Mathematical Modeling of Routing in DHTs

Peter Kersch and Robert Szabo

Abstract Although most Distributed Hash Table (DHT) overlays are structurally similar to the "small-world" navigation model of Kleinberg – architectural and algorithmic details of different DHT variants differ significantly. Lookup performance of DHTs depends on a sets of different and often incompatible parameters, which makes analytical comparison rather difficult. The objective of this chapter is to review existing analytical models for DHT routing performance and to introduce a novel framework for the per-hop routing progress analysis of long-range DHT connections based on a logarithmic transformation. With the logarithmic transformation of the DHT metric space we analyze the distribution of the per-hop routing progress in general and also for the special cases of the deterministic and probabilistic power-law routing overlays. Based on the proposed analytical framework routing performance of DHTs can be described by a triple: the long-range connection density, its coefficient of variation and the number of short-range connections. Finally, we derive upper bound on the expected number of routing hops as a function of network size and the parameter triple.

1 Introduction

A plethora of Distributed Hash Table (DHT) concepts have been proposed and analyzed in the past 6–7 years to provide scalable and robust distributed storage and

Peter Kersch
High Speed Networks Laboratory, Department of Telecommunications and Media Informatics, Budapest University of Technology and Economics, Budapest, Hungary,
e-mail: kersch@tmit.bme.hu

Robert Szabo
High Speed Networks Laboratory, Department of Telecommunications and Media Informatics, Budapest University of Technology and Economics, Budapest, Hungary,
e-mail: szabo@tmit.bme.hu

X. Shen et al. (eds.), *Handbook of Peer-to-Peer Networking,*
DOI 10.1007/978-0-387-09751-0_14,

lookup systems [6, 12, 14, 16, 17, 19–22], etc. Although architectural and algorithmic details of these DHT proposals can differ significantly, the foundations of lookup mechanisms are very similar for most of them. There are several empirical studies (based on simulations) comparing static and dynamic performance of different DHT routing mechanisms using various parameter settings [5, 13]. There exist also detailed analytical models for some DHTs, however these models are usually restricted to one specific DHT implementation. Finally, some aspects of DHT routing are covered by generic models, e.g., static resilience of DHT routing against failure [10] or the impact of lookup strategy, lookup parallelism and replication on DHT routing performance under churn [23]. However, to the best of our knowledge, there exist no generic analytical models capturing the relationship between overlay structure and routing performance of DHTs in static networks. In this chapter, we try to fill this gap proposing a generic stochastic model of DHT overlays and overlay routing covering a large family of DHTs.

The proposed analytical model builds on the fact that most DHT overlays are structurally similar to the "small-world" model of Kleinberg [8] and the sequence of long-range connections of a DHT node becomes linear after logarithmic transformation of distances in the DHT metric space. More specifically, we have identified a large subclass of DHT overlays (regular power-law routing overlays) where this transformed sequence can be described for each node as independently selected random samples from an infinite renewal process. Using this renewal process model, we analyze the distribution of the per-hop routing progress in general and also for the special cases of the deterministic and probabilistic power-law routing overlays. Furthermore, we introduce the λ long-range connection density and the c_v long-range connection density coefficient of variation parameters to characterize long-range connection distribution of an overlay. Finally, using renewal theory, we derive upper bounds on the expected number of routing hops as a function of network size and the above overlay parameters.

The rest of this chapter is structured as follows. First, we give a brief overview of DHTs in general. Then, in Section 2, we discuss challenges of modeling DHT routing, revisit applied mathematical tools from renewal theory and introduce modeling assumptions and notations used in the upcoming sections. In Section 3, we present the concept of logarithmically transformed view for long-range connections. Finally, in Section 4, we describe the proposed stochastic model based on this transformed view and tools from renewal theory.

1.1 DHTs Revisited

From the point of view of an application, Distributed Hash Tables provide similar functionality than ordinary "in memory" hash tables. An application can insert and remove key-value mappings, and given a key, it can retrieve the associated value (in the context of a peer-to-peer system, a key is an identifier used to refer to a shared resource while the associated value is the resource itself or the locator of the re-

source). All of these operations are performed quickly and efficiently and scale well for large amounts of data in both "in memory" and distributed hash tables. However, as opposed to ordinary hash tables, storage of key-value pairs is distributed over all nodes of the DHT and all hash table methods can be issued from any of these nodes (see Fig. 1). Consequently, internal operation of a DHT differs significantly from the operation of ordinary "in memory" hash tables. To present the architecture and operation of distributed hash tables, we used the terminology and formalism proposed in [1].

One of the key conceptual components of a DHT is the common metric space into which nodes and resources are mapped to. All distributed hash tables use a virtual identifier space $\mathscr{I}$ which possesses a closeness metric $d : \mathscr{I} \times \mathscr{I} \to \mathbf{R}$ so that $(\mathscr{I}, d)$ is a metric space or a quasi-metric space[1] Both the group of peers forming the DHT and the set of all shared resources are mapped to this ID space $\mathscr{I}$ (see Fig. 1). Mapping of peers can be described by a function $F_P : \mathscr{P} \to \mathscr{I}$ where $\mathscr{P}$ is the set of peers forming the DHT. F_P is usually implemented by either drawing a random identifier according to uniform distribution over $\mathscr{I}$ or by applying a hash function to the public key of the peer. Resources are mapped to $\mathscr{I}$ using a function $F_K : \mathscr{K} \to \mathscr{I}$ where K is the set of keys used to refer to shared resources. F_K is most often implemented by applying a hash function to the keys.

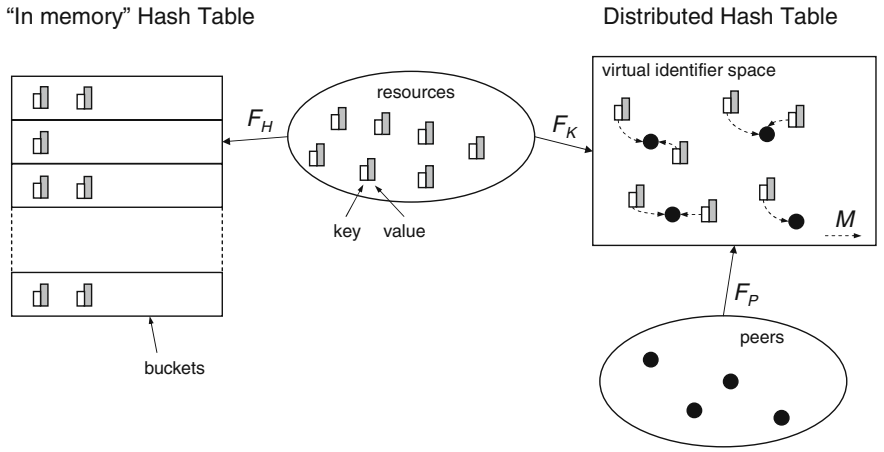

Fig. 1 Comparison of "in memory" hash tables and DHTs

Peers responsible for a given resource are determined based on the above mappings to the common metric space $(\mathscr{I}, d)$. A key-value pair describing a resource is usually stored by the peer (or the set of peers) whose image in $(\mathscr{I}, d)$ is the closest to the image of the given resource in $(\mathscr{I}, d)$. Formally, this can be described using a function $\mathscr{M} : \mathscr{I} \to 2^{\mathscr{P}}$ and a constraint $\forall i \in \mathscr{I} : \forall p \in \mathscr{M}(i), \forall q \notin \mathscr{M}(i)$:

[1] A quasi-metric space does not satisfy the symmetry requirement of metric spaces.

$d(F_P(p),i) \leq d(F_P(q),i)$ on this function. As a result, locating a key-value pair (which describes a shared resource in a DHT) corresponds to finding one of the closest peers to the image of the resource in $(\mathscr{I},d)$.

The function $\mathscr{M}$ is usually complete, which means that each identifier of $\mathscr{I}$ is under the responsibility of at least one peer. To provide fault tolerance $\mathscr{M}$ typically contains more than one element and the cardinality of $\mathscr{M}$ is typically constant, which means that each key-value pair is replicated to the same number of peers (Fig. 1 shows the simplest case when each key-value pair is stored by only one peer).

Comparing distributed hash tables to ordinary hash tables, peers correspond to buckets and the function $\mathscr{M}$ corresponds to the hash function F_H in ordinary hash tables. Changing the number of buckets in an ordinary hash table implies changing the hash function F_H and this usually requires relocation of most key-value pairs. For "in memory" hash tables, bucket size is usually constant (or changes only rarely when reaching a capacity threshold), hence this is not a problem. In contrast, a peer-to-peer network is inherently dynamic and the set of peers in a DHT might change continuously, implying changes in the mapping of key-value pairs to nodes ($\mathscr{M}$). Continuous relocation of key-value pairs in a DHT would generate a huge communication overhead, hence changes in these mappings should be minimized when peers join and leave the network. DHTs address this problem by selecting responsible peers based on proximity in the metric space $(\mathscr{I},d)$ as mentioned above. Consequently, changes in key-value pair $\rightarrow$ responsible peers mappings are restricted to the neighborhood of the joining or leaving peer in $(\mathscr{I},d)$. (This concept is also called *consistent hashing* [7] and had been proposed for distributed web caching before the era of distributed hash tables.)

Another benefit of selecting responsible peers by proximity in $(\mathscr{I},d)$ is that all DHT operations can be easily implemented on top of a routing algorithm which locates the closest peers to a given point in $\mathscr{I}$. To realize this routing process, DHTs create and maintain an overlay network. An overlay network can be modeled by a directed graph $G = (\mathscr{P},\mathscr{E})$ where $\mathscr{P}$ denotes the set of vertices (peers) while $\mathscr{E}$ denotes the set of edges (overlay connections[2]). Routing in the overlay is typically based on a simple greedy algorithm: a request for a given point in $\mathscr{I}$ is forwarded via the connection pointing to the peer which is the closest to this given point in $(\mathscr{I},d)$.

Overlay topology depends heavily on distances between the images of peers in the metric space $(\mathscr{I},d)$. In most DHT overlays, connections can be categorized into short-range (local) and long-range connections. Each node has short-range connections to some specific subset of the closest peers in $(\mathscr{I},d)$. Additionally, they have long-range connections to some distant nodes so that the distribution of these connections is structurally similar to the family of small-worlds graphs introduced by Kleinberg in [8]. In this small-worlds graph family, the probability of having a long-

[2] A connection from node v_1 to a peer node v_2 means that node v_1 knows the address of node v_2 (this is usually in the form of a pair of ID + IP address / port number). Different algorithms use different names for connections, e.g., Chord [22] calls them successors, predecessors and finger pointers, while in Pastry [21], they are called leaf set and routing table entries, etc.

range connection between to nodes is inversely proportional to the Dth power of their distance (where D is the dimension of the metric space), and Kleinberg has shown that this is necessary to provide efficient distributed search based solely on local information.

The role of short-range and long-range connections in the overlay is complementary. Short-range connections guarantee success of greedy forwarding: since each node is connected to its closest neighbors in $(\mathscr{I},d)$, it is always possible to forward requests at least a small step closer to the target. In contrast, long-range connections are not critical for successful routing but they expedite the lookup process and usually provide $O(\log n)$ bounds on the average number of lookup hops. This is achieved by ensuring that the distance from the target decreases by a constant factor in expected value after each routing step.

2 Modeling DHT Routing

Defining models and metrics to describe performance of different DHT routing architectures is not a trivial task. An application using the DHT lookup service is mostly interested in lookup latencies and in the ratio of successful lookups. A user running DHT implementations might also be concerned by resource usage (CPU, memory, storage, bandwidth, etc...) while a network operator is only interested in the overall traffic (lookup + control) generated in the network. Since most of these describe conflicting objectives, comparison only makes sense if conflicting performance metrics are analyzed together describing fundamental trade-offs.

Some of the commonly used performance metrics (e.g., overlay network diameter, node state) are not directly relevant for neither applications nor users nor network operators. In [25], the author investigates the trade off between node state and overlay network diameter. Loguinov et al. also use network diameter as the primary metric for routing in [15].

Node state affects primarily memory usage at nodes. However, the amount of memory required to keep track of connections is typically far from being a bottleneck in current systems. Node state can also influence the maintenance bandwidth (e.g., in DHTs using per connection periodic keep-alive messages to detect connection failures). However, it cannot be used as a general metric to characterize maintenance traffic.

Overlay network diameter can be used to derive only lower bounds on the worst-case number of routing hops for a lookup in a given overlay structure. Short paths between nodes do not guarantee that a distributed routing algorithm is also able to find them [8]. Hence, the distribution or the average number of routing hops is a more informative performance metric which also allows to derive [23] lookup latency – a key performance metric from a user perspective.

Analytical comparison of a performance metric (e.g., the number of routing hops) of different DHTs is usually described by asymptotic notation, commonly used to characterize algorithm complexity. E.g., CAN [19] with a D dimensional identifier

space provides lookups in $O(\frac{D}{2}n^{\frac{1}{D}})$ hops in a network of n nodes. Although this is a useful and simple way to determine scalability of a particular algorithm, it has its limitations. Due to potentially different unknown constants hidden within the notation, it is not possible to compare two different algorithms with the same asymptotic behavior (e.g., $O(\log n)$ hop count is typical for many DHTs). Furthermore, it is also possible that an algorithm with better asymptotic behavior performs worse for practical network sizes.

Asymptotic notation may even be misleading when not used carefully. The paper presenting Koorde [6] (a DHT based on de Bruijn graphs) is a good example of such a misuse. Using a base-k de Bruijn graph, Koorde completes routing in $O(log_k\ n)$ hops. Based on this, the authors claim that choosing $k = \log n$, routing cost is $O(\log n/\log k) = O(\log n/\log\log n)$. However, the base of the underlying de Bruijn graph cannot be changed on the fly as the network grows since this would require rebuilding the whole DHT from the scratch. Therefore the parameter k should not be treated as a function of network size. (Similarly, the dimension D of a CAN [19] network is not expressed as a function of network size because this is also a parameter that cannot be changed without rebuilding the whole system.) As a consequence, the number of routing hops for Koorde using base-k de Bruijn graphs is in fact $O(\log n/\log k)$.

For a few DHT architectures, there are some exact analytical results: e.g., the average number of routing hops for Chord [22] is $\frac{1}{2}log_2 n$. [23] is one of the few papers which provide a generic analytical framework for the performance comparison of different DHTs. Given the average number of routing hops in static networks, the authors analyze the influence of three key factors on routing performance under churn: lookup strategy, lookup parallelism and replication. Our results on the expected number of routing hops in static networks can be potentially used as an input for this analytical framework to derive these additional performance metrics.

Finally – although not directly related to distributed hash tables – the "small-world" navigation model of Kleinberg [8] is a fundamental contribution to theory of routing in distributed systems. A network is said to be "small-world" when there exists a short path between any two nodes, although most nodes are not directly connected. This low network diameter is a necessary but not sufficient property for efficient distributed routing. In [8], Kleinberg investigates requirements on overlay topology for efficient distributed routing based solely on local information in small-world networks. Similarly to DHT overlays, he defines a graph (embedded into a metric space) with short-range connections to the closest nodes and long-range connection(s) to some distant nodes. As in DHTs, Kleinberg's routing is greedy: requests are forwarded via the peer node being the closest to the target node in the metric space. As the main finding of the paper, Kleinberg shows that distributed routing will achieve the best asymptotical performance when the probability of having a long-range connection to another node is inversely proportional to the D^{th} power of distance of the two nodes (where D is the dimension of the metric space embedding the small-world graph).

Most DHT routing architectures – although not inspired by Kleinberg's work – can be related to the one dimensional Kleinberg small-world model.

2.1 Renewal Processes Revisited

Renewal processes are a special class of stochastic processes used to model independent identically distributed occurrences. Let $X_1, X_2, X_3, ...$ be independent identically distributed (*i.i.d*) and positive random variables defined by the distribution function $P(X < x) = F(x)$. Furthermore, let T_n be defined as $T_n = \sum_{i=1}^{n} X_i$. Then the counting process $Y(t) = \max\{n : T_n \leq t\}$ is a renewal process ($t \geq 0$).

Renewal processes are usually defined in the time domain. In the time domain, $Y(t)$ denotes the number of events until time t, T_n corresponds to the occurrence time of the n^{th} event and the random variables X_i correspond to inter-arrival times between subsequent events. The name renewal process is motivated by the fact that every time there is an occurrence, the process "starts all over again"; it renews itself (since the variables X_i are *i.i.d*).

In contrast to the general usage, renewal processes in my dissertation are not defined in the time domain but in the distance domain of a one dimensional metric space. Furthermore occurrences are not events but the images of long-range connections in this metric space and the random variables X_i correspond to distances between the images of subsequent long-range connections.

In the followings, I briefly list the results of renewal theory that I use in the upcoming sections (for further reading, see [4, 9, 11]). Note that the vocabulary of renewal theory traditionally assumes a time domain for renewal processes. However, all results are equally valid for the distance domain too.

Renewal function The expected value of the number of arrivals in function of the elapsed time is called renewal function: $m(t) = E[Y(t)]$.

Residual life Picking a random point in time (t), the random variable corresponding to the time from this point until the next event (at time $T_{Y(t)+1}$) in a renewal process is called residual life:

$$V(t) = T_{Y(t)+1} - t \tag{1}$$

Residual life is also called *residual lifetime*, *residual time* or *forward recurrence time*.

Expected value of asymptotic residual life The expected value of asymptotic residual life in a renewal process can be expressed as

$$\lim_{t\to\infty} E[v] = \frac{\mu_2}{2\mu}. \tag{2}$$

where $\mu = E[x]$ is the expected value of inter-arrival times and $\mu_2 = E[x^2]$ is the second moment of inter-arrival times.

Distribution of asymptotic residual life Considering a renewal process with an inter-arrival time distribution $F(x)$, the probability density function of asymptotic residual life can be expressed as

$$\lim_{t\to\infty} g(v) = \frac{1 - F(v)}{\mu}, \tag{3}$$

where $\mu = E[x]$ is the expected value of inter-arrival times.

Length of a randomly selected renewal period Picking a random point in time (t) in a renewal process, the *pdf* of the length of the renewal period marked by this point ($T_{Y(t)+1}$ - $T_{Y(t)}$) is asymptotically:

$$\lim_{t\to\infty} h(x') = \frac{f(x')x'}{\mu}, \tag{4}$$

where $f(x)$ is the *pdf* of inter-arrival times and $\mu = E[x]$ is the expected value of inter-arrival times in the renewal process. It is important to note that the distribution of x' and x are not the same since a random point in time will select longer periods at higher probability than shorter periods.

Note that considering a random sample from a renewal process, the above formulas are also valid in general, not only for the asymptotic case.

Lorden bound The renewal function of a renewal process is upper bounded by

$$m(t) \leq \frac{t}{\mu} + \frac{\mu_2}{\mu^2} + 1, \tag{5}$$

where μ is the expected value of inter-arrival times and μ_2 is the second moment of inter-arrival times in the renewal process (see [4], page 110.)

Poisson processes are a special class of renewal processes. Inter-arrival times in a Poisson process are exponentially distributed. A Poisson process can be characterized by the λ parameter of this exponential distribution which is also called the intensity of the process.

A Poisson process of intensity λ can also be defined as a pure birth process: the probability that an arrival occurs during an infinitesimally small interval dt is λdt (independent of arrivals outside this interval) and the probability that more than one arrival occurs is $o(dt)$. This definition is equivalent with the renewal process definition.

Random sampling Random and independent sampling of events with probability p from a Poisson process of rate λ results into a Poisson process of rate $p\lambda$.

Superposition Superposition of two Poisson process of rate λ_1 and λ_2 respectively results into a Poisson process of rate $\lambda_1 + \lambda_2$.

2.2 Assumptions and Notations

To describe the routing overlay of distributed hash tables, we reuse the terminology and reference model defined in [1]. Let's consider a one dimensional Euclidean metric space within the interval $[0,1)$ that wraps around (this can be represented

as a ring, see Fig. 2). Distance between two nodes in this metric space is defined as their distance along the ring in clockwise direction,[3] formally: $d(x,y) = y - x + I_{x>y}$.

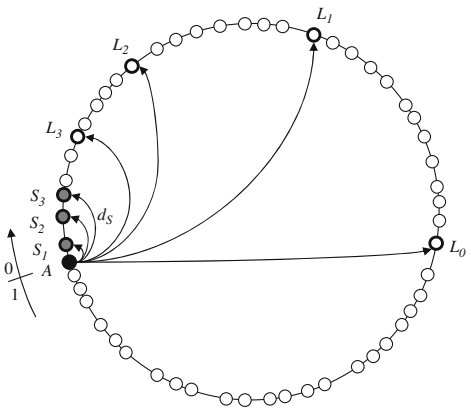

Fig. 2 Model of unidirectional DHT overlays (example)

Each node has two different types of connections to other nodes: short-range connections (called "local" connection in [1]) to a fixed number (N_S) of closest nodes (in clockwise direction) and long-range connections to some distant nodes. These nodes are called short-range and long-range peers of the node, respectively. Fig. 2 shows short-range connections (S_1, S_2, S_3) and long-range connections (L_0, L_1, L_2, L_3) of node (A). The distance of the node and its farthest short-range peer is denoted by d_S.

Routing is assumed to be greedy: a node forwards a lookup request to its peer being the closest to the target node in the metric space of the DHT (without overshooting it). This greedy routing process can be described by the following pseudo-code algorithm:

[3] This definition implies that the metric space is in fact only a quasi-metric space, since it does not satisfy the symmetry requirements. Extending the model to bidirectional routing where distance is defined as the shortest path along the ring (in any of the two directions), a real metric space can be obtained.

```
1 while node ≠ target do
2 |   proxy ← GetClosestPeer(node,target);
3 |   if Distance(proxy,target) < Distance(node,target) then
4 |   |   node ← proxy;
5 |   else
6 |   |   error
7 |   end
8 end
```

Algorithm 15: Greedy overlay routing

Routing overlay of many DHT implementations (Chord [22], Pastry [21], Symphony [16], Accordion [12] etc.) can be described (or approximated) using the above system model (e.g., routing in Pastry is more complex but is based on the same greedy algorithm). However, there are a few exceptions, for example DHTs using multidimensional metric spaces (e.g., CAN [19]) or non-Euclidean metric spaces (e.g., Kademlia [17]).

2.2.1 Degree of Randomness

Since randomness and flexibility in the choice of long-range connections plays an important role in both analysis and maintenance of overlays, let us first define two extreme DHT overlay categories:

Definition 1 (Probabilistic power-law routing overlay (PPLRO)). A routing overlay is called probabilistic power-law routing overlay when the choice of long-range connections is not deterministic and they only have to satisfy the following requirements: the probability of having a long-range connection to another overlay node is inversely proportional to the Dth power of the distance between the two nodes in the D dimensional metric space $(\mathscr{I},d)$ where the DHT maps node identifiers [8]. Join algorithm of probabilistic power-law routing overlays create initial long-range connections of joining nodes according to this distance distribution and the choice of long-range connections is mutually independent of each other.

Definition 2 (Deterministic power-law routing overlay (DPLRO)). A routing overlay over a one-dimensional metric space[4] is called deterministic power-law routing overlay if long-range connections are determined by the power series of the distances $d_i = \frac{q}{c^i}$ where c and q are constant so that $c > 1$ and $0 < q \leq 1$. For unidirectional overlays, the ith long-range connection is chosen as the first node, whose distance exceeds d_i while for bidirectional overlays, the ith connection is the node closest to the point at distance d_i.

[4] Extending this definition to multidimensional metric spaces on the analogy of probabilistic power-law routing overlays is not trivial.

Symphony [16], Accordion [12] and the routing scheme proposed in [2] are using probabilistic routing overlays while a deterministic power-law routing overlay can be thought of as a generalization of the Chord [22] overlay (for Chord, $c = 2$).

It is important to note that the term "power-law" is also used to denote the overlay of unstructured P2P systems structurally similar to scale-free random graphs [3]. In that context, it refers to distribution of node degree. However, in Definitions 1 and 2, the term "power-law" refers to distance distribution of long-range connections.

2.2.2 Bidirectional Overlay Model

The unidirectional overlay and routing model presented above can be easily extended to bidirectional overlays with bidirectional routing. In a bidirectional overlay, both short-range and long-range connections are bidirectional. Another important difference is the distance metric of the space $(\mathscr{I}, d)$. Using the ring representation, distance of two nodes is defined as their shortest distance along the ring, formally: $d_b(x,y) = +min\,[d(x,y), d(y,x)]$. Hence in contrast to unidirectional overlays, this is a real metric space also satisfying the symmetry requirements. As a result, from the point of view of distances, each node can split the DHT metric space $(\mathscr{I}, d)$ into two symmetrical partitions. Connections of a node are created independently in both of these partitions. Figure 3 shows short-range and long-range connections of node A in both partitions of the metric space for an example bidirectional overlay.

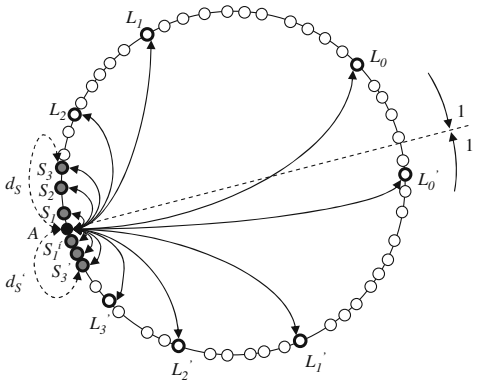

Fig. 3 Model of bidirectional DHT overlays (example)

Setting the circumference of the ring to 2 units for bidirectional overlays, it is easy to derive the bidirectional equivalent of any unidirectional overlay (where the circumference of the ring is set to 1 unit). Connections of the bidirectional overlay obey the distance distribution of the corresponding unidirectional overlay separately in both partitions of the metric space $(\mathscr{I}, d)$. Hence, the definition of probabilistic and deterministic power-law routing overlay can be extended for bidirectional

connections by applying either Defintions 1 or 2 separately for both partitions of the metric space.

As a consequence of the definition of distance metric for bidirectional overlays, greedy routing also becomes bidirectional; requests can be forwarded in both directions depending on the position of the peer node being closest to the target.

2.2.3 Node ID Distribution

The distribution of node identifiers in the metric space $(\mathscr{I},d)$ also affects mathematical analysis of routing in the overlay. The ID space of node identifiers is discrete and finite for most real systems, hence nodes can only be mapped to a finite subset of points in the Euclidean metric space $(\mathscr{I},d)$. However, the granularity of this finite ID space is so fine (the size of the ID space varies between $2^{128}-2^{256}$), that node ID mapping can be considered continuous in $(\mathscr{I},d)$.

Furthermore, we assumed that given a network of n nodes, node identifiers are drawn independently at random according to a uniform distribution over the range $[0,1)$ in $(\mathscr{I},d)$ (this is a reasonable assumption in most cases). This implies that distances between adjacent IDs on the ring will be exponentially distributed. In a few cases (explicitly noted), we assumed that peer identifiers partition the metric space $(\mathscr{I},d)$ deterministically in equal partitions. This is not a realistic scenario, but simplifies considerably mathematical analysis. In these later cases, we have always compared analytical results using deterministic identifier assignment to simulation results using random uniform distribution of peer identifiers.

Finally, for long-range connection selection in probabilistic power-law routing overlays, we assume that it is possible to find a peer node at any given distance (drawn according to a given distribution) in the metric space. This is not realistic in a real system composed of a finite number of nodes. In practice, the closest existing node to the given point is used instead. However, the resulting error between these theoretical and real distances is inversely proportional to the size of the network, hence this is negligible for large networks (which are in the main scope of this analysis).

3 Transformed View of Long-Range Connections

Transformation is a widely used mathematical concept in many disciplines to reveal, analyze and exploit hidden system characteristics. One of the best known examples of the application of a transformation method is JPEG encoding where discrete cosine transform maps a 8x8 pixel area into spatial frequency components [18]. In this example, transformation is used to exploit “hidden characteristic” of human vision being much more sensitive to small variations in color and in brightness for lower spatial frequencies than for higher frequencies. Hence higher spatial frequency components can be encoded at smaller resolutions. In our analysis, we ap-

ply a logarithmic transformation to distances between node identifiers in the metric space of a DHT to reveal "hidden characteristics" of DHT routing.

Definition 3 (Logarithmically transformed view). Let $(\mathscr{I}, d)$ be the metric space of a DHT (see Section 1.1) where the distance between the image $x_0 = F_P(p_0)$ of a node p_0 and the image $x_i = F_P(p_i)$ of another node p_i is defined as $d(x_0, x_i)$. Then, using the transformation function $f_t(u) = -\ln u$, the distance of p_0 and p_i in the logarithmically transformed view of p_0 is defined as:

$$d'(x_0, x_i) = f_t\left[d(x_0, x_i)\right] = -\ln\left[d(x_0, x_i)\right]. \tag{6}$$

It is important to note that $d'(x_0, x_1)$ is not a distance metric since it does not obey the three metric space properties. However, $d'(x_0, x_1)$ is not used as a distance metric; transformation of distances is only a mathematical tool within the concept of logarithmically transformed view.

The transformed view of a base node p_0 can be used to characterize distances between p_0 and a set of other nodes in a DHT. This transformed view can be represented along a half-line as follows: the base node p_0 itself is at the end of the half-line while other DHT nodes p_i (e.g., peers of the base node, or the target node of a lookup process) are represented along this half-line at distance $d'(x_0, x_i)$ from p_0.

3.1 Long-Range Connection Density

Figure 4 represents long-range connections of a Chord [22], Pastry [21] and Kademlia [17] node as well as long-range connections of a node in a probabilistic power-law routing overlay (e.g., Symphony [16] or Accordion [12]). For Pastry, the parameter b is the bit length of numbers in the routing table, for Kademlia, the parameter k is the maximum size of buckets and for Chord, the parameter c is the parameter used in the definition of deterministic power-law routing overlays (see Definition 2). For each of these DHTs, the upper line shows long-range peers of the node in the real metric space[5] (to ease graphical representation, the ring geometry of the metric space has been straightened) while the lower line shows these peers in the logarithmically transformed view of the node. In the real metric space, the represented node is in point 0. In the transformed view, this point corresponds to $+\infty$. Finally, long-range connections in the transformed view span within the range $[0, -\ln d_S)$, where d_S is the distance from the farthest short-range peer of the node (represented by grey circle in Fig. 4).

Consider now the segment $(d_S, 1)$ covered by long-range connections in the real metric space which corresponds to the segment $(0, -\ln d_S)$ in the transformed view

[5] Since Kademlia uses a XOR metric, long-range peers of the Kademlia node are represented based on their XOR distance from the node. Note that this is different from ID-based placement along the ring.

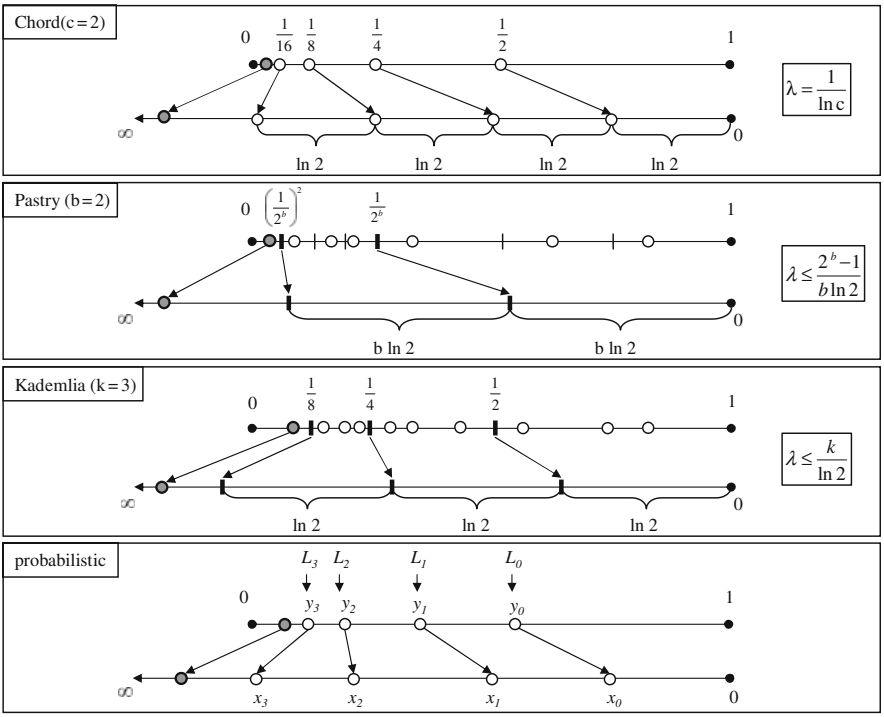

Fig. 4 Comparison of the routing table of some well known DHTs

of the node. Although long-range connection distribution differs for each of the above DHT implementations, Fig. 4 shows that it is possible to partition this segment in the transformed view into equally sized partitions of length Δx so that the number of long-range connections $N_L(\Delta x)$ be the same inside each of these partitions (either deterministically or in expected value). Based on this observation, one can define a $\lambda_{\Delta x}$ long-range connection density parameter as

$$\lambda_{\Delta x} = \frac{E\left[N_L(\Delta x)\right]}{\Delta x}. \tag{7}$$

In general, the choice of Δx is not arbitrary. In order to obtain constant long-range connection density in the entire long-range connection domain of the transformed view, Δx might need to be set to a DHT specific value (see again Fig. 4). However, for a large subclass of DHTs (including regular power-law routing overlays defined in the next subsection), long-range connection density can be defined independent of the size Δx of partitions as:

$$\lambda = \lim_{\Delta x \to 0} \frac{E\left[N_L(\Delta x)\right]}{\Delta x}. \tag{8}$$

The main advantage of the proposed λ (or $\lambda_{\Delta x}$) parameter is that it provides a simple and generic way to characterize long-range connection distribution. Furthermore, in $O(\log n)$ node state DHTs, λ characterizes the overlay independent of network size. For DHTs with constant node degree (e.g., Symphony [16]), λ depends on the size of the network ($\lambda \sim \frac{1}{\ln n}$).

Many DHT implementations have one or more tunable system parameter which affects long-range connection distribution, e.g., the bucket size k for Kademlia or the bit length b of numbers in the routing table for Pastry. The λ long-range connection density parameter allows easy comparison of overlay structure for these DHT implementations despite their mutually incompatible sets of system parameters. Figure 4 shows the λ parameter for each DHT as a function of their tunable system parameters. Note that in many cases (e.g., Chord, Pastry or Kademlia), the theoretical long-range connection values are only upper bounds of the actual long-range connection density since some routing table entries may be empty (especially for shorter distances).

3.2 Regular Power-Law Routing Overlays

Definition 4 (Regular power-law routing overlays (RPLRO)). A power-law routing overlay is called regular if the sequence of long-range connections of a node in its transformed view correspond to a randomly chosen sample of length $-\ln d_S$ from an infinite renewal process and these random samples are chosen independently for each node.

Hence, from the definition of renewal processes, distances between subsequent long-range connections of a node in its transformed view are *i.i.d* in a RPLRO. Furthermore, the distribution function $F(x)$ of these *i.i.d* random variables identifies unambiguously a regular power-law routing overlay. Random sampling from an infinite process (instead of defining long-range connections as a renewal process starting from point 0 in the transformed view) ensures uniformity of long-range connection density in the whole long-range connection range (also including the first part of this range).

Theorem 1. *Long-range connection density of a regular power-law routing overlay is uniformly $\lambda = \frac{1}{\mu}$, where μ is the mean distance between subsequent long-range connections of a node in its transformed view.*

Proof. Let $f(x)$ be the *pdf* and $F(x)$ be the *cdf* of renewal intervals (corresponding to the distances between subsequent long-range connections of a node in its transformed view). As a result of the random sampling property of RPLROs, the starting point of each partition of length $\Delta x \to 0$ can be considered as a randomly chosen point in the corresponding infinite renewal process. Hence this interval of length $\Delta x \to 0$ contains at least one renewal point (long-range connection) either when the corresponding renewal interval (selected by this randomly chosen point)

is smaller than Δx or when this interval is larger than Δx but the distance between the randomly selected point and the next renewal is less than Δx. Since the *pdf* of the length of the renewal period selected by a random point is $\frac{f(x)x}{\mu}$, the probability that such an interval contains at least one renewal (long-range connection) can be written as:

$$p_1 = \int_0^{\Delta x} \frac{f(x)x}{\mu} dx + \int_{\Delta x}^{\infty} \frac{f(x)x}{\mu} \frac{\Delta x}{x} dx. \tag{9}$$

The probability that this interval contains k or more arrivals (where $k > 1$) can be upper bounded as follows:

$$p_k \leq p_1 F^{k-1}(\Delta x). \tag{10}$$

The expected number of renewals (long-range connections) within a randomly selected period of length Δx can be written as:

$$E[N_L(\Delta x)] = \sum_{i=1}^{\infty} p_i. \tag{11}$$

Hence, combining (10) and (11):

$$p_1 \quad \leq \quad E[N_L(\Delta x)] \quad \leq \quad p_1 \left[1 + \sum_{i=1}^{\infty} F^i(\Delta x) \right]. \tag{12}$$

Dividing by Δx and applying Equation(7):

$$\frac{p_1}{\Delta x} \quad \leq \quad \lambda_{\Delta x} \quad \leq \quad \frac{p_1}{\Delta x} \left[1 + \sum_{i=1}^{\infty} F^i(\Delta x) \right]. \tag{13}$$

Applying $\Delta x \to 0$ to Equation (9) and the (reasonable) assumptions[6] that $\lim_{\Delta x \to 0} F(\Delta x) = 0$ and $\lim_{\Delta x \to 0} f(\Delta x) < \infty$:

$$\lim_{\Delta x \to 0} \frac{p_1}{\Delta x} = \frac{f(0)\Delta x}{2\mu} + \frac{1 - F(\Delta x)}{\mu} = 0 + \frac{1}{\mu}. \tag{14}$$

Finally, substituting Equation(14) into the Inequality (13):

$$\frac{1}{\mu} \leq \lambda \leq \frac{1}{\mu} \quad \to \quad \lambda = \frac{1}{\mu}. \tag{15}$$

Probabilistic power-law routing overlays are regular (see Section 3.3). Pastry and Kademlia are not regular but are close to being regular with only small distortions. Finally, Chord and deterministic power-law routing overlays in general are not regular, but, they can be made regular: Considering the transformed view of a node in a

[6] These assumptions can be made because 0 distance between subsequent long-range connections does not make sense in an overlay.

DPLRO, its first long-range peer is always located at $\ln c$. Substituting the constant q in Defition 2 by a random variable so that this first long-range peer be evenly distributed in the range $[0, \ln c]$, the overlay becomes regular.

To characterize regular power-law routing overlays, we also introduce a c_v long-range connection coefficient of variance parameter describing the relative variance of distances between long-range connections in the transformed view. $c_v = \frac{\sigma}{\mu}$, where σ is the standard deviation while μ is the mean of distances between consecutive long-range connections in the transformed view of a node. In Section 4.1, we show that using the λ and c_v parameters it is possible to derive a lower bound on routing performance.

3.3 Probabilistic Power-Law Routing Overlays

It is interesting to compare the degree of randomness in the choice of long-range connection for different DHT implementations in Fig. 4. In Chord, each connection is deterministic. Pastry is somewhat more flexible, each routing table entry may contain any node of the network from a given ID range, increasing the degree of randomness. Kademlia goes one small step further in flexibility and randomness and allows the choice of any nodes (up to a maximum number of k) from a given range.

However, the choice of long-range connections can be made "even more random" within the family of routing overlays for which long-range connection density can be defined. For the "most random" routing overlays out of this family, long-range connections of a node in its transformed view correspond to a random and independent placement of points in the range $(0, -\ln d_S)$ according to a uniform distribution, which is equivalent to a truncated (spatial) Poisson process. In the following, we show that this family of "most random" routing overlays is equivalent to the family of probabilistic power-law routing overlays over a one dimensional metric space (see Definition 1).

Theorem 2. *Consider a truncated Poisson process of rate λ in the range $(0, -\ln d_S)$. Furthermore consider a routing overlay where the sequence of long-range connections in the transformed view of each node is defined as a random realization of this Poisson process. Then this routing overlay is a probabilistic power-law routing overlay of long-range connection density λ.*

Proof. Consider a small range $[x, x+\Delta x]$ in the transformed view. Inverse transforming this range back to the real metric space using $y = f_t^{-1}(x) = e^{-x}$ results into the range $[e^{-x-\Delta x}, e^{-x}] = [y-\Delta y, y]$ in the real metric space.

Using the birth process definition of Poisson processes, the probability of having an arrival (long-range connection) in the range $[x, x+\Delta x]$ of the transformed view is $\lambda \Delta x$ when $\Delta x \to 0$. Since the inverse transformation function $f_t^{-1}(x)$ is strictly monotone decreasing, the probability of having a long-range connection in the corresponding range $[y-\Delta y, y]$ of the real metric space is the same.

Using the derivative of the transformation function $f_t(y) = -\ln y$, it is possible to express the relationship between the length of these ranges when they are infinitesimally small:

$$\lim_{\Delta y \to 0} \Delta x = -f_t'(y)\Delta y = \frac{\Delta y}{y}. \tag{16}$$

Hence the probability of having a long-range connection in an infinitesimally small range of length $\Delta y \to 0$ at a distance y from this node is $\lambda \frac{\Delta y}{y}$. This is equivalent to the long-range connection distribution requirement of Definition 1 for one dimensional metric spaces. Being generated from a Poisson process, long-range connections also satisfy the independence requirement of Definition 1, hence the generated overlay is a probabilistic power-law routing overlay.

Finally, the expected number of arrivals in a Poisson process of rate λ for an interval of length Δx is $\lambda \Delta x$, hence substituting into Equation(7) any positive value of Δx interval length, the obtained long-range connection density value for this overlay equals to the λ rate of the generating Poisson process.

Theorem 3. *Consider a probabilistic power-law routing overlay in a one dimensional metric space with a long-range connection density λ. Then the sequence of long-range connections of any node in its transformed view correspond to a random realization of a truncated Poisson of rate λ in the range $(0, -\ln d_S)$.*

Proof. According to Definition 1, the probability of having a long-range connection in an small range $[y-\Delta y, y]$ of a one dimensional metric space is $c\frac{\Delta y}{y}$ when $\Delta y \to 0$ (c is a positive constant).

Transforming this range into the transformed view using $x = f_t(y) = -\ln y$ results into the range $[-\ln y, -\ln(y-\Delta y)] = [x, x+\Delta x]$. Since the transformation $f_t(y)$ is strictly monotone decreasing, the probability of having a long-range connection in the corresponding range $[x, x+\Delta x]$ of the transformed view is the same.

Using the derivative of the inverse transformation function $f_t^{-1}(x) = e^{-x}$, it is possible to express the relationship between the length of these ranges when they are infinitesimally small:

$$\lim_{\Delta x \to 0} \Delta y = -f_t^{-1\prime}(x)\Delta x = e^{-x}\Delta x = y\Delta x. \tag{17}$$

Hence the probability of having a long-range connection in an infinitesimally small range of length $\Delta x \to 0$ in the transformed view is $c\frac{\Delta y}{y} = c\Delta x$, independent of the value of x within the range $(0, -\ln d_S)$. Since long-range connections of a probabilistic power law routing overlay are also independent of each other according to Definition 1, the sequence of long-range connections of any node in its transformed view correspond to a random realization of a truncated Poisson of rate c in the range $(0, -\ln d_S)$.

Finally, using the definition of long-range connection density and the assumption that the long-range connection density of the given overlay is λ, it is deducible that $c = \lambda$.

Since a Poisson process is a special renewal process, from Theorem 2, it follows that probabilistic power-law routing overlays belong to the subclass of regular power-law routing overlays.

3.4 Distortions in the Transformed View

In Section 3.3, we assumed that a node can find (and create a connection to) a peer node at any given point of the metric space $(\mathscr{I},d)$. In reality, given a network of n nodes, a connection can be created only to $n-1$ points in $(\mathscr{I},d)$ corresponding to the images of all the other nodes in $(\mathscr{I},d)$. Hence in real systems, a connection is established to the peer node whose images is the closest to the "theoretical" point in $(\mathscr{I},d)$ drawn according to the required distribution. This introduces small distortions to theoretical distance distribution.

Similar distortions exist for deterministic power-law routing overlays. E.g., in Chord, long-range connections (fingers) point to the first node whose distance is not smaller than $\frac{1}{2},\frac{1}{4},\frac{1}{8},\ldots$. This results into a sequence of distances $\frac{1}{2}+\varepsilon_1,\frac{1}{4}+\varepsilon_2,\frac{1}{8}+\varepsilon_3,\ldots$. In a network of n nodes uniformly distributed in the range $[0,1)$ of the metric space $(\mathscr{I},d)$, ε will be a random variable with exponential distribution of parameter n.

While the distribution of this small offset is the same for all distances in the real metric space, it depends strongly on the distance in the transformed view of a node. For large real distances, this offset is negligible, however, it increases exponentially and can be considerable for small real distances in the transformed view.

4 Stochastic Analysis of Routing

The role of short-range and long-range connections in the routing process is complementary. While short-range contacts ensure the success of greedy forwarding, long-range contacts expedite routing and provide $O(\log n)$ bounds on the number of routing hops. For deterministic routing geometries, the routing process can be clearly separated into a first phase using only long-range contacts and a second phase using only short-range contacts. For non-deterministic routing geometries, the first routing hops usually take place via long-range connections while the last hops usually take place via short-range connections and the probability of routing via a short-range peer increases monotonously approaching to the target. Nevertheless, for non-deterministic routing geometries, it is not possible to separate routing process into distinct long-range and short-range routing phases.

Analytical study of this dual routing process is rather complicated. Analysis becomes much easier if forwarding is restricted to either only short-range or only long-range connections.

Restricting forwarding to short-range connections, progress toward the target becomes linear. Assuming that node identifiers are drawn independently and at random according to a uniform distribution from the interval $[0,1)$ of the metric space $(\mathscr{I},d)$ (see Section 2.2), each routing hop has the same length in expected value independent of the current distance from the target.[7] The length of consecutive routing hops can be described by a series of independent Erlang distributed random variable with rate n (number of nodes in the network) and shape parameter N_S (number of short-range connections per node). Obviously, routing via only short-range contacts degrades routing performance from $O(\log n)$ to $O(n)$.

In Section 4.1, we show that using logarithmic transformation, analysis is also possible when restricting forwarding to long-range connections; progress toward the target will be linear in the transformed view of the target. However, simply forbidding forwarding via short-range contacts may cause routing failures. Therefore, to analyze long-range only forwarding, we use an imaginary routing overlay where the sequence of long-range connections in the transformed view of a node is infinite instead of being truncated after reaching the short-range connection domain. For regular power-law routing overlays, this means that long-range connections correspond to infinite random samples from a renewal process instead of random samples of length $-\ln d_S$.

In the real routing overlay, forwarding takes place via a short-range peer only when the target is closer than the closest long-range peer of the forwarding node. When the real routing overlay is forwarding via a short-range peer, the modified long-range only model is forwarding through an imaginary long-range peer at a smaller distance

Figure 5 demonstrates the difference between real forwarding (via a short-range connection) and long-range only forwarding via an imaginary long-range connection. The upper line in the figure represent the real metric space while the lower line shows the transformed view of the forwarding node. For a real overlay, forwarding occurs via a short-range peer. However, when restricting forwarding to long-range connections in order to simplify analysis, forwarding takes place via the imaginary long-range peer being the closest to the target node.

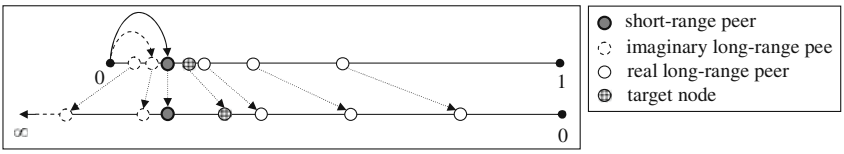

Fig. 5 Forwarding through imaginary long-range connections

Forwarding via an imaginary long-range connection always results in less progress than the real forwarding would result in via a short-range connection. Therefore

[7] If $N_S > 1$, the expected value of the last hop reaching the target is smaller.

results on routing progress obtained from analysis restricted to long-range forwarding can be used as a lower bound on real routing progress.

Modeling long-range only forwarding, a routing process is terminated when the distance of the current forwarding node from the target decreases below d_S (the distance between the target node and the node whose farthest short-range peer is the target node[8]). Let M_l be the number of routing hops for long-range only routing until the termination and let M be the number of routing hops for the real routing process. Since routing progress of long-range only forwarding is always equal to or less than routing progress of real forwarding, the real routing process will always reach a peer with direct short-range connection to the target node in M_l or less hops. Hence $M_l + 1$ can be used as an upper bound on the real number of routing hops:

$$M \leq M_l + 1. \tag{18}$$

The rest of this section is structured as follows: Section 4.1 analyzes progress of the routing process in the transformed view of the target using this long-range only forwarding model. Then Section 4.2 uses the obtained long-range only results to derive upper bounds on the number of routing hops for the real routing process (using both short-range and long-range connections).

4.1 Analysis of Routing in the Transformed view

Analysis of routing in the transformed view can be best introduced through an example. Figure 6a shows one hop of an example routing process: a request reaches forwarding node F_k in step k and node F_k forwards this request to its long-range peer F_{k+1} being the closest to the target node T without overshooting it. Figure 6b shows distances in the real metric space (upper line) and the transformed view (lower line) of node F_k while Fig. 6c shows the same distances as seen in the real and transformed view of the target. Note that the default direction of the ring is reversed in Fig. 6c in order to represent remaining distances from the perspective of the target node.

d_k and d_{k+1} is the distance from the target in step k and $k+1$ respectively, while d'_k and d'_{k+1} are the same distances in the transformed view of the target. To analyze per hop routing progress in the transformed view of the target node, let's express the progress $u_k = d'_{k+1} - d'_k$ toward the target after step k as a function of the distance v_k from the next-hop node in the transformed view of forwarding node F_k. Applying transformation to distances in Fig. 6c:

[8] Assuming uniform distribution of node identifers in the DHT metric space, d_S will be a random variable with Erlang distribution of rate n and shape parameter N_S (where n is the number of nodes in the DHT and N_S is the number of short-range connections per node.)

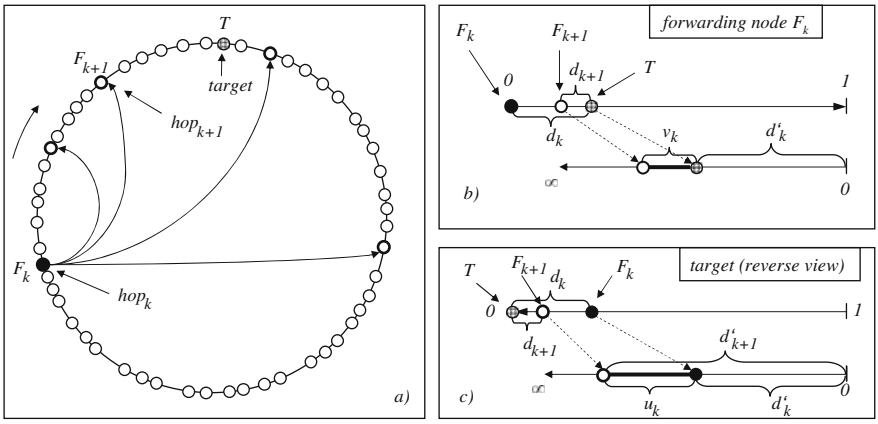

Fig. 6 Routing from hop k to hop $k+1$

$$u_k = -\ln d_{k+1} - d'_k. \tag{19}$$

Inverse transforming distances in Fig. 6/b:

$$d_{k+1} = e^{-d'_k} - e^{-d'_k - v_k} = e^{-d'_k}(1 - e^{-v_k}). \tag{20}$$

Finally, substituting Equation(20) into (19):

$$u_k = -\ln\left[e^{-d'_k}\left(1 - e^{-v_k}\right)\right] - d'_k = -\ln\left(1 - e^{-v_k}\right). \tag{21}$$

Hence the progress u_k after routing hop k in the transformed view of the target can be expressed as a function of the distance v_k from the next-hop node in the transformed view of forwarding node F_k:

$$u = h(v) = -\ln\left(1 - e^{-v}\right). \tag{22}$$

Note: as Section 2.2 mentions, routing in DHTs with non-Euclidean metric spaces cannot be analyzed using this model. The reason is that Equation(20) uses the assumption that $d(x,z) = d(x,y) + d(y,z)$ which holds only for the one dimensional Euclidean metric space. For any other metric spaces: $d(x,z) \leq d(x,y) + d(y,z)$ (triangle inequality).

In the followings, we analyze routing in regular power-law routing overlays in general. Using the above transformation results, we derive a lower bound on routing performance as a function of λ and c_v. Then we analyze two special cases of regular power-law routing overlays in more details: the probabilistic and "regularized" deterministic overlays. We derive exact analytical results on their routing performance and compare these results to the generic lower bound.

4.1.1 Regular Power-Law Routing Overlays

Theorem 4. *Considering the routing process via long-range connections in a regular power law routing overlay, the length of the per-hop progress u_k in the transformed view of the target is* i.i.d *for subsequent hops. Furthermore, the expected value of this per-hop routing progress is lower bounded by*

$$E[u_k] \geq -\ln\left[1 - e^{-\frac{1+c_v^2}{2\lambda}}\right]. \tag{23}$$

where λ is the long-range connection density and c_v is the long-range connections density coefficient of variation in the overlay.

Proof. In regular power-law routing overlays, the sequence of long-range connections of different nodes in their transformed view correspond to independently selected random samples from an infinite renewal process. Therefore, the target node can be considered as a uniformly distributed random point in the transformed view of a forwarding node (see Fig. 6b). As a consequence, the random variable v_k corresponds to the distance of a random point from the next renewal (long-range connection) in the renewal process of long-range connection (residual life). Hence, the series of the random variables v_k will be *i.i.d*, and applying Equation (22), the series of the random variables u_k will be also *i.i.d*.

Let $\mu = E[x]$ be the expected value and $\mu_2 = E[x^2]$ be the second moment of the length of renewal periods (corresponding to distances between subsequent long-range connection in the transformed view of a node). Then, from renewal theory, the mean residual life in this renewal process (corresponding to $E[v]$) can be expressed as:

$$E[v] = \frac{E[x^2]}{2E[x]} = \frac{Var(x) + E^2[x]}{2E[x]} = \frac{E[x]}{2}(1 + c_v^2). \tag{24}$$

Using Theorem 1, $E[x] = \frac{1}{\lambda}$ for any RPLRO, hence:

$$E[v] = \frac{1 + c_v^2}{2\lambda}. \tag{25}$$

Using Equation 22, the distribution of the per-hop routing progress in the transformed view of the target (u) can be expressed as a convex function $h(v)$ of the random variable v. Therefore the Jensen inequality can be applied as follows:

$$E[u] = E[h(v)] \geq h(E[v]) = -\ln\left[1 - e^{-\frac{1+c_v^2}{2\lambda}}\right]. \tag{26}$$

4.1.2 Probabilistic Power-Law Routing Overlays

According to Theorem 3 on PPLROs, the sequence of long-range connections in the transformed view of a node can be described as a realization of a stationary

Poisson process of rate λ, where λ is the long-range connection density of this overlay. Hence, in the transformed view of the forwarding node F_k, the target of the lookup process corresponds to an arbitrary point while the image of the next-hop node F_{k+1} corresponds to the next arrival in this Poisson process. The random variable v_k describes the distance between these two points in the transformed view of F_k (see Fig. 6b).

As a consequence of the memoryless property of Poisson processes, picking an arbitrary point in the process, the distance to the next arrival will always be exponentially distributed with parameter λ. Hence the distribution of v_k is the same for each step of the routing process (via long-range connections) and the *pdf* of v v_k is:

$$g_{prob}(v) = \lambda e^{-\lambda v}. \tag{27}$$

Another consequence of the memoryless property of Poisson processes is that the random variables v_k and v_{k+1} are independent, hence the series v_k are *i.i.d* random variables. Since u_k (the progress toward the target in the k^{th} routing hop) can be derived from v_k using Equation (22), u_k is also a series of *i.i.d* random variables (since PPLROs are regular, this could be derived also applying Theorem 4). The *pdf* of u can be obtain by transforming the *pdf* of v using the function $h(v)$:

$$f_{prob}(u) = g_{prob}\left(h^{-1}(u)\right)\left|\frac{dh^{-1}(u)}{du}\right| = \lambda(1-e^{-u})^{(\lambda-1)}e^{-u}. \tag{28}$$

Hence:

$$F_{prob}(u) = \int_0^u f_{prob}(t)dt = (1-e^{-u})^{\lambda}. \tag{29}$$

Finally, the expected value of the length of one routing hop in the transformed view of the target[9]:

$$E_{prob}[u] = \int_0^{\infty} f_{prob}(u)u\,du = H_{\lambda}, \tag{30}$$

where H_x is the harmonic number [24] (generalized for real numbers) of x. For practical λ values, the following approximation can be used[10]:

$$H_{\lambda} \approx \ln\left[(e-1)\lambda + 1\right]. \tag{31}$$

The above results can be transformed back from the transformed view of the target node to the real metric space of the DHT as follows.

Theorem 5. *Consider the routing process in a probabilistic power-law routing overlay of long-range connectiondensity λ. Then the series of random variables*

[9] Calculated using the Mathematica software from Wolfram Research Inc. (http://www.wolfram.com)

[10] This approximation provides less than $\pm 1\%$ relative error if $\lambda > 0.5$ and less than $+5\%$ relative error if $0 < \lambda < 0.5$. Note that λ is typically larger than $\frac{1}{\ln 2} \approx 1.41$ for most DHTs (see Section 3.1)

$w_k = \frac{d_{k+1}}{d_k}$ describing the ratio of distances from the target after and before a routing hop via a long-range connection are i.i.d *and the* pdf *and expected value of w_k are:*

$$f^w_{prob}(w) = \lambda(1-w)^{(\lambda-1)} \quad \text{if} \quad 0 < w < 1 \quad \text{and0otherwise} \tag{32}$$

and

$$E[w] = \frac{1}{1+\lambda}. \tag{33}$$

Proof. Since $u_k = d'_{k+1} - d'_k$ in the transformed view of the target and since transformed distances can be obtained as $d'_k = -\ln d_k$ and $d'_{k+1} = -\ln d_{k+1}$ from distances in the real metric space, u_k can be expressed as:

$$u_k = -\ln d_{k+1} - (-\ln d_k) = -\ln \frac{d_{k+1}}{d_k}. \tag{34}$$

Hence defining, the random variable $w_k = \frac{d_{k+1}}{d_k}$ as the ratio of distances after and before a routing hop via a long-range connection, this random variable w_k can be expressed as a function $w_k = \Phi(u_k) = e^{-u_k}$ of the random variable u_k. According to Theorem 4, u_k is a series of *i.i.d* random variables, therefore w_k will be also a series of *i.i.d* random variables (to simplify notation, u_k and w_k are denoted simply by u and v hereafter). $\Phi(u) = e^{-u}$ is a strictly monotone decreasing function. Hence the *cdf* of w can be expressed from the *cdf* of u as:

$$F^w_{prob}(w) = 1 - F_{prob}\left(\Phi^{-1}(w)\right) = 1 - \left[1 - e^{-(-\ln w)}\right]^{\lambda} = 1 - (1-w)^{\lambda}. \tag{35}$$

The *pdf* of w can be obtained as the derivative of $F^w_{prob}(w)$:

$$f^w_{prob}(w) = \frac{dF^w_{prob}}{dw} = \lambda(1-w)^{(\lambda-1)}. \tag{36}$$

As a result of greedy routing, $d_{k+1} < d_k$ holds for each routing step, hence $0 < w_k < 1$ and the expected value of the random variable w is:

$$E[w] = \int_0^1 f^w_{prob}(w) w dw = \left[\frac{(1-w)^{\lambda}(1+\lambda w)}{1+\lambda}\right]_0^1 = \frac{1}{1+\lambda}. \tag{37}$$

Hence the distance to the target decreases in expected value by a factor of $1+\lambda$ after each routing hop via a long-range connection. Since these distance decrease ratios in subsequent routing hops are independent, the expected value of distance decrease after i routing hops via long-range connections can be expressed as:

$$E\left[\frac{d_{k+i}}{d_k}\right] = \left(\frac{1}{1+\lambda}\right)^i. \tag{38}$$

4.1.3 Deterministic Power-Law Routing Overlays

Definition 2 introduces deterministic power-law overlays which can be considered as a generalization of the Chord overlay. These overlays are not regular because the sequence of long-range connections in the transformed view of nodes correspond to the same renewal process for each node and lacks the random sampling property of regular power law routing overlays. However, DPLROs can be made regular substituting the constant q in Definition 2 with a random variable so that the first long-range peer is evenly distributed over the range $[0, \ln c]$ in the transformed view of the node. In this subsection, we analyze these "regularized" deterministic power-law routing overlays.

To ease comparison with PPLROs and regular power-law routing overlays in general, the generic λ long-range connection density is used during the analysis instead of the parameter c in the definition of deterministic power law routing overlays. The relationship between these two parameters can easily be determined from Fig. 4:

$$\lambda = \frac{1}{\ln c} \quad \Leftrightarrow \quad c = e^{\frac{1}{\lambda}}. \tag{39}$$

As for any regular power-law routing overlay, target nodes in the transformed view of forwarding nodes can be considered as uniformly distributed random points. Hence the *pdf* of the random variable v_k for DPLRO:

$$g_{det}(v) = \begin{cases} \lambda & \text{if} \quad 0 < v < \dfrac{1}{\lambda} \\ 0 & \text{otherwise} \end{cases} \tag{40}$$

Transforming this distribution using Equation(22), the *pdf* of the random variable u_k:

$$f_{det}(u) = g_{det}\left(h^{-1}(u)\right)\left|\frac{dh^{-1}(u)}{du}\right| = \begin{cases} \lambda \dfrac{e^{-u}}{1-e^{-u}} & \text{if} \quad u > -\ln\left[1-e^{\frac{1}{\lambda}}\right] \\ 0 & \text{otherwise} \end{cases} \tag{41}$$

Hence the expected value of the per-hop routing progress in the transformed view of the target node can be obtained as[11]:

$$E_{det}[u] = \int_{-\ln\left[1-e^{-\frac{1}{\lambda}}\right]}^{\infty} \lambda \frac{e^{-u}}{1-e^{-u}} u du. \tag{42}$$

4.1.4 Comparison of Per-Hop Routing Progress in the Transformed View

In the previous subsections, we analyzed routing via long-range connections in regular power-law routing overlays. In Theorem 4, we show that the length of the

[11] The analytical form of this integral is too complicated, Fig. 7 represents the results numerically

per-hop progress in the transformed view of the target (distance between the images of subsequent forwarding nodes in this transformed view) is *i.i.d.* In other words, lookup approaches toward the target in its transformed view at constant "speed" in expected value for any regular power-law routing overlay.

Furthermore, knowing the λ long-range connection density and the c_v long-range connection density coefficient of variation parameters of the overlay, it is possible to derive a lower bound on the expected value of the length of this per-hop progress in the transformed view of the target (see Equation (23)).

We have also derived analytically the expected value of this per hops progress for two special cases, namely the probabilistic and the "regularized" deterministic power-law routing overlays. Figure 7 compares these expected values as well as their generic lower bounds (derived using Inequality 23) as a function of the long-range connection density. The result shows that deterministic overlays provide better per-hop progress than probabilistic overlays for all values of λ. This is a consequence of different coefficients of variation (c_v) for distances between subsequent long-range connections. According to Equation (23), the lower bound on $E[u]$ decreases monotonically with increasing c_v values. For DPLROs, where the distance between subsequent long-range connections is constant in the transformed view of a node:

$$c_v^{det} = 0, \tag{43}$$

while for PPLROs, where these distances are exponentially distributed:

$$c_v^{prob} = \frac{\sigma}{\mu} = \frac{\frac{1}{\lambda}}{\frac{1}{\lambda}} = 1. \tag{44}$$

Although Equation (23) can be used to express lower bounds on the expected value for any regular power-law routing overlay, the distribution of per-hop routing progress can differ significantly for different overlays. Figure 8 compares the *pdf* $f(u)$ for different long-range connection density values both for probabilistic and deterministic power-law routing overlays. As it can be expected, the deterministic routing overlay guarantees a minimum progress for each routing hop. In probabilistic routing overlays, there is no such lower bound on the length of one single hop.

Equation (29) reveals another interesting property of the distribution of u for probabilistic power-law routing overlays: $F_{prob}(u)$ can be obtained by raising the *cdf* of an exponential distribution to the power λ. As a result, $\lambda = 1$ is a very special long-range connections density value where the length of one routing hop in the transformed view of the target node is exponentially distributed with parameter 1. This means that the sequence of routing hops in the transformed view of the target node corresponds to the same stochastic process as the sequence of long-range connections in the transformed view of any nodes; both can be described by a Poisson process of rate 1. For any other λ values, the length of routing hops is not exponentially distributed. Figure (8) shows well the difference in the shape of

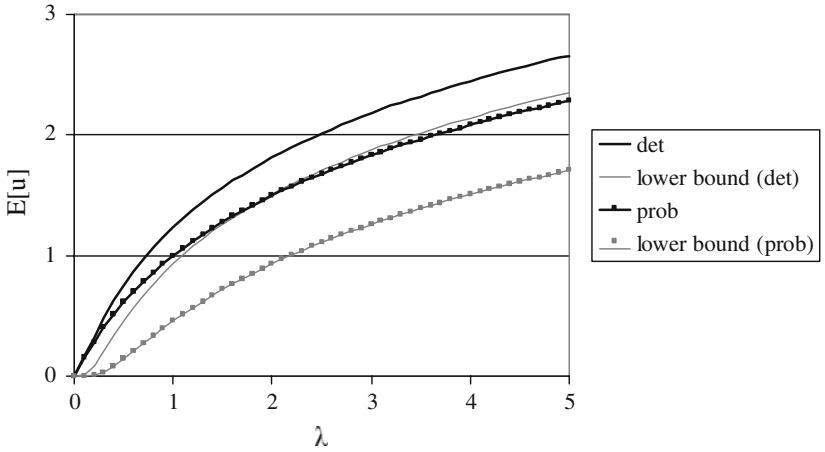

Fig. 7 Expected value of u and its estimated lower bound as a function of λ

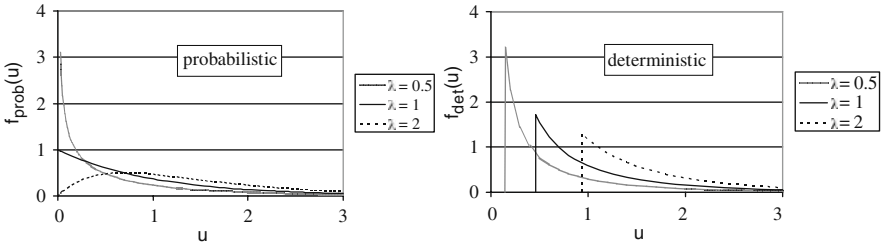

Fig. 8 Pdf of u for different λ values

the probability density function for long-range connection density values $\lambda = 1$, $\lambda < 1$ and $\lambda > 1$.

The sequence of long-range connections in the transformed view of a node can also be approximated by a renewal process for many "non-regular" DHTs (see Section 3.1). However, irregularities in the distribution of distances between successive long-range connections induces distortions to the "constant speed" progress in the transformed view of the target and make mathematical analysis difficult.

For example, in Pastry, long-range connection density have a slight periodic variation. The length of this period is $b \ln 2$ in the transformed view (where b is the bit length of numbers in the routing table). This slight periodic fluctuation is visible on the graph showing the expected number of routing hops as a function of network size (see Fig. 4 in paper [21]).

4.2 Upper Bound on the Expected Number of Routing Hops

In the previous subsections, we have analyzed routing via long-range connections in the transformed view of target nodes. We have shown that the per-hop routing progress u_k is *i.i.d* for regular power-law routing overlays, hence the images of forwarding nodes in the transformed view of the target can be described as a renewal process. Furthermore, we proposed a lower bound on the expected value of this per-hop progress as a function of the λ and c_v overlay parameters (see Theorem 4).

Although these results cannot be used directly to characterize overlay performance, the proposed renewal process model allows analytical derivation of an important overlay performance metric: an upper bound on the expected number of routing hops as a function of network size and the overlay parameters (λ, c_v and N_S). The computation of this overlay performance metric is based on the upper bound of Lorden for renewal processes (see [4], page 110):

$$U_L(t) \leq \frac{t}{\mu} + \frac{\mu_2}{\mu^2} - 1, \tag{45}$$

where $U_L(t)$ is the renewal function (the expected value of renewals until time t), μ is the mean of renewal periods and μ_2 is the second moment of renewal periods. Applying this bound to the renewal process corresponding to the sequence of forwarding nodes in the transformed view of the target node, the variables in Inequality(45) correspond to the followings:

- $U_L(t)$ corresponds to the expected value of the number of routing hops for long-range only routing;
- t corresponds to the length of the long-range routing path in the transformed view of the target node (from the image of the initiator node to the image of the first node having a direct short-range connection to the target node);
- μ corresponds to $E[u]$, for which, Theorem 4 gives a lower bound as a function of the overlay parameters λ and c_v;
- μ_2 corresponds to $E[u^2] = \int_0^\infty u^2 f(u)du$.

Lemma 1. *Considering the routing process via long-range connections in a regular power law routing overlay, the second moment of the length of the per-hop progress u in the transformed view of the target node is upper bounded by*

$$E[u^2] \leq 2.41\lambda, \tag{46}$$

where λ is the long-range connection density of the overlay

Proof. The random variable u can be obtained from the random variable v using Equation (22) while the variable v itself corresponds to the residual life in the renewal process corresponding to the sequence of long-range connections in the transformed view of a node (see proof of Theorem 4). From renewal theory, the *pdf* of the residual life v can be expressed as follows:

$$g(v) = \frac{1 - F(v)}{E(x)} = \lambda(1 - F(v)), \tag{47}$$

where $F(x)$ is the *cdf* of renewal intervals (corresponding to the distances between subsequent long-range connections of a node in its transformed view), $E[x]$ is the expected length of these intervals and we have used $\frac{1}{E[x]} = \lambda$ from Theorem 1.

Since a *cdf* is always a non-decreasing function and $0 \leq F(x) \leq 1$, the *pdf* $g(v)$ is a non-increasing function upper bounded by $g(v) \leq g(0) = \lambda$.

This upper bound can be used to derive an upper bound on $f(u)$. Using Equation(22) to transform v to u:

$$f(u) = g(-\ln(1 - e^{-u}))\frac{e^{-u}}{1 - e^{-u}} \quad \leq \quad \lambda \frac{e^{-u}}{1 - e^{-u}}. \tag{48}$$

Hence

$$E[u^2] = \int_0^\infty u^2 f(u) du \quad \leq \quad \int_0^\infty \lambda \frac{e^{-u}}{1 - e^{-u}} u^2 du \quad \leq \quad 2.41\lambda. \tag{49}$$

Lemma 2. *Assuming uniform distribution of node identifiers in the metric space of the DHT and uniform selection of target nodes for the routing process, the expected length of the routing path in the transformed view of the target is:*

$$E[t] = \ln n + \gamma - 1 - H_{N_S - 1} + \varepsilon, \tag{50}$$

where N_S is the number of short-range connections per node, n is the number of nodes in the overlay, H_k is the k^{th} harmonic number, γ is the Euler-Mascheroni constant [12] *and ε is a small positive error term*

$$\varepsilon \in o\left(\frac{n^{N_S}}{e^n}\right) \tag{51}$$

negligible except for very small network sizes.

Proof. When using the long-range only forwarding model, a routing process (as seen in the transformed view of the target) starts from the image of the initiator node and ends at the image of the first node having a direct short-range connection to this target node. The expected length $E[t]$ of this routing path can be obtained as the difference between the expected location of these start and end points.

Assuming that nodes of the overlay are uniformly distributed in the range $[0,1)$ of the DHT metric space and target nodes are selected also uniformly by initiator nodes, the distance between initiator and target nodes will be uniformly distributed over the interval $(0,1)$. Applying logarithmic transformation to this distribution according to Equation (6) results into exponentially distributed distances in the transformed view of the target with the *pdf*: $f'(x) = e^{-x}$. Hence the expected value of the distance between the target and initiator nodes in the transformed view of the target

[12] The Euler-Mascheroni constant is defined as $\gamma = \lim_{k\to\infty}(H_k - \ln k) \approx 0.5772$.

will be

$$E[L_{start}] = \int_0^\infty e^{-x} x dx = 1. \tag{52}$$

Assuming again that nodes of the overlay are uniformly distributed in the metric space of the DHT, the distance between a node and its farthest short-range peer will be Erlang distributed with rate n and shape parameter N_S, where n is the number of nodes in the overlay and N_S is the number of short-range connections per node. Hence the *pdf* of this distance distribution will be:

$$f_{Erl}(y) = \frac{n^{N_S} y^{N_S-1} e^{-ny}}{(N_S-1)!}. \tag{53}$$

Transforming the *pdf* of this Erlang distribution according to Equation (6), the *pdf* in the transformed view will be:

$$f'_{Erl}(x) = f_{Erl}\left(f_t^{-1}(x)\right) \frac{df_{Erl}^{-1}}{dx} = \frac{e^{-\left[ne^{-x}+x(N_S-1)\right]} n^{N_S}}{(N_S-1)!}. \tag{54}$$

Hence the expected value of the distance between a node and its farthest short-range peer in its transformed view[13]:

$$E[L_{end}] = \int_0^\infty f'_{Erl}(x) x dx = \ln n + \gamma - H_{N_S-1} + \varepsilon, \tag{55}$$

where H_k is the kth harmonic number, γ is the Euler-Mascheroni constant and ε is a small positive error term upper bounded by

$$\varepsilon < e^{-n}\left[1 + \frac{(n+N_S)^{N_S-2}}{(N_S-1)!}\right]. \tag{56}$$

Hence ε can be expressed using the following asymptotic bound:

$$\varepsilon \in o\left(\frac{n^{N_S}}{e^n}\right). \tag{57}$$

Typically, N_S is a small number, hence – except for very small network sizes – ε is negligibly small.

Theorem 6. *The expected number of routing hops U in a regular power-law routing overlay is upper bounded by:*

$$U(n,\lambda,c_v,N_S) \quad \leq \quad \frac{\ln n - H_{N_S-1} - 0.42}{-\ln\left[1 - e^{-\frac{1+c_v^2}{2\lambda}}\right]} + \frac{2.41\lambda}{\ln^2\left[1 - e^{-\frac{1+c_v^2}{2\lambda}}\right]} + \varepsilon, \tag{58}$$

[13] Calculated using the Mathematica software from Wolfram Research Inc. (http://www.wolfram.com)

where n is the number of nodes in the overlay, λ is the long-range connection density, c_v is the long-range connection density coefficient of variation, N_S is the number of short-range connections per node and ε is a small positive error term

$$\varepsilon \in o\left(\frac{n^{N_S}}{e^n}\right) \tag{59}$$

negligible except for very small network sizes.

Proof. Substituting the results of Lemma 1, Lemma 2 and Theorem 4 into the Lorden bound (Inequality 45), an upper bound can be obtained on the expected number of routing hops for long-range only forwarding:

$$U_L(t) \quad \leq \quad \frac{\ln n - H_{N_S-1} - 0.42 + \varepsilon}{-\ln\left[1 - e^{-\frac{1+c_v^2}{2\lambda}}\right]} + \frac{2.41\lambda}{\ln^2\left[1 - e^{-\frac{1+c_v^2}{2\lambda}}\right]} - 1. \tag{60}$$

Then, the upper bound on the number of routing hops for real routing (via both short-range and long-range connections) can be obtained using $U(t) < U_L(t) + 1$ from Inequality (18).

When the first and second moments of the per-hop progress u in the transformed view of the target are known, the upper bound of Theorem 6 can be further tightened:

Theorem 7. *The expected number of routing hops U in a probabilistic power-law routing overlay is upper bounded by:*

$$U(n,\lambda,N_S) \quad \leq \quad \frac{\ln n - H_{N_S-1} - 0.42}{H_\lambda} + \frac{1.645 - \psi'(1+\lambda)}{H_\lambda^2} + 1 + \varepsilon, \tag{61}$$

where n is the number of nodes in the overlay, λ is the long-range connection density, and N_S is the number of short-range connections per node, $\psi'(x)$ is the first derivative of the digamma function and ε is a small positive error term

$$\varepsilon \in o\left(\frac{n^{N_S}}{e^n}\right) \tag{62}$$

negligible except for very small network sizes.

Proof. Using Equation (30), the first moment of u is $\mu = H_\lambda$, where H_x is the harmonic number generalized for real numbers. The second moment of u can be derived from the *pdf* of u given by Equation (28):

$$\mu_2 = \int_0^\infty f_{prob}(u)u^2 du = \frac{\pi^2}{6} + H_\lambda^2 - \psi'(1+\lambda). \tag{63}$$

Substituting μ, μ_2 and the result of Lemma 2 into the Lorden bound (Inequality 45) an upper bound can be obtained on the expected number of routing hops for long-range only forwarding:

$$U_L(t) \quad \leq \quad \frac{\ln n - H_{N_S-1} - 0.42 + \varepsilon}{H_\lambda} + \frac{\frac{\pi^2}{6} + H_\lambda^2 - \psi'(1+\lambda)}{H_\lambda^2} - 1. \qquad (64)$$

Performing simplifications, the upper bound on the number of routing hops for real routing (using both short-range and long-range connection) can be obtained using $U(t) < U_L(t) + 1$ from Inequality (18).

5 Summary

Although most DHT overlays are structurally similar to the "small-world" navigation model of Kleinberg [8] – architectural and algorithmic details of different DHT variants differ significantly. Furthermore, lookup performance depends on a sets of different and often incompatible parameters which makes analytical comparison rather difficult. The objective of this chapter was to propose a general analytical model that can be used to investigate and compare static routing performance of most DHT implementations as a function of their overlay structure.

To capture the above mentioned common foundations of overlay structure, we have introduced the concept of logarithmically transformed view, where distances between a reference node and other nodes are represented after a logarithmic transformation. We have shown that long-range peers of a node form a linear sequence in this transformed view for most DHTs. Furthermore, we have identified an important subclass of DHT overlays – regular power-law routing overlays – where this sequence can be described as a random sample from an infinite renewal process. Based on this stochastic model, we have introduced the λ long-range connection density and c_v long-range connection density coefficient of variation parameters. For $O(\log n)$ node state, these parameters characterize long-range connection distribution independent of network size.

Using the renewal process model of long-connections, we have analyzed stochastically the progress of lookup process via long-range connections. We have shown that the sequence of intermediate forwarding nodes in the transformed view of the target node can be also described as a renewal process. Additionally, we have derived (i) the distribution of this per-hop routing progress for the spacial cases of probabilistic and "regularized" deterministic power-law routing overlays (ii) a generic upper bound on the per-hop routing progress in the transformed view of the target as a function of the λ and c_v long-range connection distribution parameters.

Finally, using the renewal process model of the routing process, we have derived closed form upper bounds on the expected number of routing hops as a function of network size and the overlay parameters λ, c_v and N_S.

The above model and results can be applied directly to any DHT using probabilistic power-law routing overlays (e.g., Symphony [16], Accordion [12], etc.). Additionally, overlay structure and static routing performance of any DHT using a one-dimensional metric space and being structurally similar to the "small-world"

navigation model of Kleinberg can be approximated applying this model (e.g., Chord [22] and its variants, Pastry [21], Bamboo [20], Kademlia [17], etc.)

References

1. Aberer, K., Alima, L.O., Ghodsi, A., Girdzijauskas, S., Haridi, S., Hauswirth, M.: The essence of P2P: A reference architecture for overlay networks. In: Proceedings of the 5th IEEE International Conference on Peer-to-Peer Computing, pp. 11–20. Konstanz, Germany (2005)
2. Aspnes, J., Diamadi, Z., Shah., G.: Fault tolerant routing in peer-to-peer systems. In: Proceedings of the 21st Annual Symposium on Principles of Distributed Computing (PODC '02), pp. 223–232. Monterey, CA, USA (2002)
3. Barabási, A.L., Albert, R.: Emergence of scaling in random networks. Science **286**(5439), 509–512 (1999)
4. Beichelt, F., Fatti, L.P.: Stochastic Processes and Their Applications. CRC Press (2001)
5. Gummadi, K., Gribble, S., Ratnasamy, S., Shenker, S., Stoica, I.: The impact of DHT routing geometry on resilience and proximity. In: Proceedings of ACM Sigcomm, pp. 381–394. Karlsruhe, Germany (2003)
6. Kaashoek, F., Karger, D.: Koorde: A simple degree-optimal distributed hash table. In: Proceedings of the 2nd International Workshop on Peer-to-Peer Systems, LNCS 2735, pp. 98–107. Berkeley CA, USA (2003)
7. Karger, D., Lehman, E., Leighton, T., Panigrahy, R., Levine, M., Lewin, D.: Consistent hashing and random trees: distributed caching protocols for relieving hot spots on the World Wide Web. In: Proceedings of the 29th annual ACM symposium on theory of computing (STOC '97), pp. 654–663. El Paso, TX, USA (1997)
8. Kleinberg, J.M.: The small-world phenomenon: an algorithmic perspective. In: Proceedings of the 32nd Annual ACM Symposium on Theory of Computing, pp. 163–170. Portland, OR, USA (2000)
9. Kleinrock, L.: Queueing Systems, vol. 1. Wiley (1975)
10. Kong, J.S., Bridgewater, J.S.A., Roychowdhury, V.P.: A general framework for scalability and performance analysis of DHT routing systems. In: Proceedings of the International Conference on Dependable Systems and Networks (DSN'06), pp. 343–354. Philadelphia, PA, USA (2006)
11. Lawler, G.F.: Introduction to Stochastic Processes. CRC Press (2006)
12. Li, J., Stribling, J., Morris, R., Kaashoek, M.F.: Bandwidth-efficient management of DHT routing tables. In: Proceedings of the 2nd Conference on Symposium on Networked Systems Design & Implementation, pp. 99–114. Boston, MA, USA (2005)
13. Li, J., Stribling, J., Morris, R., Kaashoek, M.F., Gil, T.M.: A performance vs. cost framework for evaluating DHT design tradeoffs under churn. In: Proceedings of INFOCOM '05, pp. 225–236. IEEE, Cambridge, MA, USA (2005)
14. Locher, T., Schmid, S., Watternhofer, R.: eQuus: A provably robust and locality-aware peer-to-peer system. In: Proceedings of the 6th IEEE International Conference on Peer-to-Peer Computing, pp. 3–11. Cambridge, UK (2006)
15. Loguinov, D., Kumar, A., Rai, V., Ganesh, S.: Graph-theoretic analysis of structured peer-to-peer systems: Routing distances and fault resilience. In: Proceedings of ACM SIGCOMM'03, pp. 395–406. Karlsruhe, Germany (2003)
16. Manku, G.S., Bawa, M., Raghavan, P.: Symphony: Distributed hashing in a small world. In: Proceedings of the 4th USENIX Symposium on Internet Technologies and Systems, pp. 127–140. Seattle, WA, USA (2003)
17. Maymounkov, P., Mazieres, D.: Kademlia: A peer-to-peer information system based on the xor metric. In: Proceedings of the 1st International Workshop on Peer-to-Peer Systems, LNCS 2429, pp. 53–65. Cambridge, MA, USA (2002)

18. Rao, K.R., Yip, P.: Discrete cosine transform: algorithms, advantages, applications. Academic Press, San Diego, CA, USA (1990)
19. Ratnasamy, S., Francis, P., Handley, M., Karp, R., Shenker, S.: A scalable content-addressable network. In: Proceedings of ACM SIGCOMM, pp. 161–172. San Diego, CA, USA (2001)
20. Rhea, S., Geels, D., Roscoe, T., Kubiatowicz, J.: Handling churn in a DHT. Tech. Rep. UCB/CSD-3-1299, UC Berkeley, Computer Science Division, UC Berkeley, USA (2003)
21. Rowstron, A., Druschel, P.: Pastry: Scalable, distributed object location and routing for large-scale peer-to-peer systems. In: Proceedings of the 18th IFIP/ACM International Conference on Distributed Systems Platforms, LNCS 2218, pp. 329–350. Springer, Heidelberg, Germany (2001)
22. Stoica, I., Morris, R., Karger, D., Kaashoek, M.F., Balakrishnan, H.: Chord: Scalable peer-to-peer lookup service for internet applications. In: Proceedings of ACM SIGCOMM, pp. 149–160. San Diego, CA, USA (2001)
23. Wu, D., Tian, Y., Ng, K.W.: Analytical study on improving DHT lookup performance under churn. In: Proceedings of the 6th IEEE International Conference on Peer-to-Peer Computing, pp. 249–258. IEEE, Cambridge, UK (2006)
24. Harmonic number definition. `http://mathworld.wolfram.com/HarmonicNumber.html`
25. Xu, J., Kumar, A., Yu, X.: On the fundamental tradeoffs between routing table size and network diameter in peer-to-peer networks. IEEE Journal on Selected Areas in Communications **22**(1), 151–163 (2004)