Near Optimal Routing in a Small-World Network with Augmented Local Awareness

Jianyang Zeng[1] and Wen-Jing Hsu[1] Jiangdian Wang[2]

[1] Center for Advanced Information Systems, Nanyang Technological University, Singapore.
Email: zengjy@gmail.com, hsu@ntu.edu.sg.
[2] School of Electrical Electronic Engineering, Nanyang Technological University, Singapore.
Email: wang0059@pmail.ntu.edu.sg.

Abstract. In order to investigate the routing aspects of small-world networks, Kleinberg [13] proposes a network model based on a d-dimensional lattice with long-range links chosen at random according to the d-harmonic distribution. Kleinberg shows that the greedy routing algorithm by using only local information performs in $O(\lg^2 n)$ expected number of hops, where n denotes the number of nodes in the network. Martel and Nguyen [17] have found that the expected diameter of Kleinberg's small-world networks is $\Theta(\lg n)$. Thus a question arises naturally: Can we improve the routing algorithms to match the diameter of the networks while keeping the amount of information stored on each node as small as possible?

Existing approaches for improving the routing performance in the small-world networks include: (1) Increasing the number of long-range links [2, 15]; (2) Exploring more nodes before making routing decisions [14]; (3) Increasing the local awareness for each node [10, 17]. However, all these approaches can only achieve $O\big((\lg n)^{1+\epsilon}\big)$ expected number of hops, where $\epsilon > 0$ denotes a constant. We extend Kleinberg's model and add two augmented local links for each node, which are connected to nodes chosen randomly and uniformly within $\lg^2 n$ Mahattan distance. Our investigation shows that these augmented local connections can make small-world networks more navigable.

We show that if each node is aware of $O(\lg n)$ number of neighbors via the augmented local links, there exist both non-oblivious and oblivious algorithms that can route messages between any pair of nodes in $O(\lg n \lg \lg n)$ expected number of hops, which is a near optimal routing complexity and outperforms the other related results for routing in Kleinberg's small-world networks. Our schemes keep only $O(\lg^2 n)$ bits of routing information on each node, thus they are scalable with the network size. Our results imply that the awareness of $O(\lg n)$ nodes through augmented links is more efficient for routing than via the local links [10, 17].

Besides adding new light to the studies of social networks, our results may also find applications in the design of large-scale distributed networks, such as peer-to-peer systems, in the same spirit of Symphony [15].

1 Introduction

A well-known study by Milgram in 1967 [18] shows the *small-world phenomenon* [9], also called "six degree of separation", that any two people in the world can be connected by a chain of six (on the average) acquaintances, and people can deliver messages efficiently to an unknown target via their acquaintances. This study is repeated by Dodds, Muhamad, and Watts [8] recently, and the results show that it is still true for today's social network. The small-world phenomenon has also been shown to be pervasive in networks from nature and engineering systems, such as the World Wide Web [21, 1], peer-to-peer systems [2, 16, 15, 22], etc.

Recently, a number of network models have been proposed to study the small-world properties [19, 21, 13]. Watts and Strogatz [21] propose a random rewiring model whose diameter is a poly-logarithmic function of the size of the network. The model is constructed by adding a small number of random edges to nodes uniformly distributed on a ring, where nodes are connected densely with their near neighbors. A similar approach can also be found in Ballabás and Chung's earlier work [6], where the poly-logarithmic diameter of the random graph is achieved by adding a random matching to the nodes of a cycle. However, these models fail to capture the algorithmic aspects of a small-world network [13]. As commented by Kleinberg in [13], the poly-logarithmic diameter of some graphs does not imply the existence of efficient routing algorithms. For example, the random graph in [6] yields a logarithmic diameter, yet any routing using only local information requires at least $\sqrt{n}$ expected number of hops (where n is the size of the network) [13].

In order to incorporate routing or navigating properties into random graph models, Kleinberg [13] develops a new model based on a d-dimensional torus lattice with long-range links chosen randomly from the d-harmonic distribution, i.e., a long-range link between nodes u and v exists with probability proportional to $Dist(u,v)^{-d}$, where $Dist(u,v)$ denotes the Mahattan distance between nodes u and v. Based on this model, Kleinberg then shows that routing messages between any two nodes can be achieved in $O(\lg^2 n)$ [3] expected number of hops by applying a simple greedy routing algorithm using only local information. This bound is tightened to $\Theta(\lg^2 n)$ later by Barrière et al. [3] and Martel et al. [17]. Further research [16, 14, 17, 10] shows that in fact the $O(\lg^2 n)$ bound of the original greedy routing algorithm can be improved by putting some extra information in each message holder. Manku, Naor, and Wieder [16] show that if each message holder at a routing step takes its own neighbors' neighbors into account for making routing decisions, the bound of routing complexity can be improved to $O(\frac{\lg^2 n}{q \lg q})$, where q denotes the number of long-range contacts for each node. Lebhar and Schabanel [14] propose a routing algorithm for 1-dimensional Kleinberg's model, which visits $O(\frac{\lg^2 n}{\lg^2(1+q)})$ nodes on expectation before routing the message, and they show

[3] The logarithmic symbol lg is with the base 2, if not otherwise specified. Also, we remove the ceiling or floor for simplicity throughout the paper.

that a routing path with expected length of $O(\frac{\lg n(\lg\lg n)^2}{\lg^2(1+q)})$ can be found. Two research groups, Fraigniaud et al. [10], and Martel and Nguyen [17], independently report that if each node is aware of its $O(\lg n)$ closest local neighbors, the routing complexity in d-dimensional Kleinberg's small-world networks can be improved to $O(\lg n \lg^{1+1/d} n)$ expected number of hops. The difference is that [17] requires keeping additional state information, while [10] uses an oblivious greedy routing algorithm. Fraigniaud et al. [10] also show that $\Theta(\lg^2 n)$ bits of topological awareness per node is optimal for their oblivious routing scheme. In [17], Martel and Nguyen show that the expected diameter of a d-dimensional Kleinberg network is $\Theta(\lg n)$. As such, there is still some room for reducing the routing complexity, which motivates our work.

Other small-world models have also been studied. In their recent paper [20], Nguyen and Martel study the diameters of variants of Kleinberg's small-world models, and provide a general framework for constructing classes of small-world networks with $\Theta(\lg n)$ expected diameter. Aspnes, Diamadi, and Shah [2] find that the greedy routing algorithms in directed rings with a constant number of random extra links given in *any* distribution requires at least $\Omega(\lg^2 n/\lg\lg n)$ expected number of hops. Another related models are the small-world percolation models [16, 4, 7, 5]. The diameters of these models are studied by Benjamin et al. [4], Coppersmith et al. [7] and Biskup [5]. The routing aspects of the percolation models, such as the lower bound and upper bound of greedy routing algorithms with 1-lookahead, are studied in [16].

Applications of small-world phenomenon in computer networks include efficient lookup in peer-to-peer systems [16, 2, 15, 22], gossip protocol in a communication network [12], flooding routing in ad-hoc networks [11], and the study of diameter of World Wide Web [1], etc.

1.1 Our Contributions

We extend Kleinberg's structures of small-world models with slight change. Besides having long-range and local links on the grid lattice, each node is augmented with two extra links connected to nodes chosen randomly and uniformly within $\lg^2 n$ Mahattan distance. Based on this extended model, we present near optimal algorithms for decentralized routing with $O(\lg n)$ augmented awareness. We show that if each node is aware of $O(\lg n)$ number of nodes via the augmented neighborhood, there exist both non-oblivious and oblivious routing algorithms that perform in $O(\lg n \lg\lg n)$ expected number of hops (see Theorem 1 and Theorem 2). Our investigation constructively show that the augmented local connections can make small-world networks more navigable.

Scheme	#bits of awareness	#steps expected	Oblivious or Non-oblivious?
Kleinberg's greedy [13, 2, 15]	$O(q \lg n)$	$O(\lg^2 n/q)$	Oblivious
NoN-greedy [16]	$O(q^2 \lg n)$	$O(\lg^2 n/(q \lg q))$	Non-oblivious
Decentralized algorithm in [14]	$O\big(\lg^2 n/\lg(1+q)\big)$	$O\big((\lg n)^2/\lg^2(1+q)\big)$	Non-oblivious
Decentralized algorithm [17]	$O(\lg^2 n)$	$O\big((\lg n)^{1+1/d}\big)$	Non-oblivious
Indirect-greedy algorithm [10]	$O(\lg^2 n)$	$O\big((\lg n)^{1+1/d}\big)$	Oblivious
Our algorithms for the model with augmented awareness	$O(\lg^2 n)$	$O(\lg n \ \lg \lg n)$	Both are provided

Table 1. Comparisons of our decentralized routing algorithms with the other existing schemes. In the first three schemes (in [13, 2, 15, 16, 14]), we suppose that each node has q long-range contacts, while in the next three schemes (in [17, 10] and this paper), we suppose that each node has one long-range contact. A routing protocol is *oblivious* if the message holder makes routing decisions only by its local information and the target node, and independently of the previous routing history, otherwise, it is said to be *non-oblivious*.

A comparison of our algorithm with the other existing schemes is shown in Table 1. Our decentralized routing algorithms assume that each node can compute a shortest path among a poly-logarithmic number of known nodes. Such an assumption is reasonable since each node in a computer network is normally a processor and can carry out such a simple computation. Our schemes keep $O(\lg^2 n)$ bits of routing information stored on each node, thus they are scalable with the increase of network size. Our investigation shows that the awareness of $O(\lg n)$ nodes through the augmented links is more efficient for routing than via the local links [10, 17].

We note that besides adding new light to the studies of social networks such as Milgram's experiment [18], our results may also find applications in the design of large-scale distributed networks, such as peer-to-peer systems, in the same spirit of Symphony [15]. Since the links in our extended model are randomly constructed according to the probabilistic distribution, the network may be less vulnerable to adversarial attacks, and thus provide good fault tolerance.

1.2 Organization

The rest of the paper is organized as follows. Section 2 gives notations for Kleinberg's small-world model and its extended version with augmented local connections. Section 3 gives some preliminary notations for decentralized routing. In Section 4, we propose both non-oblivious and oblivious routing algorithms with near optimal routing complexity in our extended model. Section 5 gives the experimental evaluation of our schemes. Section 6 briefly concludes the paper.

2 Definitions of Small-World Models

In this section, we will give the definition of Kleinberg's small-world model and its extended version in which each node has extra links. For simplicity, we only consider the one-dimensional model with *one* long-range contact for each node. In addition, we assume that all links are directed, which is consistent with the real-world observation, for example, person x knows person y, but y may not know x.

Definition 1. (Kleinberg's Small-World Network (KSWN) [13]) *A Kleinberg's Small-World Network, denoted as $\mathcal{K}$, is based on a one-dimensional torus (or ring) $[n] = [0, 1, \cdots, n]$. Each node u has a directed local link to its next neighbor $(u+1) \mod n$ on the ring. We refer to this local link as* ***Ring-link*** *(or* ***R-link*** *for short), and refer to node $(u+1) \mod n$ as the* ***R-neighbor*** *of node u. In addition, each node has* one *long-range link to another node chosen randomly according to the 1-harmonic distribution, that is, the probability that node u sends a long-range link to node v is $\Pr[u \rightarrow v] = \frac{1}{Z_u \cdot Dist(u,v)}$, where $Dist(u, v)$ denotes the ring distance* [4] *from u to v, and $Z_u = \sum_{z \neq u} \frac{1}{Dist(u,z)}$. We refer to this long-range link as the* ***Kleinberg-link*** *(or* ***K-link*** *for short), and refer to node v as a* ***K-neighbor*** *of node u if a K-link exists from u to v.*

Our extended structure introduces several extra links for each node. Its definition is given below.

Definition 2. (KSWN with Augmented Local Connections (KSWN*)) *A Kleinberg's Small-World Network with Augmented Local Connections, denoted as $\mathcal{K}^*$, has the same structure of KSWN, except that each node u in $\mathcal{K}^*$ has two extra links to nodes chosen randomly and uniformly from the interval $(u, u + \lg^2 n]$. We refer to these two links as the* ***augmented local links*** *(or* ***AL-links*** *for short), and refer to node v as a* ***AL-neighbor*** *of node u if a AL-link exists from u to v.*

There are in total four links for each node in a KSWN*: one R-link, one K-link, two AL-links. We refer to all nodes linked directly by node u as the **immediate neighbors** of u. Our extended structure retains the same $O(1)$ order of node degree as that of Kleinberg's original model.

3 Decentralized Routing Algorithms

Based on the original model, Kleinberg presents a class of decentralized routing algorithms, in which each node makes routing decisions by using local information and in a greedy fashion. In other words, the message holder forward the message to its immediate neighboring node, including its K-neighbor, which is closest to the destination in terms of the Mahattan distance. Kleinberg shows that such a simple greedy algorithm performs in $O(\lg^2 n)$ expected number of hops. The other

[4] or Mahattan distance for multi-dimensional models.

existing decentralized routing algorithms [2, 15, 14, 10, 17, 16] mainly rely on three approaches to improve routing performance: (1) Increasing the number of long-range links [2, 15]; (2) Exploring more nodes before making routing decisions [14]; (3) Increasing the local awareness for each node [10, 17]. However, so far using these approaches can only achieve $O\big((\lg n)^{1+\epsilon}\big)$ expected number of hops in routing, where $\epsilon > 0$. Although the scheme in [16], where each node makes routing decision by looking ahead its neighbors's neighbors, can achieve an optimal $O(\lg n/\lg\lg n)$ bound, their result depends on the fact that each node has at least $\Omega(\lg n)$ number of K-links.

There are normally two approaches for decentralized routing: oblivious and non-oblivious schemes [10]. A routing protocol is *oblivious* if the message holder makes routing decisions only by its local information and the target node, and independently of the previous routing history. On the other hand, if the message holder needs to consider certain information of the previous routing history to make routing decisions, the protocol is referred to as *non-oblivious*. The non-oblivious protocol is often implemented by adding a header segment to the message packet so that the downstream nodes can learn the routing decisions of upstream nodes by reading the message header information. The scheme in [10] is oblivious, while the schemes in [14] and [17] are non-oblivious.

We refer to the message holder as *the current node*. For the current node x, we define a sequence of node sets $T_0, T_1, \cdots, T_i, \cdots$, where $T_0 = \{x\}$, $T_1 = \{$ u's AL-neighbors, $\forall u \in T_0\}$, $T_2 = \{u\text{'s AL-neighbors}, \forall u \in T_1\}$, and so on. We refer to T_i as the set of nodes in *the ith level of AL neighborhood*, and let $H_i = \bigcup_{j \le i} T_j$ denote the set of all nodes in *the first i levels of AL neighborhood*. At a certain level i of AL neighborhood, we may also refer to H_{i-1} as the set of *previously known nodes*. Let $L_i = T_i - H_{i-1}$ denote the set of *new* nodes discovered during the ith level of AL neighborhood. Let $A_x(k) = H_k$ denote the *augmented local awareness* (or *AL awareness* for short) of a given node in a KSWN*, where each node is aware of the first k levels of its AL neighborhood.

In Section 4, we will show that there exists a sufficiently large constant σ such that $|A_x(\lg\lg n)| \ge \lg n/\sigma$, based on which we propose both non-oblivious and oblivious routing algorithms running in $O(\lg n \lg\lg n)$ expected number of hops and requiring $O(\lg^2 n)$ bits of information on each node.

Our near optimal $O(\lg n \lg\lg n)$ bound on the routing complexity outperforms the other related results for Kleinberg's small-world networks. To our knowledge, our algorithms achieve the best expected routing complexity while requiring at most $O(\log^2 n)$ bits of information stored on each node.

4 Near Optimal Routing with $O(\lg n)$ Awareness

4.1 Augmented Local Awareness of $O(\lg n)$

In this subsection, we will show that $|A_x(\lg\lg n)|$, the number of distinct nodes that node x is aware of via the first $\lg\lg n$ levels of AL neighborhood, is not less than $\lg n/\sigma$ for a constant σ, which, as will be shown in Lemma 3, is sufficiently large to guarantee that $A_x(\lg\lg n)$ contains a K-link that jumps over half distance (Suppose that the destination node is at a certain large distance from the current node). These results are useful for the subsequent analysis of our oblivious and non-oblivious routing schemes.

Lemma 1. *Let $A_x(\lg\lg n)$ denote the AL awareness of node x in a KSWN* $\mathcal{K}^*$, where each node is aware of $\lg\lg n$ levels of AL-neighbors. Then*

$$\Pr[\ |A_x(\lg\lg n)| \geq \frac{\lg n}{\sigma}\ ] > \psi,$$

where σ denotes a sufficiently large constant and ψ denotes a positive constant.

Proof: Throughout the proof, we assume that $|H_i| < \frac{\lg n}{\sigma}$ for all $1 \leq i \leq \lg\lg n$, otherwise, the lemma already holds, since $|A_x(\lg\lg n)| = |H_{\lg\lg n}| > \lg n/\sigma$. We will show that at each level of AL neighborhood, the probability that each AL-link points to previously known nodes is small so that a large number of distinct nodes will be found via the first $\lg\lg n$ levels of AL neighborhood.

Consider the construction of a AL-link for the current node x. By definition of KSWN*, each AL-link of x is connected to a node randomly and uniformly chosen from the interval $(x, x + \lg^2 n]$, that is, each AL-link of x points to a node in the interval $(x, x + \lg^2 n]$ with probability $(\lg n)^{-2}$. By assumption, there could be no more than $\lg n/\sigma$ previously known nodes in the interval $(x, x + \lg^2 n]$. Thus, the probability for a AL-link of a given node to point to a previously known node is at most $(\lg n/\sigma) \cdot (\lg n)^{-2} = (\sigma \lg n)^{-1}$. Thus, the probability for a AL-link of x to point to a new node is at least $1 - (\sigma \lg n)^{-1}$. There are in total at most $2 \cdot |H_{\lg\lg n}| \leq 2\lg n/\sigma$ number of AL-links, so the probability for all AL-links to point to new nodes is at least $(1 - (\sigma \lg n)^{-1})^{2\lg n/\sigma} \geq 1 - \frac{2}{\sigma^2}$ for sufficiently large n. Here we use the fact $(1 + x)^a \geq 1 + ax$ for $x > -1$ and $a \geq 1$. When σ is a sufficiently large constant, we have $\Pr[\ |A_x| \geq \frac{\lg n}{\sigma}\ ] > \psi$ for a positive constant $\psi = 1 - \frac{2}{\sigma^2} > 0$. Thus, the proof of Lemma 1 is completed. ∎

4.2 Non-Oblivious Decentralized Routing

Our non-oblivious routing algorithm is given as follows: Initially the source node s finds in its AL awareness $A_s(\lg\lg n)$ an intermediate node z that is closest to the destination, and then computes a shortest path π from s to z in $A_s(\lg\lg n)$. Before routing the message, s adds the information about shortest path π to the message header. Once the message passes a node on the shortest path π, the next stop is read off the header stack. When the message reaches node z, node z can tell that it is an intermediate target by reading the message header and then route the message to its K-neighbor. Such processes are repeated until the message reaches a certain node close enough to the destination node. After that, Kleinberg's plain greedy algorithm can be used to route the message effectively to the target node. Given a message M, a source node s and a target node t in a KSWN* $\mathcal{K}^*$, the pseudocodes of our non-oblivious algorithm running on the current node x are given in Algorithm 1.

Algorithm 1

Input: the source s, the target t and the message M.
Initialization:
  Current node $\leftarrow s$.
  Set the header stack of the message M to be empty.
while Distance between the current node and the destination $\geq (\lg n)^2 \lg\lg n$ **do**
  if the header stack of the message M is empty **then**
    Route the message M to x's K-neighbor y.
    Find an intermediate node z in $A_y(\lg\lg n)$ whose K-neighbor is closest to t (ties are broken arbitrarily).
    Compute a shortest path $\pi : x_0 = y, x_1, \cdots, x_t = z$ from y to z, and push the shortest path information $\pi : x_1, \cdots, x_t = z$ into the header stack of the message M.
  else
    Pop up the first node x_i from the header stack and route the message M to node x_i.
  end if
end while
Final phase (Kleinberg's greedy algorithm):
  Route the message M to an immediate neighbor of x that is closest to the target t, until it reaches t.

Next we will analyze the performance of the Algorithm 1. We first give a basic lemma, which provide a lower bound and an upper bound on the probability of the existence of a K-link in Kleinberg's small-world networks. Its proof can be found in Appendix A.

Lemma 2. *Let* $\Pr[u \xrightarrow{K} v]$ *denote the probability that node* u *sends a K-link to node* v *in a KSWN** $\mathcal{K}^*$. *Suppose that* $a \leq Dist(u,v) \leq b$, *then* $\frac{c_1}{b \lg n} \leq \Pr[u \xrightarrow{K} v] \leq \frac{c_2}{a \lg n}$, *where* c_1 *and* c_2 *are constants independent of* n.

In Lemma 1, we have shown that $\Pr[\, |A_x(\lg\lg n)| \geq \lg n/\sigma \,]$ is at least a positive constant for a sufficiently large constant σ. Based on this result, Lemma 3 shows that the probability for $A_x(\lg\lg n)$ to contain a K-link jumping over half distance is at least a positive constant.

Lemma 3. *Suppose that the distance between the current node x and the target node t in a KSWN* $\mathcal{K}^*$ is $Dist(x,t) \geq \lg^2 n \lg\lg n$. Then with probability at least a positive constant, node x's AL awareness $A_x(\lg\lg n)$ contains a K-neighbor within $Dist(x,t)/2$ distance to the target node t .*

Proof: Let $\mathcal{A}$ denote the event that $|A_x(\lg\lg n)| \geq \frac{\lg n}{\sigma}$. By Lemma 1, we have $\Pr[\mathcal{A}] > \psi$ for a constant $\psi > 0$.

Let $B_l(t)$ denote the set of all nodes within l ring distance to t. Let $\Pr[x \xrightarrow{K} B_l(t)]$ denote the probability that x's K-neighbor is inside the ball $B_l(t)$.

Let $m = Dist(x,t)$. By Lemma 2, the probability for a K-link to point to a given node inside the ball $B_{\frac{m}{2}}(t)$ is at least $\frac{c_1}{m \lg n}$, so we have

$$\Pr[x \xrightarrow{K} B_{\frac{m}{2}}(t)] \geq |B_{\frac{m}{2}}(t)| \cdot \frac{c_1}{m \lg n} = \frac{m}{2} \cdot \frac{c_1}{m \lg n} \geq \frac{c_3}{\lg n},$$

where c_3 is a constant.

Since $Dist(x,t) \geq \lg^2 n \lg\lg n$ and each AL-link spans a distance no more than $\lg^2 n$, the nodes in AL awareness $A_x(\lg\lg n)$ are all between the current node x and the target node t. Let $\Pr[A_x(\lg\lg n) \xrightarrow{K} B_{\frac{m}{2}}(t)]$ denote the probability that at least one node in $A_x(\lg\lg n)$ has a K-neighbor in $B_{\frac{m}{2}}(t)$. Then we have

$$\begin{aligned}
\Pr[A_x(\lg\lg n) \xrightarrow{K} B_{\frac{m}{2}}(t)] &\geq \Pr[A_x(\lg\lg n) \xrightarrow{K} B_{\frac{m}{2}}(t) \mid \mathcal{A}] \cdot \Pr[\mathcal{A}] \\
&\geq \big(1 - (1 - \frac{c_3}{\lg n})^{\frac{\lg n}{\sigma}}\big) \cdot \psi \\
&\geq \psi(1 - e^{-\frac{c_3}{\sigma}}),
\end{aligned}$$

which is larger than a positive constant. At the last step, we obtain $(1 - \frac{c_3}{\lg n})^{\frac{\lg n}{\sigma}} \leq e^{-\frac{c_3}{\sigma}}$ by using the fact that $(1 + \frac{b}{x})^x \leq e^b$ for $b \in \mathbb{R}$ and $x > 0$. ■

Lemma 4. *Suppose that the distance between the current node x and the target node t in a KSWN* $\mathcal{K}^*$ is $Dist(x,t) \geq \lg^2 n \lg\lg n$. Then after at most $O(\lg n \lg\lg n)$ expected number of hops, Algorithm 1 will reduce the distance to within $\lg^2 n \lg\lg n$.*

Proof: Since $Dist(x,t) \geq \lg^2 n \lg\lg n$, all known nodes in x's AL awareness $A_x(\lg\lg n)$ are between the current node x and the target node t. We can apply the result in Lemma 3 to analyze Algorithm 1.

We refer to the routing steps from a given node x to any node within $A_x(\lg\lg n)$ as an indirect phase. The routings in different indirect phases are independent from each other. By Lemma 3, the probability that node x's AL awareness $A_x(\lg\lg n)$

contains a K-neighbor within $Dist(x,t)/2$ distance to the target node t is at least a positive constant, so after at most $O(1)$ expected number of indirect phases, Algorithm 1 will find an intermediate node whose K-link jumps over half distance. Since each indirect phase takes at most $\lg\lg n$ hops and the maximum distance is n, after at most $O(\lg n\ \lg\lg n)$ expected number of hops, the message will reach a node within $\lg^2 n \lg\lg n$ distance to the target node t. ■

Lemma 5. *Suppose that the distance between the current node x and the target node t in a *KSWN *$\mathcal{K}$ is $Dist(x,t) \leq \lg^2 n \lg\lg n$. Then using the final phase of Algorithm 1 (i.e. using Kleinberg's greedy algorithm) can route the message to the target node t in $O(\lg n)$ expected number of hops.*

Proof: When the distance $Dist(x,t) \leq \lg^2 n \lg\lg n$, the final phase in Algorithm 1 is executed. By Kleinberg's results in [13], after at most $O\big(\lg^2(\lg^2 n \lg\lg n)\big) = O(\log n)$ expected number of steps, the message will be routed to the destination node. ■

Combining the above lemmas, it is not difficult for us to obtain the routing complexity of Algorithm 1.

Theorem 1. *In a KSWN* $\mathcal{K}^*$, Algorithm 1 performs in $O(\lg n\ \lg\lg n)$ expected number of hops.*

4.3 Oblivious Decentralized Routing

In our oblivious scheme, when the distance is large, the current node x first finds in $A_x(\lg\lg n)$ whether there is an intermediate node z, which contains a K-neighbor within $Dist(x,t)/2$ distance to the target node, and is closest to node x in terms of AL-links (any possible tie is broken arbitrarily). Next, node x computes a shortest path π from x to z among the AL awareness $A_x(\lg\lg n)$, and then routes the message to its next AL-neighbor on the shortest path π. When the distance is small, Kleinberg's plain greedy algorithm is applied.

Given a message M, a source s and a target t in a KSWN* $\mathcal{K}^*$, the pseudocodes of our oblivious algorithm running on the current node x are given in Algorithm 2.

Lemma 6. *Suppose that the distance between the current node x and the target node t in a KSWN* $\mathcal{K}^*$ is $Dist(x,t) \geq c(\lg n)^2 \lg\lg n$, where c is a sufficiently large constant. Then after at most $O(\lg\lg n)$ expected number of hops, Algorithm 2 will reduce the distance to within $Dist(x,t)/2$.*

Algorithm 2

Input: the source s, the target t and the message M.
Initialization:
Current node $\leftarrow s$.
while Distance between the current node and the destination $\geq c(\lg n)^2 \lg\lg n$ **do** (c is a sufficiently large constant and will be given later)
$z \leftarrow$ a node in $A_x(\lg\lg n)$ that contains a K-neighbor within $Dist(x,t)/2$ distance to t, and is closest to node x in terms of AL-links (ties are broken arbitrarily).
if node z does not exist **then**
Route the message M to an immediate neighbor closest to node t.
else
Compute a shortest path π from x to z among $A_x(\lg\lg n)$.
if π consists of only node x itself **then**
Route the message M to the K-neighbor.
else
Route the message M to the next AL-neighbor on the shortest path π.
end if
end if
end while
Final phase (Kleinberg's greedy algorithm):
Route the message M to an immediate neighbor of x that is closest to the target t, until it reaches t.

Proof: As shown in Figure 1, node r is the midpoint of $\overline{xt}$, and node r' is between r and t such that $Dist(r,r') = \lg^2 n \lg\lg n$. Let z be an intermediate node in $A_x(\lg\lg n)$ that contains a K-neighbor between r and t, and is closest to x in terms of AL-links. We refer to a node z in x's AL awareness $A_x(\lg\lg n)$ as a *good* intermediate node if it satisfies the following two conditions: (1) has a K-neighbor within $Dist(x,t)/2$ to the target node; (2) is closest to node x in terms of AL-links. Let $\pi : x_0 = x, x_1, \cdots, x_t = z$ denote a shortest path that x finds from itself to z among the AL awareness $A_x(\lg\lg n)$. We divide the next routing into two cases according to the different locations of z's K-neighbor.

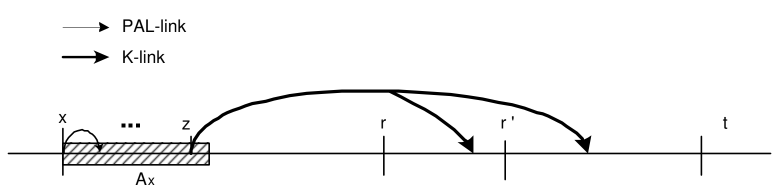

Fig. 1. Diagram for oblivious decentralized routing. The shade area represents node x's AL awareness $A_x(\lg\lg n)$. The target node t is on the right side of x. Node r is the midpoint of $\overline{xt}$. Node r' is between nodes r and t such that $Dist(r,r') = \lg^2 n \ \lg\lg n$. Node z is an intermediate node in $A_x(\lg\lg n)$ that contains a K-neighbor in $\overline{rt}$ (in $\overline{rr'}$ or $\overline{r't}$) and is closest to x in terms of AL-links.

In the first case, z's K-neighbor is within $\overline{r't}$. Since the distance between x and the right most node in $A_x(\lg\lg n)$ is at most $\lg^2 n \lg\lg n$, z's K-neighbor is also within $Dist(x_i,t)/2$ to the target node for every x_i on the shortest path π, that is, node z always satisfies the first condition of a good intermediate node for every node x_i. Also, if z is an intermediate node closest to x, it is also a closest intermediate node to every x_i on the shortest path π, that is, z also satisfies the second condition of a good intermediate node for every node x_i. Therefore, node z will become a fixed good intermediate node for all nodes x_i on the shortest path. When this case happens, Algorithm 2 will route the message along a shortest path from x to z in an oblivious routing fashion. Thus, in this case, after at most $\lg\lg n$ number of hops, the message will reach a good intermediate node and the routing distance will be reduced by half [5]. In the second case, z's K-neighbor is within $\overline{rr'}$. When this happens, the intermediate node z may change for each x_i on the shortest path $\pi : x_0 = x, x_1, \cdots, x_t = z$, and the message may not be routed along the shortest path as expected by the previous node x. However, we will show that the latter case will not happen very likely, since the length of $\overline{rr'}$ is relatively small when $Dist(x,t) \geq c(\lg n)^2 \lg\lg n$ for a sufficiently large constant c.

Let $\mathcal{F}_1$ denote the event that $A_x(\lg\lg n)$ contains a K-neighbor in $\overline{r't}$. By using a similar technique in Lemma 3, we can easily obtain that $\mathcal{F}_1$ occurs with probability at least a positive constant.

Let $\mathcal{F}_2$ denote the event that $A_x(\lg\lg n)$ contains a K-neighbor in $\overline{rr'}$. For any node y in $A_x(\lg\lg n)$, we have $Dist(y,r) \geq \frac{1}{3}c(\lg n)^2 \lg\lg n$ when c is a sufficiently large constant. By Lemma 2, the probability for a node y in $A_x(\lg\lg n)$ to send a K-link to a node in $\overline{rr'}$ is at most $\frac{3c_2}{c(\lg n)^2(\lg\lg n)\cdot\lg n}$. Because there are in total $\lg^2 n \lg\lg n$ nodes in $\overline{rr'}$, a node in $A_x(\lg\lg n)$ has a K-neighbor in $\overline{rr'}$ with probability at most $\frac{3c_2}{c(\lg n)^2(\lg\lg n)\cdot\lg n} \cdot \lg^2 n \lg\lg n = \frac{3c_2}{c\lg n}$. Since $|A_x(\lg\lg n)| \leq 1 + 2 + 2^2 + \cdots + 2^{\lg\lg n} \leq 2\lg n$, the event $\mathcal{F}_2$, i.e., $A_x(\lg\lg n)$ has a K-neighbor in $\overline{rr'}$, occurs with probability at most $\frac{3c_2}{c\lg n} \cdot 2\lg n = \frac{6c_2}{c}$, which is smaller than a certain constant when c is a sufficiently large constant. Thus, we have $Pr[\neg\mathcal{F}_2] > \gamma$ for a constant $\gamma > 0$, if we choose a sufficiently large constant c.

Therefore, $\Pr[\neg\mathcal{F}_2 \bigcap \mathcal{F}_1]$ is larger than a positive constant, if we choose a sufficiently large constant c. Thus, after at most $c' \lg\lg n$ expected number of hops for a constant c', the event $\neg\mathcal{F}_2 \bigcap \mathcal{F}_1$ will occur, that is, a message will be routed to a node x whose AL awareness $A_x(\lg\lg n)$ contains a K-neighbor in $\overline{r't}$, but no K-neighbor in $\overline{rr'}$. When such a node x is reached, the intermediate node z is fixed for every node x_i on a shortest path $\pi : x_0 = x, x_1, \cdots, x_t = z$ in an oblivious routing fashion. Then after at most $\lg\lg n$ number of hops, the message will be

[5] There may be more than one good intermediate nodes z when a tie happens. However, even when this happens, the message will still reach one of good intermediate nodes along a shortest path finally. Hereinafter, we focus on the case in which the good intermediate node z is unique for the current node x. The analysis for the case with multiple good intermediate nodes can be easily obtained.

routed to the fixed intermediate node z, which has a K-link jumping over half distance.

Therefore, after at most $c' \lg\lg n + \lg\lg n = O(\lg\lg n)$ expected number of hops, the distance will be reduced by half. ■

Lemma 7. *Suppose that the distance between the current node x and the target node t in a KSWN* $\mathcal{K}^*$ *is $Dist(x,t) \geq c\lg^2 n \lg\lg n$, where c is a sufficiently large constant. Then after at most $O(\lg n \lg\lg n)$ expected number of hops, Algorithm 2 will reduce the distance to within $c\lg^2 n \lg\lg n$.*

Proof: The proof is similar to that of Lemma 4, and hence is omitted here. ■

Lemma 8. *Suppose that the distance between current node x and the target node t in a KSWN* $\mathcal{K}^*$ *is $m < c(\lg n)^2 \lg\lg n$, where c is a sufficiently large constant. Then using the final phase of Algorithm 2 (i.e. using Kleinberg's greedy algorithm) can route the message to the target node t in $O(\lg n)$ expected number of hops.*

Proof: The proof is similar to that of Lemma 5, and hence is omitted here. ■

Combining the above lemmas, we can easily obtain the following theorem.

Theorem 2. *In a KSWN* $\mathcal{K}^*$*, Algorithm 2 performs in $O(\lg n \;\lg\lg n)$ expected number of hops.*

5 Experimental Evaluation

In this section, we will conduct experiments to evaluate our schemes and other existing routing schemes for Kleinberg's small-world networks.

We focus on the following four schemes: (a) The original greedy routing algorithm [13] in Kleinberg's small-world network with only one long-range contact per node. Each node forwards the message to its immediate neighbor closest to the destination; (b) The greedy routing algorithm in Kleiberg's small-world network with two long-range contacts per node [2, 15]. In the experimental study, we would like to learn how much the additional number of long-range links can help routing. (c) The decentralized routing scheme with $O(\lg n)$ local awareness [10, 17]. With this scheme, we intend to evaluate the degree at which the local awareness improve the routing efficiency. (d) Our near optimal routing scheme proposed in this paper. We note that most schemes have both non-oblivious and oblivious versions. Here we only focus on the non-oblivious version for each scheme.

5.1 Experimental Setup

Network Construction: We construct the small-world network based on a ring $[0, 1, \cdots, n]$. Each node i is connected to its immediate neighbors $(i+1) \mod n$. Let $H_n = \sum_{i=1}^{n} 1/i$ denote the harmonic normalization factor. We then generate a sequence of intervals $I(i)$, which we call the probability intervals, where $1 \le i \le n-1$. Let $0 < I_1 \le 1/H_n$, and $\frac{1}{(i-1)H_n} < I_i \le \frac{1}{iH_n}$, where $2 \le i \le n-1$. Each node i uniformly generates a random number x in $(0, 1]$, and then finds the interval that contains x. Suppose that x is located in the interval I_k. Node i then forms a long-range link connected to a node with the distance k. When each node has multiple long-range contacts, it just generates more than one random numbers, and sets up the connections in the same way.

In the extension of Kleinberg's small-world networks, each node uniformly and randomly chooses two nodes within the Manhattan distance $\lg^2 n$ as its augmented local neighbors.

Messages Generation and Evaluation Metrics: We let each node generate a query message with a random destination, and then evaluate the following metrics.

(1) **Average length of routing path** is the average number of hops travelled by the messages from the source to the destination.
(2) **Storage requirement** for each node is the number of information bits required to be stored on each node.

5.2 Experimental Results

We vary the number of nodes in the network from 5,000 to 25,000, and evaluate different routing schemes, as shown in Figures 2 and 3. For large n, the greedy algorithm with increasing number of long-range contacts [2, 15], the decentralized routing algorithm with local awareness [10, 17], our near optimal and algorithm all improve Kleinberg's original greedy algorithm. Our near optimal scheme can find a shorter routing path than the decentralized routing schemes with local awareness [10, 17], while keep almost the same storage space on each node.

6 Conclusion

We extend Kleinberg's small-world network with augmented local links, and show that if each node participating in routing is aware of $O(\lg n)$ neighbors via augmented links, there exist both non-oblivious and oblivious decentralized algorithms that can finish routing in $O(\lg n \lg \lg n)$ expected number of hops, which is a near

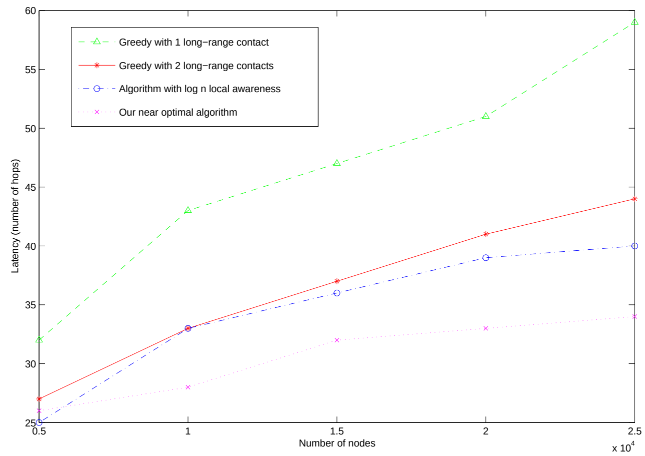

Fig. 2. Average length of routing paths for different routing schemes.

optimal routing complexity. Our investigation shows that the awareness of $O(\lg n)$ nodes through the augmented links will be more efficient for routing than via the local links [10, 17].

Our extended model may provide an important supplement for the modelling of small-world phenomenon, and may better approximate the real-world observation. For example, each person in a human society is very likely to increase his/her activities randomly within some certain communities, and thus is aware of certain levels of "augmented" acquaintances. This augmented awareness would surely help delivery the message to an unknown target in the society.

Our results may also find applications in the design of large-scale distributed networks, such as distributed storage systems. Unlike most existing deterministic frameworks for distributed systems, our extended small-world networks may provide good fault tolerance, since the links in the networks are constructed probabilistically and less vulnerable to adversarial attacks.

References

1. R. Albert, H. Jeong, and A.-L. Barabasi. The diameter of the World Wide Web. *Nature*, 401(9):130–131, 1999.
2. J. Aspnes, Z. Diamadi, and G. Shah. Fault-tolerant routing in peer-to-peer systems. In *Proceedings of PODC'02*, pages 223–232, 2002.
3. L. Barriére, P. Fraigniaud, E. Kranakis, and D. Krizanc. Efficient routing in networks with long range contacts. In *Proceedings of the 15th International Symposium on Distributed Computing (DISC'01)*, pages 270–784, 2001.
4. I. Benjamini and N. Berger. The Diameter of Long-Range Percolation Clusters on Finite Cycles. *Random Structures and Algorithms*, 19(2):102–111, 2001.

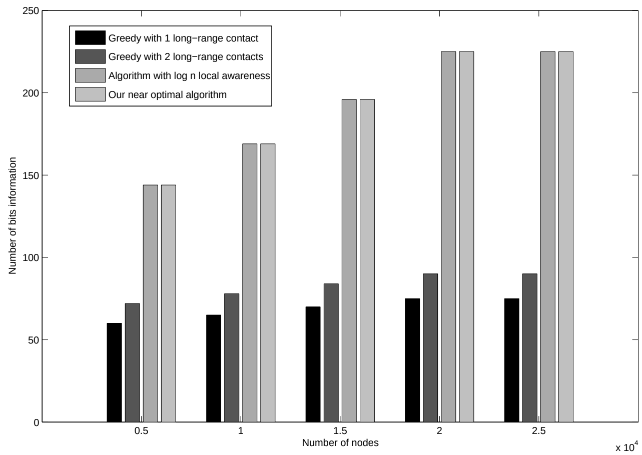

Fig. 3. Storage requirement on each node for different routing schemes.

5. M. Biskup. Graph diameter in long-range percolation. *Submitted to Electron. Comm. Probab*, 2004.
6. B. Bollobás and F.R.K. Chung. The Diameter of a cycle plus a random matching. *SIAM Journal on Discrete Mathematics*, 1(3):328–333, 1988.
7. D. Coppersmith, D. Gamarnik, and M. Sviridenko. The diameter of a long-range percolation graph. *Random Structures and Algorithms*, 21(1):1–13, 2002.
8. P. Dodds, R. Muhamad, and D. Watts. An experimental study of search in global social networks. *Science*, 301:827–829, 2003.
9. M. Kochen Ed. *The small world (Ablex, Norwood)*. 1989.
10. P. Fraigniaud, C. Gavoille, and C. Paul. Eclecticism shrinks even small worlds. In *PODC 2004*, pages 169–178, 2004.
11. C.M. Homan and G. Istrate. Small worlds, locality, and flooding on landscapes. Research Report TR-2003-796, Department of Computer Science, University of Rochester, USA, 2003.
12. D. Kemper, J. Kleinberg, and A. Demers. Spatial gossip and resource location protocols. In *Proceedings of STOC*, pages 163–172, 2001.
13. J. Kleinberg. The Small-World Phenomenon: An Algorithmic Perspective. In *Proceedings of the 32nd ACM Symposium on Theory of Computing*, pages 163–170, 2000.
14. E. Lebhar and N. Schabanel. Almost optimal decentralized routing in long-range contact networks. In *ICALP 2004*, pages 894–905, 2004.
15. G. S. Manku, M. Bawa, and P. Raghavan. Symphony: Distributed hashing in a small world. In *Proceedings of the 4th USENIX Symposium on Internet Technologies and Systems*, pages 127–140, 2003.
16. G.S. Manku, M. Naor, and U. Wieder. Know thy neighbor's neighbor: The power of lookahead in randomized p2p networks. In *Proceedings of STOC 2004*, pages 54–63, 2004.
17. C. Martel and V. Nguyen. Analyzing Kleinberg's (and other) small-world models. In *PODC 2004*, pages 179–188, 2004.
18. S. Milgram. The small world problem. *Psychology Today*, 61, 1967.
19. M.E.J. Newman. Models of the small world. *J. Stat. Phys.*, 101, 2000.
20. V. Nguyen and C. Martel. Analyzing and Characterizing Small-World Graphs. In *SODA 2005*, 2005.
21. D. Watts and S. Strogatz. Collective dynamics of small-world networks. *Nature*, 393:440–442, 1998.

22. H. Zhang, A. Goel, and R. Govindan. Using the small-world model to improve Freenet performance. In *Proceedings of IEEE INFOCOM 2002*, pages 1228–1237, 2002.

Appendix A. Proof of Lemma 2

Lemma 2. *Let* $\Pr[u \xrightarrow{K} v]$ *denote the probability that node* u *sends a K-link to node* v *in a KSWN** $\mathcal{K}^*$. *Suppose that* $a \leq Dist(u,v) \leq b$, *then* $\frac{c_1}{b \lg n} \leq \Pr[u \xrightarrow{K} v] \leq \frac{c_2}{a \lg n}$, *where* c_1 *and* c_2 *are constants independent of* n.

Proof: The probability that node v is a K-neighbor of node u is $\Pr[u \xrightarrow{K} v] = \frac{1}{Dist(u,v) Z_v}$, where $Dist(u,v)$ is the ring distance between nodes u and v, and $Z_v = \sum_{z \neq v} \frac{1}{Dist(v,z)}$.

Observe that $Z_v = \sum_{i=1}^{n} \frac{|U_i|}{i}$, where $|U_i|$ is the set of all nodes at distance i away to node v. Since $|U_i| = \Theta(1)$, we have $Z_v = \sum_{i=1}^{n} \frac{\Theta(1)}{i} = \Theta(\lg n)$.

Since $a \leq Dist(u,v) \leq b$, we have $\frac{c_1}{b \lg n} < \Pr[u \xrightarrow{K} v] < \frac{c_2}{a \lg n}$, for some constants c_1 and c_2 independent of n. Thus the lemma follows. ∎