Horizon** $(G, t, k)$

*Input*: An weighted graph $G = (V, E, \ell)$, a TTL-bound $t$, and an integer $k$.

*Output*: A $k$-subset of nodes $S \subset V(G)$

*Objective*: Maximize the size of the message horizon $\mathrm{H}(S, t)$

Initialize $S$ to empty set
For $i = 1$ to $k$ do
Find $s_i$ so that $\mathrm{H}(S \cup s_i, t)$ is maximized
Set $S = S \cup s_i$

As in Section 3 we assumed an independence property among message broadcasts using multiple sources. That is, by giving unique $(uid_s)$ tags for each message source, messages originating from different source nodes cannot interfere with each other. This assumption is sufficient to show that the above greedy approach will give good constant-factor approximation for maximizing the message horizon, since we can reduce the problem to the classical maximum coverage problem. The greedy algorithm for this problem is well known to yield a constant-factor approximation (see, for example, [5,11]). Thus, we have the following theorem.

**Theorem 9.** *The algorithm* **Greedy k-Maximum t-Horizon** *yields a polytime constant ratio* $1 - 1/e$*-approximation algorithm.*

## 5. Empirical studies

Our original interest in the effects of short-circuiting arose out of experimental evidence associated with the performance of a large-scale P2P file sharing application called Gnutella. Gnutella applies flooding in its search strategy,

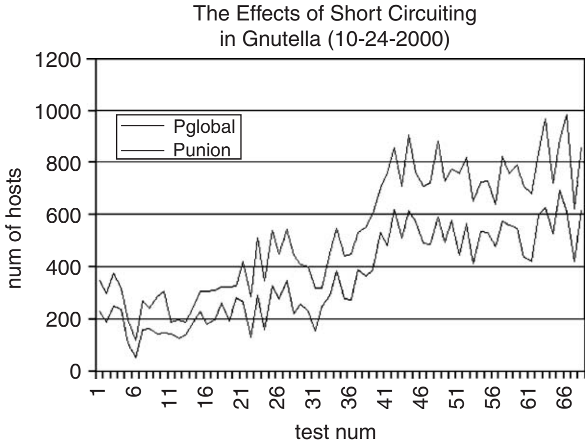


Fig. 3. The results of level-1 short-circuiting effects on the broadcast horizon on the Gnutella network. The *y*-axis represents the broadcast horizon size, and the *x*-axis labels each of 70 broadcast trials. The lower line (Pglobal) is the graph of the number of nodes reached by a broadcast message from a single source node $v_0$. The top line (Punion) is the resulting number of nodes reached in a union of sets of nodes reached from multiple distinct broadcasts from the source node $v_0$, where we used distinct message IDs for each neighbor. The discrepancy or distance between these two lines represents "level-1 short-circuiting" effects.

and this flooding is governed in its current implementation by TTL- and UID-Mechanisms as described in Section 1. Gnutella client software generally enforces limits on TTL values to at most 7, and its routing protocol makes it potentially vulnerable to the short-circuiting effects we have described. During experiments that involved crawling and mapping the entire Gnutella network, we noted that the number of reachable hosts reported by a client was substantially less than that estimated by an off-line analysis of the generated topology map. We consistently noted discrepancies of at least 50% as compared to the estimated size of a reachable set. After conjecturing that short-circuiting may play a substantial role in such discrepancies, we attempted to try to prove this empirically. We found it quite difficult to obtain highly accurate real-time statistics about the global topology of Gnutella. Hence, we devised an experimental method of discovering a subset of the effects of short-circuiting which relied only upon communication with a set of immediate neighbors. We called this restricted effect *level*-1 *short-circuiting*.

### 5.1. Level-1 short-circuiting in Gnutella

In our experiments we compared the 7-horizon of a message broadcast from $v_0$ with the 6-horizon of distinct message broadcasts from the neighbors of $v_0$. As noted, distinct ID labels prevent messages from interfering with each other, and thereby allows us to measure a subset of the total short-circuiting effect. The actual number of nodes reached by the broadcast of the shared message is compared to the union of node sets (message horizons) reached by the distinct broadcast messages. We noted, empirically, that short-circuiting was suggested by comparing the hop counts of messages responding to the different broadcasts. Fig. 3 shows the results of 70 trials of an experiment measuring level-1 short-circuiting. We note that the observed reductions in network broadcast coverage averaged over 55%.

### 5.2. Network simulation studies

We now turn our attention to a series of network simulation studies in which we were able to precisely isolate the effects of short-circuiting on synthetic network topologies. We report experiments ranging over a number of graph topologies, including snapshots of the Gnutella network, random small-world graphs, uniform random graphs, and some highly structured graphs, including the mesh and hypercube networks.

Snapshots of the Gnutella network were obtained by a network crawler that performed near real-time topology discovery in parallel. Details of the crawler, statistics, and pictures of the Gnutella topology can be found in [6]. Analysis of the obtained data reveals interesting structural properties of the network, including strong "small-world"

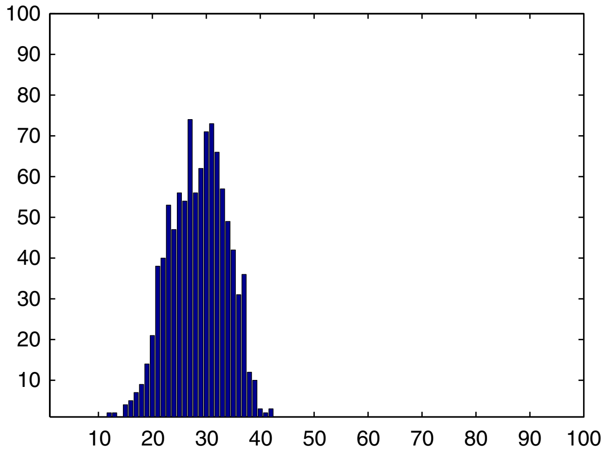


Fig. 4. Histogram of trials with $TTL = 10$ from experiment on the Watts–Strogatz topology, see Table 1. The *y*-axis represents the number of trials, and the *x*-axis represents short-circuit broadcast ratio in percentage. Note the narrow variance around the mean broadcast ratio (MBR) of approximately 28%.

characteristics and a power law distribution of node degrees and latencies. A class of random small-world networks were obtained using the Watts–Strogatz model [14].

To model network latencies we used several classes of weights representing various commonly used Internet connection bandwidths. We conducted our experiments by using uniform random distributions of these weights. We applied the SC-distance (discussed in Section 4) to measure the message horizon sizes and used this value to compute the short-circuit broadcast ratios.

### *5.3. Statistical table*

Each row of the following Table 1 represents results from 100 trials, broadcasting from a fixed, random source node. In each trial we used a new set of randomly generated latencies. In each row of Table 1 we report for a fixed $TTL = t$, the worst, average, and best observed message horizon sizes. We give the worst-case short-circuit-broadcast ratio (**WBR**), which is obtained by dividing the worst message horizon size by the size of the reachable network, and the mean short-circuit-broadcast ratio (**MBR**), which is obtained by dividing the average message horizon size by the reachable network size.

The histogram shown in Fig. 4 represents the distribution of short-circuit-broadcast ratio values over the set of 100 trials using a TTL value of $t = 10$. Note that the distribution is strongly clustered around the mean of approximately 28%.

## 6. Observations and conclusions

We have observed the most significant impact of short-circuiting on “small-world” topologies such as our Gnutella snapshots and Watts–Strogatz network models. For certain values of *t*, for these graphs, we have seen average reduction in message horizon sizes of over 70%, and in the worst case this reduction is as large as 90%. Furthermore, in such graphs the difference between the worst observed message horizon and the best observed message horizon can differ by factors of 4 or more.

In our experimental studies we have observed that both random graphs and highly structured graphs such as the mesh and hypercube tend to have, on average, less pronounced short-circuiting effects, as compared with “small-world” graphs. In general, for a fixed $TTL = t$, the distribution of message horizon sizes tend to be normally distributed with small variance, independent of network topology, as illustrated in Fig. 4.

Table 1
This table provides statistics on the impact of short-circuiting effects in various networks as measured by message horizon sizes and short-circuit broadcast ratios. The table reports the results of experiments on five different network topologies: Gnutella (using topology obtained from network crawl, see [6]), Waats–Strogatz model (using parameters $k = 3$, $p = 0.2$, see [14]), random graph model (using a uniform edge model with average degree = 8), 2-dimensional mesh, and hypercube network (order-13). The number of nodes (indicated in parentheses) for each network represents the subgraph of the original topology reachable within the hop-bound given by the TTL. Each row of the table reports for a fixed $TTL = t$, the worst, average, and best observed message horizon sizes, the worst-case short-circuit-broadcast ratio (**WBR**), and the mean short-circuit-broadcast ratio (**MBR**)

| Topology (# nodes) | TTL | Worst | Mean | Best | WBR | MBR |
|---|---|---|---|---|---|---|
| Gnutella (843) | 6 | 246 | 589 | 1107 | 22% | 53% |
| Gnutella (1124) | 7 | 419 | 806 | 1040 | 37% | 72% |
| Gnutella (1125) | 8 | 566 | 915 | 1071 | 50% | 81% |
| Watts–Strogatz (1399) | 7 | 278 | 498 | 723 | 20% | 36% |
| Watts–Strogatz (2771) | 8 | 434 | 819 | 1364 | 16% | 30% |
| Watts–Strogatz (5018) | 9 | 765 | 1388 | 2307 | 15% | 28% |
| Watts–Strogatz (7729) | 10 | 977 | 2148 | 3420 | 13% | 28% |
| Random (9021) | 5 | 4686 | 5986 | 6875 | 52% | 66% |
| Random (9998) | 6 | 6557 | 8143 | 8809 | 66% | 81% |
| Random (10000) | 7 | 8113 | 9060 | 9443 | 81% | 91% |
| 2-D Mesh (28) | 6 | 18 | 23 | 28 | 64% | 82% |
| 2-D Mesh (36) | 7 | 21 | 29 | 36 | 58% | 81% |
| 2-D Mesh (45) | 8 | 25 | 36 | 45 | 56% | 80% |
| Hypercube (2380) | 5 | 1120 | 1750 | 2139 | 47% | 74% |
| Hypercube (5812) | 7 | 2796 | 4422 | 5298 | 48% | 76% |
| Hypercube (7099) | 8 | 3970 | 5813 | 6644 | 56% | 82% |
| Hypercube (7814) | 9 | 6023 | 6844 | 7424 | 77% | 88% |

## Acknowledgements

A preliminary version of this work appeared in 13th ACM SPAA Conference 2001. This research was supported in part by NSF Grant No. CCR-9877139, and an Ohio Board of Regents' Research Investment Grant. The authors would like to thank an anonymous referee for comments that helped to improve the presentation.

## References

[1] D. Bertsekas, R. Gallager, Data Networks, Prentice-Hall, New York, 1992.
[2] A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tomkins, J. Wiener, Graph structures in the web, Comput. Networks 33 (1–6) (2000) 309–320.
[3] M. Castro, M. Costa, A. Rowstron, Should we build gnutella on a structured overlay?, in: Proc. of HotNets 2003, November 2003.
[4] Y. Chawathe, S. Ratnasamy, L. Breslau, N. Lanham, S. Shenker, Making Gnutella-like P2P systems scalable, in: Proc. of SIGCOMM, 2003.
[5] D. Hochbaum, Approximation Algorithms for NP-Hard Problems, PWS Publishing Co., MA, 1997.
[6] M. Jovanovic, F. Annexstein, K. Berman, Scalability issues in large peer-to-peer networks—a case study of gnutella, Technical Report, University of Cincinnati, January 2001.
[7] J. Kleinberg, The small-world phenomenon: an algorithmic perspective, in: 32nd ACM Symposium on the Theory of Computing, 2000.
[8] J. Lipman, P. Boustead, J. Chicharo, J. Judge, Optimised flooding in ad hoc networks, in: Proc. of Workshop on the Internet, Telecomm. and Signal Proc. (WITSP'03), December 2003.
[9] B.T. Loo, R. Huebsch, I. Stoica, J. Hellerstein, The case for a hybrid P2P search infrastructure, in: The Third Internat. Workshop on Peer-to-Peer Systems (IPTPS'04), 2004.
[10] Q. Lv, P. Cao, E. Cohen, K. Li, S. Shenker, Search and replication in unstructured peer-to-peer networks, in: Proc. 16th Internat. Conf. on Supercomputing, 2002.
[11] G. Nemhauser, L. Woolsey, Maximizing submodular set functions: formulations and analysis of algorithms, in: Studies of Graphs and Discrete Programming, North-Holland, Amsterdam, 1981, pp. 279–301.
[12] M. Portmann, P. Sookavatana, S. Ardon, A. Seneviratne, The cost of peer discovery and searching in the gnutella peer-to-peer file sharing protocol, in: Ninth IEEE Internat. Conf. on Networks, 2001.
[13] The Gnutella Homepage, ⟨http://www.gnutella.com⟩.
[14] D.J. Watts, S.H. Strogatz, Collective dynamics of small-world networks, Nature 393 (1998) 440–442.
[15] B. Williams, T. Camp, Comparison of broadcasting techniques for mobile ad hoc networks, in: Proc. of MOBIHOC, 2002, June 9–11.