

## 8 Decentralized overlay networks I

In the next two sections we present overlay networks that are completely decentralized, i.e., they do not depend on a supervisor. We assume that, in principle, every peer has the right to initiate the integration of new peers into the system and that every peer knows at least one peer currently in the system so that publicly available entry points such as a supervisor are not necessary any more. In this section, we will focus on overlay networks that are based on the continuous-discrete approach [5], and in the next section overlay networks are presented that are based on so-called skip graphs [2]. Each each section, we will first assume that all peers are reliable and honest, and then we show how to remove this assumption.

### 8.1 Overlay networks for reliable, honest peers

Recall the hierarchical decomposition approach in Section 7. Our basic approach for the decentralized case is that each time a new peer  $p$  enters the system, a *random* label  $\ell(p) \in \{0, 1\}^*$  is generated for it, and this will be used to assign it to some place in the decomposition tree so that the following invariant is maintained (which only differs from the corresponding invariant in Section 7 by the missing item 3, which demands that peers are located at close-by levels).

#### Invariant 8.1

1. *The subcubes of the (nodes assigned to the) peers are disjoint and*
2. *the union of the subcubes of the peers gives the entire set  $U$ .*

In order preserve this invariant, the following is done when peers join or leave the system.

#### Joining the system

When a new peer  $p$  joins the system, it follows down the decomposition tree (which is simulated by a proper routing scheme in the given overlay network) according to its label  $\ell(p)$  until it arrives at some node  $v$  that is currently occupied by peer  $q$ . Then  $q$  is moved to node  $v0$  and  $p$  is moved to node  $v1$  of the decomposition tree (i.e.,  $q$  splits its region into two pieces and gives one of them to  $p$ ) so that Invariant 8.1 still holds.

#### Leaving the system

When a peer  $p$  at node  $v$  in the decomposition tree leaves, the peer at the lowest position in the decomposition tree that is reachable from the sibling of  $v$ , say  $q$ , is taken to replace the position of  $p$  (and thereby takes over its region).  $q$  must have a peer  $q'$  at its prior sibling in the decomposition tree which is moved to the parent of its node to take over  $q$ 's old region.

Given the two rules, one can show the following result for  $U = [0, 1)^d$  [1].

**Lemma 8.2** *Suppose that there are  $n$  peers in the system. If the join-leave activity of the peers is independent of their labels, then the level of every peer in the decomposition tree is within  $\log n \pm (\log_d \log n + O(1))$ , w.h.p.*

The lemma implies that the sizes of the regions assigned to the peers only differ by a factor of  $O(\log n)$ . Often, using the decomposition tree approach is overly complicated, especially when  $U = [0, 1)$ . In this case, a much simpler strategy is to use the consistent hashing approach in order to partition  $[0, 1)$  among the peers:

- Every peer  $p$  is assigned to some random point  $x_p \in [0, 1)$ .
- Every peer  $p$  is responsible for the region  $R_p = [x_p, \text{succ}(x_p))$  where  $\text{succ}(x_p)$  is the closest point succeeding  $x_p$  in  $[0, 1)$  that is occupied by a peer.

Using this rule, it is obvious that the regions are pairwise disjoint and that  $\bigcup_{R_v} = [0, 1)$ , which is necessary for applying the continuous-discrete approach. For this strategy, it holds:

**Lemma 8.3** *Suppose that there are  $n$  peers in the system. If the join-leave activity of the peers is independent of their positions, then every peer is responsible for a region of size at least  $\Omega(1/n^3)$  and most  $O(\log n/n)$ , w.h.p.*

**Proof.** We first prove the upper bound. Consider any interval  $I$  of size  $(c \ln n)/n$  for some sufficiently large constant  $c > 0$ . The probability that none of the peers has its point in  $I$  is equal to

$$\left(1 - \frac{c \ln n}{n}\right)^n \leq e^{-((c \ln n)/n) \cdot n} = e^{-c \ln n} = n^{-c}.$$

Hence, when partitioning  $[0, 1)$  into  $n/(c \log n)$  such intervals, every one of these has at least one point in them, w.h.p. Thus, a peer can be responsible for a region of size at most  $O(\log n/n)$ , w.h.p.

Next we prove the lower bound. The probability that any two peer positions have a distance of less than  $1/n^3$  is at most

$$\binom{n}{2} \frac{1}{n^3} \leq \frac{1}{2n}$$

Hence, the probability is very low that such a case occurs, completing the proof.  $\square$

The upper bound is acceptable though much better results can be achieved when performing local load balancing. In fact, evenly balancing the region size among the  $\Theta(\log n)$  closest peers in  $[0, 1)$  would ensure that the region size of every peer is  $\Theta(1/n)$ , w.h.p., which is implied by the following lemma.

**Lemma 8.4** *Given  $n$  peers, every interval of size  $\Theta(\log n/n)$  has  $\Theta(\log n)$  peers in it, w.h.p.*

**Proof.** Consider some fixed interval  $I$  of size  $(c \ln n)/n$  for some sufficiently large constant  $c > 0$ . For every peer  $v$  let the binary random variable  $X_v$  be 1 if and only if  $x_v \in I$ . Let  $X = \sum_{v \in V} X_v$ . It holds that

$$\mathbb{E}[X_v] = \Pr[X_v = 1] = \frac{c \ln n}{n}$$

and from the linearity of expectation it follows that

$$\mathbb{E}[X] = \sum_{v \in V} \mathbb{E}[X_v] = n \cdot \frac{c \ln n}{n} = c \ln n.$$

Hence, when using the well-known Chernoff bounds, we obtain that

$$\Pr[X \geq (1 + \epsilon)E[X]] \leq e^{-\epsilon^2 E[X]/3} = e^{-\epsilon^2 c \ln n/3} = n^{-\epsilon^2 c/3}$$

and

$$\Pr[X \leq (1 - \epsilon)E[X]] \leq e^{-\epsilon^2 E[X]/2} = n^{-\epsilon^2 c/3}$$

for all  $0 \leq \epsilon \leq 1$ . Thus, the probability is polynomially small in  $n$  that the bound in the lemma is violated.  $\square$

Next we consider specific examples of decentralized overlay networks based on  $U = [0, 1)$ . We start with the dynamic hypercube, and then we consider the dynamic de Bruijn network.

### The dynamic hypercube

Recall the definition of the continuous form of the hypercube. According to this definition,

- $U = [0, 1)$  and
- $F = \{f_i^-, f_i^+ \mid i \geq 1\}$  with  $f_i^-(x) = (x - 1/2^i) \bmod 1$  and  $f_i^+(x) = (x + 1/2^i) \bmod 1$ .

The dynamic hypercube for the decentralized case is based on the continuous-discrete approach together with a cycle connecting each peer to its predecessor and successor in  $[0, 1)$ . If the peers are assigned to random points in  $[0, 1)$ , the results in the previous section together with Lemmas 8.3 and 8.4 imply the following result.

**Lemma 8.5** *Given  $n$  peers, the dynamic hypercube has a maximum degree of  $O(\log^2 n)$ , w.h.p.*

### Routing in a dynamic hypercube

Suppose that we want to route a message from point  $x$  to point  $y$  in  $[0, 1)$ . Let  $(x_1, x_2, \dots)$  be the binary representation of  $x$  and  $(y_1, y_2, \dots)$  be the binary representation of  $y$ . Then we use the following continuous strategy to route the message from  $x$  to  $y$ :

$$(x_1, x_2, \dots) \rightarrow (y_1, x_2, \dots) \rightarrow (y_1, y_2, x_3, \dots) \rightarrow \dots$$

If  $x$  and  $y$  have infinite binary representations, then this strategy may take an infinite amount of hops, but in the discrete world with a finite number of peers, this is not the case with the following discrete variant of the continuous routing strategy above:

The message starts at the peer  $v_0$  responsible for  $x$ . Peer  $v_0$  forwards the message to the peer  $v_1$  responsible for  $(y_1, x_2, \dots)$ , peer  $v_1$  forwards it to the peer  $v_2$  responsible for  $(y_1, y_2, x_3, \dots)$ , and so on, until the message reaches a peer  $v_\ell$  whose region or whose neighboring region contains  $y$ . From this peer the message is forwarded to the peer responsible for the region containing  $y$ .

Notice that the maximal remaining distance to  $y$  shrinks by a factor 2 in each step. Hence, once a distance equal to the smallest region is reached, the routing terminates. Thus, the following theorem immediately follows from Lemma 8.3.

**Theorem 8.6** *Using the continuous-discrete routing strategy, it takes at most  $O(\log n)$  hops until a message is routed from any point  $x$  to any point  $y$  in  $[0, 1)$ .*

Besides having a small dilation, it is also important to have a small congestion, i.e., when routing multiple messages, the maximum number of messages to be handled by a peer should be as close to optimal as possible. In order to achieve a low congestion, the following routing strategy may be used by any peer, which is a continuous version of Valiant's trick:

Suppose that a peer with position  $x$  wants to send a message to position  $y$ . Then it chooses a random point  $z \in [0, 1)$ , first routes the message from  $x$  to  $z$  and then from  $z$  to  $y$  using the continuous-discrete routing strategy above.

With this strategy, we obtain the following theorem.

**Theorem 8.7** *For every permutation routing problem, the congestion caused when using the extended continuous-discrete routing strategy above is at most  $O(\log^2 n)$ , w.h.p.*

**Proof.** Consider any permutation routing problem  $\pi$ , and consider cutting  $[0, 1)$  into  $n'$  intervals of size  $1/n'$  starting at integral multiples of  $1/n'$  where  $n'$  is chosen so that  $n'$  is a power of 2 and  $1/n' = (c \ln n)/n$  for some suitably chosen constant  $c$ . It follows from Lemma 8.4 that every interval has  $O(\log n')$  packets starting at it and  $O(\log n')$  packets aiming for it. Viewing these intervals as the nodes of a  $\log n'$ -dimensional hypercube, it follows from the analysis of Valiant's trick that at most  $O(\log^2 n')$  packets pass every node, w.h.p. Since, according to Lemma 8.3, every peer is responsible for an interval of size at most  $O(\log n/n)$ , this implies that every peer is passed by at most  $O(\log^2 n)$  packets, w.h.p., which proves the theorem.  $\square$

### Joining and leaving a dynamic hypercube

We only consider isolated executions of join and leave requests because otherwise it can be quite tricky to correctly update the network.

Suppose that a new peer  $v$  contacts some peer  $w$  already in the system to join the system. Then  $v$ 's request is first sent to the peer  $u$  owning  $x_v$  using the continuous-discrete routing strategy, which only takes  $O(\log n)$  hops according to Theorem 8.6.  $u$  forwards information about all of its outgoing edges to  $v$ , deletes all edges that it does not need any more, and informs the corresponding endpoints about this. Because  $R(v) \subseteq R(u)$  for the old  $R(u)$ , the edges reported to  $v$  are a superset of the edges that it needs to establish.  $v$  checks which of the edges are relevant for it, informs the other endpoint for each relevant edge, and removes the others.

If a peer  $v$  wants to leave the network, it simply forwards all of its outgoing edges to the peer at  $\text{pred}(x_v)$ . That peer will then merge these edges with its existing edges and notify the endpoints of these edges about the changes.

We know from Theorem 8.6 that the routing part only takes  $O(\log n)$  hops. Furthermore, Lemmas 8.3 and 8.4 imply that every peer has at most  $O(\log^2 n)$  incoming and outgoing edges. Hence, we obtain the following theorem.

**Theorem 8.8** *Join and leave require at most  $O(\log^2 n)$  work and  $O(\log n)$  communication rounds, w.h.p.*

### Data management

Suppose that we want to store data in the dynamic hypercube. Here we can simply use the consistent hashing strategy in Section 5: data items are hashed to random values in  $[0, 1)$  using a pseudo-random

hash function  $h$  (which is known to all peers), and every data item  $d$  is stored in the peer  $v$  with  $h(d) \in R_v$ .

Using this strategy, data will on expectation be evenly distributed among the peers, and on expectation, at most a factor of 2 more data than necessary has to be replaced if a node joins or leaves. (Recall the section on hashing.)

### A dynamic de Bruijn network

Recall the definition of the continuous de Bruijn graph. According to this definition,

- $U = [0, 1)$  and
- $F = \{f_0, f_1\}$  with  $f_0(x) = x/2$  and  $f_1(x) = (1 + x)/2$ .

The dynamic de Bruijn graph for the decentralized case is based on the continuous-discrete approach together with a cycle connecting each peer to its predecessor and successor in  $[0, 1)$ . It follows from Lemmas 8.3 and 8.4:

**Lemma 8.9** *Given  $n$  peers, the dynamic de Bruijn network has a maximum degree of  $O(\log n)$ , w.h.p.*

### Routing in a dynamic de Bruijn network

Suppose that we want to route a message from point  $x$  to point  $y$  in  $[0, 1)$ . Let  $(x_1, x_2, \dots)$  be the binary representation of  $x$  and  $(y_1, y_2, \dots)$  be the binary representation of  $y$ . Then we use the following continuous-discrete strategy to route the message from  $x$  to  $y$ :

The message starts at the peer  $v_0$  responsible for  $x$ . If  $y$  is not in the region of  $v_0$  or one of its neighbors,  $v_0$  chooses a random bit  $z_1$  and forwards the message to the peer  $v_1$  owning  $(z_1 x_1 x_2 \dots)$ . In general, if the message has reached a peer  $v_i$  and  $(z_i \dots z_1 y_1 y_2 \dots)$  is not in the region of  $v_i$  or one of its neighbors,  $v_i$  chooses a random  $z_{i+1}$  and forwards the message to the peer owning  $(z_{i+1} \dots z_1 x_1 x_2 \dots)$ . Once the message has reached a peer  $w_i$  owning  $(z_i \dots z_1 y_1 y_2 \dots)$ , it is forwarded to the peer  $w_{i-1}$  owning  $(z_{i-1} \dots z_1 y_1 y_2 \dots)$  and so on until it reaches the peer owning  $(y_1 y_2 \dots)$ .

All of the hops in the routing strategy can be performed along edges of the dynamic de Bruijn graph. Also, notice that the maximum distance between  $(z_i \dots z_1 x_1 x_2 \dots)$  and  $(z_i \dots z_1 y_1 y_2 \dots)$  shrinks by a factor of 2 in each hop. Hence, once a distance equal to the smallest region is reached, the routing strategy can move from a peer  $v_i$  to a peer  $w_i$ . Thus, the following theorem immediately follows from Lemma 8.3.

**Theorem 8.10** *Using the continuous-discrete routing strategy, it takes at most  $O(\log n)$  hops until a message is routing from any point  $x$  to any point  $y$  in  $[0, 1)$ .*

The continuous-discrete routing strategy does not only have a low dilation but also a small congestion because it has Valiant's trick already built into it. Hence, the following theorem holds using similar arguments as in Theorem 8.7.

**Theorem 8.11** *For every permutation routing problem, the congestion caused when using the continuous-discrete routing strategy above is at most  $O(\log^2 n)$ , w.h.p.*

## Joining and leaving a dynamic de Bruijn network

Suppose that a new peer  $v$  contacts some peer  $w$  already in the system to join the system. Then  $v$ 's request is first sent to the peer  $u$  owning  $x_v$  using the continuous-discrete routing strategy above, which only takes  $O(\log n)$  hops according to Theorem 8.10.  $u$  forwards information about all of its (incoming and) outgoing edges to  $v$ , deletes all edges that it does not need any more, and informs the corresponding endpoints about this. Because  $R(v) \subseteq R(u)$  for the old  $R(u)$ , the edges reported to  $v$  are a superset of the edges that it needs to establish.  $v$  checks which of the edges are relevant for it, informs the other endpoint for each relevant edge, and removes the others.

If a node  $v$  wants to leave the network, it simply forwards all of its outgoing edges to the peer at  $\text{pred}(x_v)$ . That peer will then merge these edges with its existing edges and notifies the endpoints of these edges about the changes.

We know from Theorem 8.10 that the routing part only takes  $O(\log n)$  hops. Furthermore, Lemmas 8.3 and 8.4 imply that every peer has at most  $O(\log n)$  outgoing edges. Hence, we obtain the following theorem.

**Theorem 8.12** *Join and leave take at most  $O(\log n)$  work and  $O(\log n)$  communication rounds, w.h.p.*

Also the dynamic de Bruijn network can be used for data management with the help of the consistent hashing approach.

## 8.2 Robustness against random faults

In order to protect against random faults in dynamic networks based on  $U = [0, 1)$ , each peer  $v$  aims at preserving the following invariant (which is similar to the supervised case).

**Invariant 8.13** *Every peer  $v$  in the system is connected to*

- $\text{pred}_i(v)$  and  $\text{succ}_i(v)$  for every  $i \in \{1, \dots, k\}$  and
- all peers  $w$  with the property that  $\Gamma(R(N_v)) \cap R(N_w) \neq \emptyset$

where  $N_v = \{v\} \cup \{\text{pred}_i(v) \mid i \in \{1, \dots, k\}\} \cup \{\text{succ}_i(v) \mid i \in \{1, \dots, k\}\}$  and for any set  $V' \subseteq V$ ,  $R(V') = \bigcup_{v \in V'} R(v)$ .

When doing this, the following result can be shown.

**Theorem 8.14** *If faults of peers are independent of their positions and only happen at a (sufficiently small) constant rate, then the peers can maintain the the Invariant 8.13 everywhere, w.h.p.*

The result holds since for a sufficiently small constant fault rate and a sufficiently large  $k$ , no entire sequence of  $k$  consecutive peers in  $[0, 1)$  will become faulty within a constant number of steps, w.h.p., and every peer only needs a constant number of communication rounds to update its connectivity information as peers among its predecessors and successors become faulty as long as this is in principle possible (i.e., not all of its predecessors and successors fail).

### 8.3 Robustness against adversarial join-leave behavior

Finally, we consider the problem of protecting an overlay network against adversarial join-leave behavior. More precisely, we consider the following scenario. There are  $n$  blue (or honest) nodes and  $\epsilon n$  red (or adversarial) nodes for some fixed constant  $\epsilon < 1$ . There is a rejoin operation that, when applied to node  $v$ , lets  $v$  first leave the system and then join it again from scratch. The leaving is done by simply removing  $v$  from the system and the joining is done with the help of a join operation to be specified by the system. We assume that the sequence of rejoin requests is controlled by an adversary, which is a typical assumption in the analysis of online algorithms. The adversary can only issue rejoin requests for the red nodes, but it can do this in an arbitrary adaptive manner. That is, at any time it can inspect the entire system and select whatever red node it likes to rejoin the system. Our goal is to find an *oblivious* join strategy, i.e., a strategy that cannot distinguish between the blue and red nodes, so that for *any* adversarial strategy above the following two conditions can be preserved for every interval  $I \subseteq [0, 1)$  of size at least  $(c \log n)/n$  for a constant  $c > 0$  and any polynomial number of rounds in  $n$ :

- *Balancing condition:*  $I$  contains  $\Theta(|I| \cdot n)$  nodes.
- *Majority condition:* the blue nodes in  $I$  are in the majority.

It is not difficult to see that the brute-force strategy of giving every node a new random place whenever a node rejoins will achieve the stated goal, with high probability, but this would be a very expensive strategy. The challenge is to find a join operation that needs as little randomness and as few rearrangements as possible to satisfy the two conditions. Fortunately, there is such a strategy, called the *cuckoo rule*. We first introduce some notation, and then we describe the strategy.

In the following, a *region* is an interval of size  $1/2^r$  in  $[0, 1)$  for some integer  $r$  that starts at an integer multiple of  $1/2^r$ . Hence, there are exactly  $2^r$  regions of size  $1/2^r$ . A *k-region* is a region of size (closest from above to)  $k/n$ , and for any point  $x \in [0, 1)$ , the *k-region*  $R_k(x)$  is the unique *k-region* containing  $x$ .

**Cuckoo rule:** If a new node  $v$  wants to join the system, pick a random  $x \in [0, 1)$ . Place  $v$  into  $x$  and move all nodes in  $R_k(x)$  to points in  $[0, 1)$  chosen uniformly and independently at random (without replacing any further nodes).

See Figure 1 for an illustration of the cuckoo rule. The following result can be shown [3]:

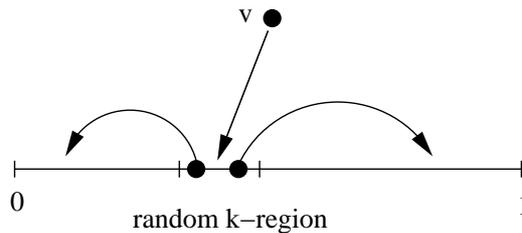


Figure 1: The cuckoo rule. (a) A new node is placed at a random point  $x$ . (b) Old nodes in  $R_k(x)$  are evicted and moved to new, random places.

**Theorem 8.15** *For any constants  $\epsilon$  and  $k$  with  $\epsilon < 1 - 1/k$ , the cuckoo rule with parameter  $k$  satisfies the balancing and majority conditions for a polynomial number of rounds, with high probability, for any adversarial strategy within our model. The inequality  $\epsilon < 1 - 1/k$  is sharp as counterexamples can be constructed otherwise.*

We just sketch the proof of the theorem. Let  $\hat{R}$  be any fixed region of size  $(c \log n) \cdot k/n$ , for some constant  $c$ , for which we want to check the balancing and majority conditions over polynomial in  $n$  many steps. Thus,  $\hat{R}$  contains exactly  $c \log n$  many  $k$ -regions. The *age* of a  $k$ -region is the difference between the current round and the last round when a new node was placed into it (and all old nodes got evicted), and the age of  $\hat{R}$  is defined as the sum of the ages of its  $k$ -regions. A node in  $\hat{R}$  is called *new* if it was placed in  $\hat{R}$  when it joined the system, and otherwise it is called *old*. The first lemma follows directly from the cuckoo rule because every  $k$ -region can have at most one new node at any time.

**Lemma 8.16** *At any time,  $\hat{R}$  contains at most  $c \log n$  new nodes.*

In order to bound the number of old nodes in  $\hat{R}$ , we first have to bound the age of  $\hat{R}$  (Lemma 8.17). Then we bound the maximum number of nodes in a  $k$ -region (Lemma 8.18) and use this to bound the number of evicted blue and red nodes in a certain time interval (Lemma 8.19). After that, we can combine all lemmas to bound the number of old blue and red nodes in  $\hat{R}$  (Lemma 8.20).

**Lemma 8.17** *At any time,  $\hat{R}$  has an age within  $(1 \pm \delta)(c \log n) n/k$ , with high probability, where  $\delta > 0$  is a constant that can be made arbitrarily small depending on the constant  $c$ .*

**Lemma 8.18** *For any  $k$ -region  $R$  in  $\hat{R}$  it holds at any time that  $R$  has at most  $O(k \log n)$  nodes, with high probability.*

Next we bound the number of blue and red nodes that are evicted in a certain time interval.

**Lemma 8.19** *For any time interval  $I$  of size  $T = (\gamma/\epsilon) \log^3 n$ , the number of blue nodes that are evicted in  $I$  is within  $(1 \pm \delta)T \cdot k$ , with high probability, and the number of red nodes that are evicted in  $I$  is within  $(1 \pm \delta)T \cdot \epsilon k$ , with high probability, where  $\delta > 0$  can be made arbitrarily small depending on  $\gamma$ .*

Combining Lemmas 8.17 to 8.19, we obtain the following lemma.

**Lemma 8.20** *At any time,  $\hat{R}$  has within  $(1 \pm \delta)(c \log n) \cdot k$  old blue nodes and within  $(1 \pm \delta)(c \log n) \cdot \epsilon k$  old red nodes, with high probability, where the lower bound on the red nodes holds if none of the red nodes has rejoined.*

Combining Lemmas 8.16 and 8.20, we can now prove when the balancing and majority conditions are satisfied.

- **Balancing condition:** From Lemmas 8.16 and 8.20 it follows that every region  $R$  of size  $(c \log n)k/n$  has at least  $(1 - \delta)(c \log n) \cdot k$  and at most  $(1 + \delta)(c \log n)k + (c \log n)\epsilon k = (1 + \delta)(c \log n)(1 + (1 + \epsilon)k)$  nodes, where the constant  $\delta > 0$  can be made arbitrarily small. Hence, the regions are balanced within a factor of close to  $(1 + \epsilon + 1/k)$ .

- **Majority condition:** From Lemmas 8.16 and 8.20 it also follows that every region of size  $(c \log n)k/n$  has at least  $(1 - \delta)(c \log n) \cdot k$  blue nodes and at most  $(1 + \delta)(c \log n + (c \log n) \cdot \epsilon k)$  red nodes, w.h.p., where the constant  $\delta > 0$  can be made arbitrarily small. These bounds are also tight in the worst case, which happens if the adversary focuses on a specific region  $R$  of size  $(c \log n)k/n$  and continuously rejoins with any red node outside of  $R$ . Hence, the adversary is not able to obtain the majority in any region of size  $(c \log n)k/n$  as long as  $(c \log n)(\epsilon k + 1) < (c \log n) \cdot k$  which is true if and only if  $\epsilon < 1 - 1/k$ .

Hence, for  $\epsilon < 1 - 1/k$  the balancing and majority conditions are satisfied, w.h.p., and this is sharp, which proves Theorem 8.15.

The cuckoo rule has the drawback that it only works if only the red nodes show adversarial join-leave behavior. What if both kinds of nodes show adversarial join-leave behavior? Then we need to extend the cuckoo rule in the following way in order to main the balancing and majority conditions.

The join operation works in the same way as the cuckoo rule. But whenever a peer wants to leave the network, we use the following leave operation:

**Leave( $v$ ):** If a peer  $v$  leaves the system, then a  $k$ -region  $R$  is chosen uniformly at random among the  $k$ -regions of  $R_{kc \log n}(x)$  for some (sufficiently large) constant  $c$ , where  $x$  is the position of  $v$ .  $R$  is flipped with a  $k$ -region  $R'$  chosen uniformly at random in  $[0, 1)$ , and then all peers in  $R$  (as well as  $v$ ) have to rejoin the system from scratch using the cuckoo rule.

Hence, the departure of a peer may spawn several join operations. We call this algorithm the *cuckoo&flip strategy*. With this strategy the balancing and majority conditions can be kept, with high probability. More precisely, one can show the following result [4]:

**Theorem 8.21** *For any constants  $\epsilon$  and  $k$  with  $\epsilon < 1/4 - (2 \log k + 1)/k$ , the cuckoo&flip strategy satisfies the balancing and majority conditions for any polynomial number of rejoin requests, with high probability, for any adversarial strategy within our model.*

Hence, as long as  $\epsilon < 1/4$ , only a constant factor overhead has to be paid (on average) compared to standard join and leave operations without any additional replacements of peers.

## References

- [1] I. Abraham, B. Awerbuch, Y. Azar, Y. Bartal, D. Malkhi, and E. Pavlov. A generic scheme for building overlay networks in adversarial scenarios. In *Proc. of the 2nd International Workshop on Peer-to-Peer Systems (IPTPS)*, 2003.
- [2] J. Aspnes and G. Shah. Skip graphs. In *Proc. of the 14th ACM/SIAM Symp. on Discrete Algorithms (SODA)*, pages 384–393, 2003.
- [3] B. Awerbuch and C. Scheideler. Towards a scalable and robust DHT. In *Proc. of the 18th ACM Symp. on Parallel Algorithms and Architectures (SPAA)*, 2006.
- [4] B. Awerbuch and C. Scheideler. Towards scalable and robust overlay networks. In *Proc. of the 6th International Workshop on Peer-to-Peer Systems (IPTPS)*, 2007.
- [5] M. Naor and U. Wieder. Novel architectures for P2P applications: the continuous-discrete approach. In *Proc. of the 15th ACM Symp. on Parallel Algorithms and Architectures (SPAA)*, pages 50–59, 2003.