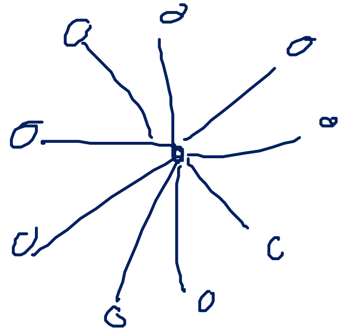
CoNe Freiburg

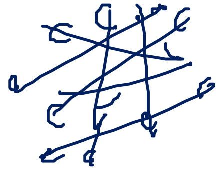
IIF INSTITUT FÜR INFORMATIK FREIBURG



Napster

Peer-to-Peer Networks

07 Degree Optimal Networks



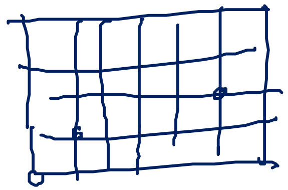
Gnutella

Christian Schindelhauer

Technical Faculty

Computer-Networks and Telematics

University of Freiburg



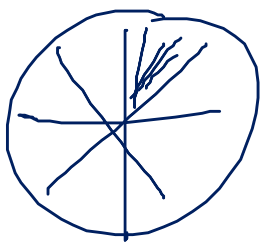
CAN

degree 4

$\sqrt{n}$

UNI FREIBURG

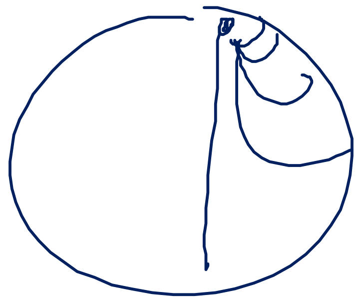
Pastry



degree $O(\log n)$

diameter $O(\log n)$

Chord



degree $O(\log n)$

diameter $O(\log n)$

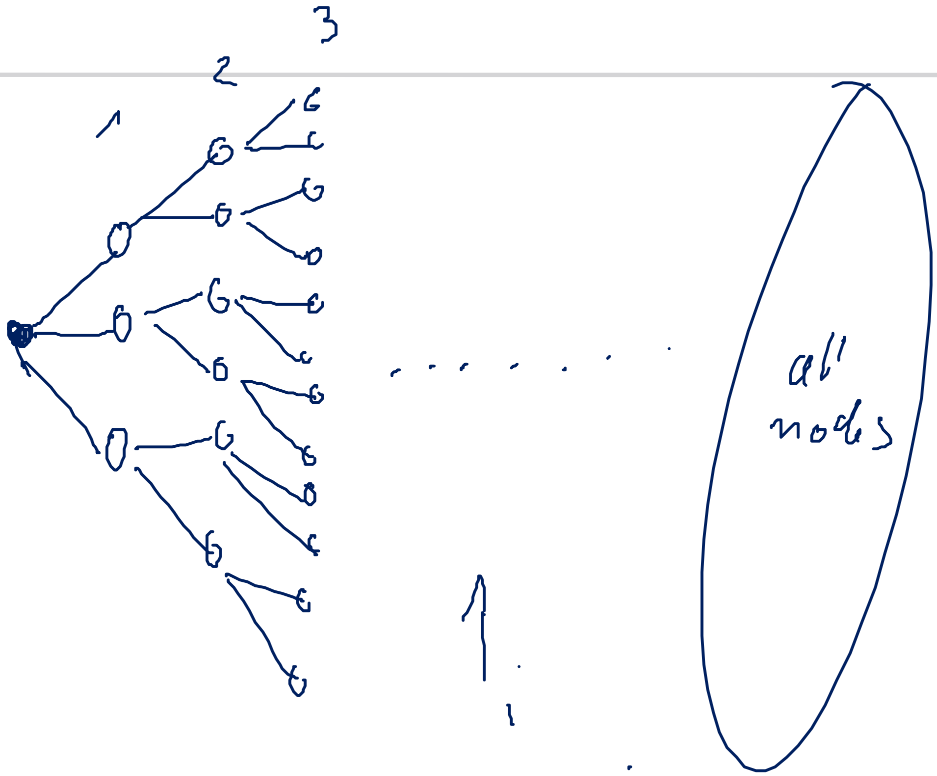
Diameter and Degree in Graphs

- CHORD:
  - degree O(log n)
  - diameter O(log n)
- Is it possible to reach a smaller diameter with degree g=O(log n)?
  - In the neighborhood of a node are at most g nodes
  - In the 2-neighborhood of node are at most g^2 nodes
  - ...
  - In the d-neighborhood of node are at most g^d nodes
- So, $$(\log n)^d = n$$
- Therefore $$d = \frac{\log n}{\log \log n}$$
- So, Chord is quite close to the optimum diameter.

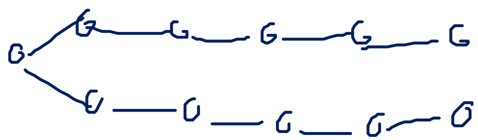
Are there P2P-Netzwerke with constant out-degree and diameter log n?

- CAN
  - degree: 4
  - diameter: $n^{1/2}$
- Can we reach diameter O(log n) with constant degree?

$1 \quad 3 \quad \overset{6}{3\cdot 2} \quad 3\cdot 4 \quad 3\cdot 8 \quad \ldots \quad n$



$3\cdot 2^{i-1}$



$$3\cdot 2^{i-1} \overset{!}{=} n$$

$$2^{i-1} \geq \frac{n}{3}$$

$$i-1 \geq \log_2 \frac{n}{3}$$

$$i \geq 1 + \underline{\log_2 n} - \log_2 3$$

CoNe Freiburg

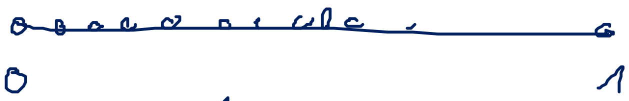
What is a Graph?

$G = (V, E)$

$E \subseteq V \times V$

$\frac{1}{2} \in V$

$V = [0, 1)$



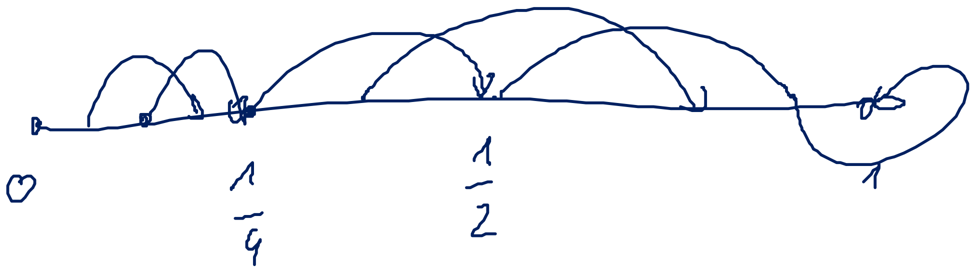
$\frac{1}{e} \in V$

$(u, 2u)$, $u \le \frac{1}{2}$

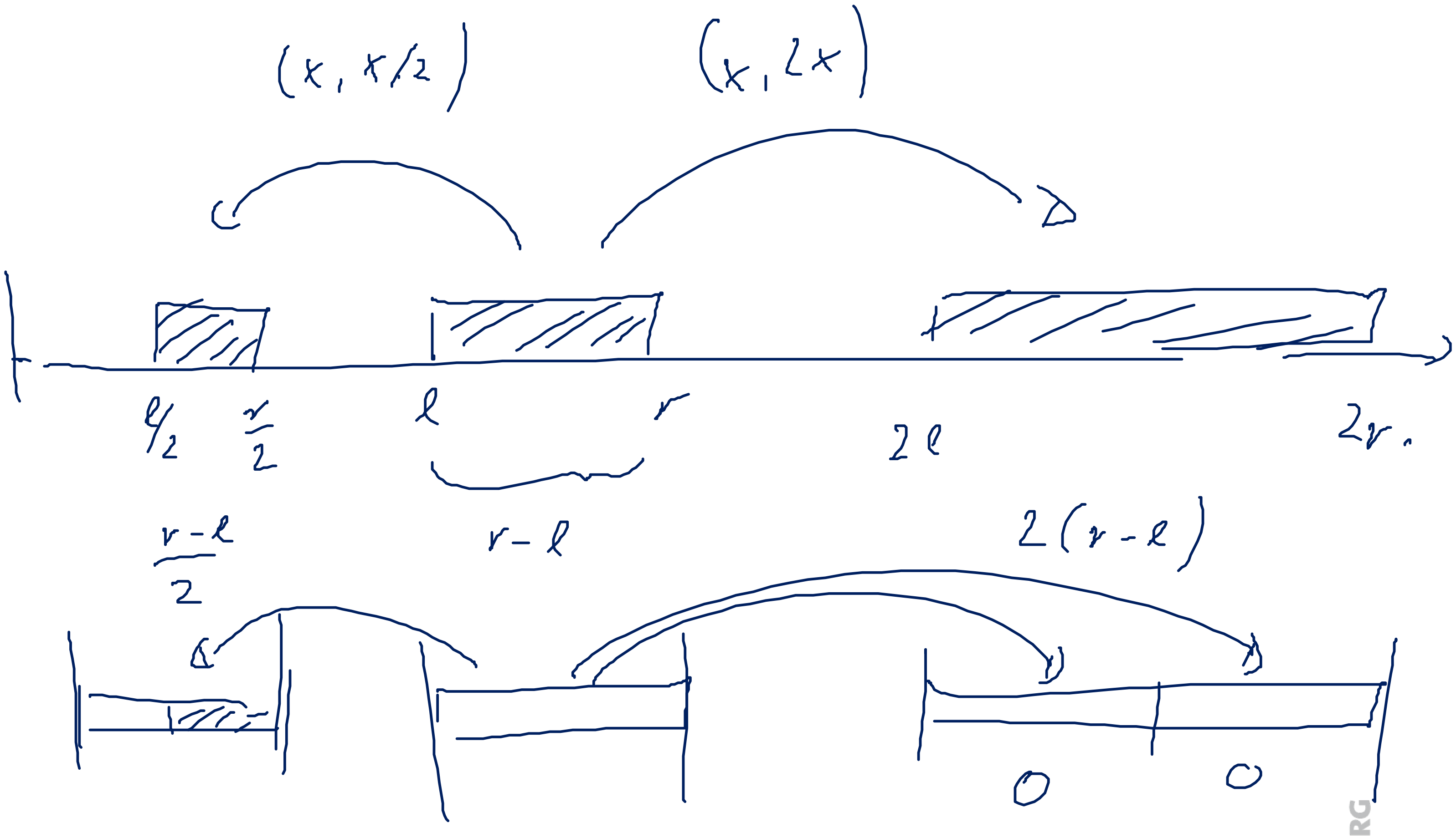
$(\underbrace{1-u}_{v}, \underbrace{1-2u}_{1-2+2v})$ $u \le \frac{1}{2}$

$(u, \frac{u}{2})$

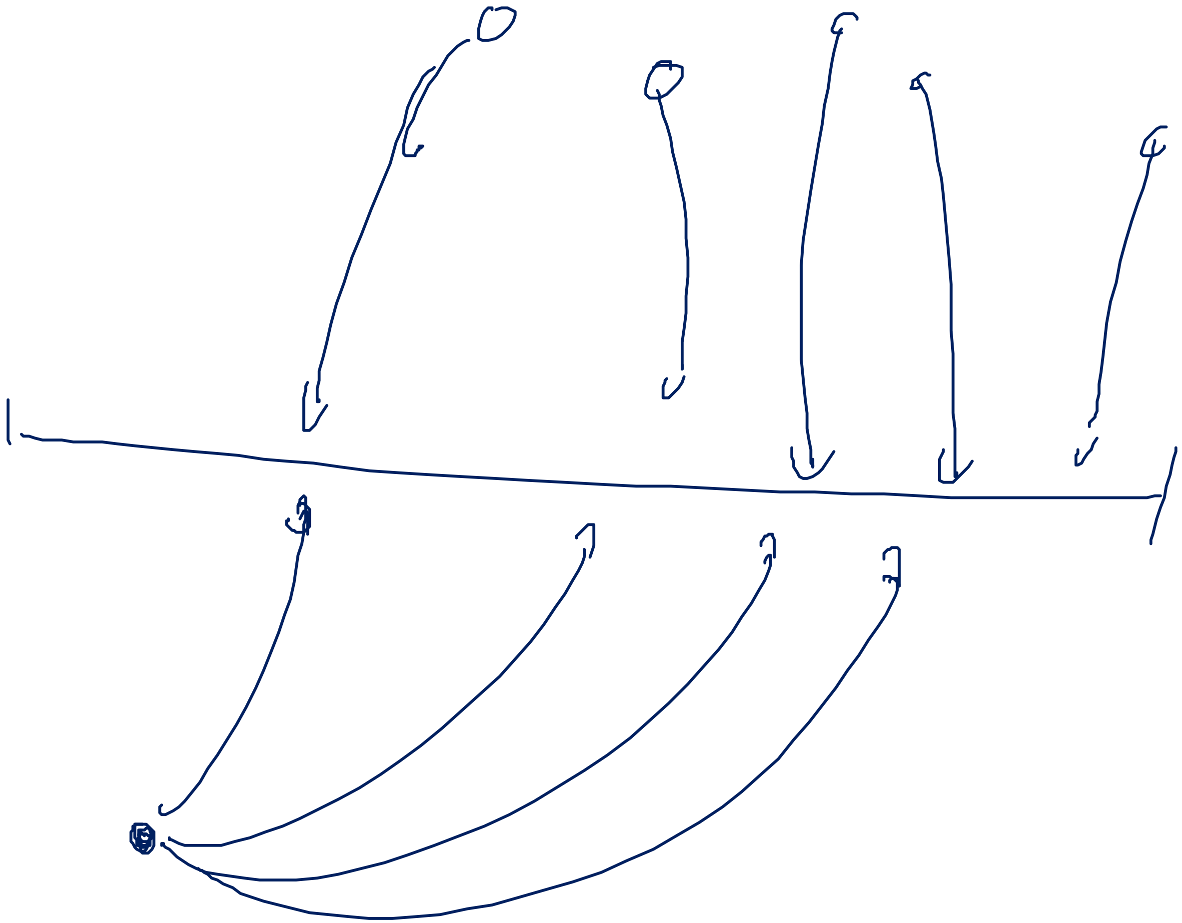
$(v, 2v-1)$ $v \ge \frac{1}{2}$



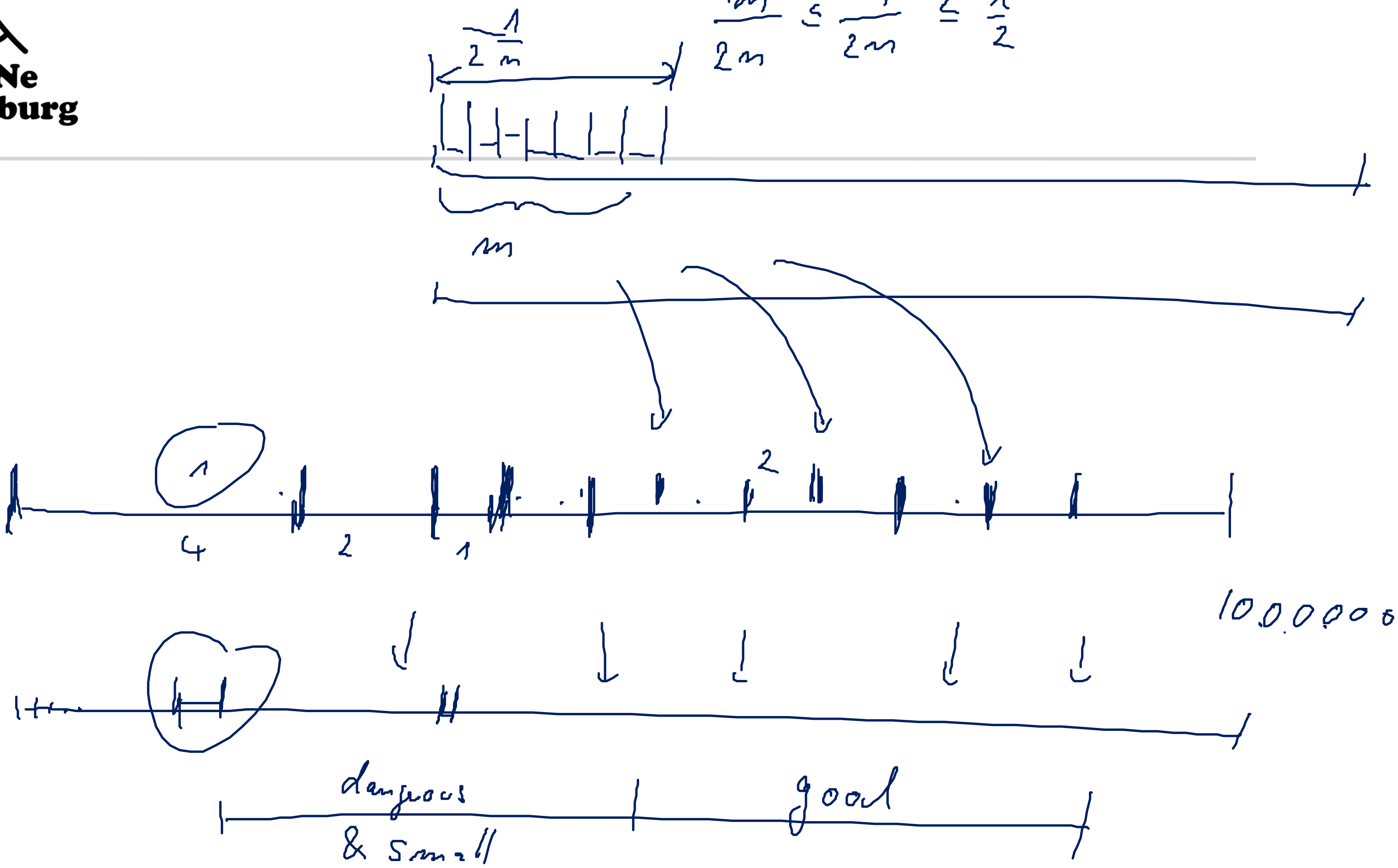
UNI FREIBURG



(x, x/2)
(x, 2x)
l/2
r/2
l
r
2l
2r.
(r-l)/2
r-l
2(r-l)
0
0



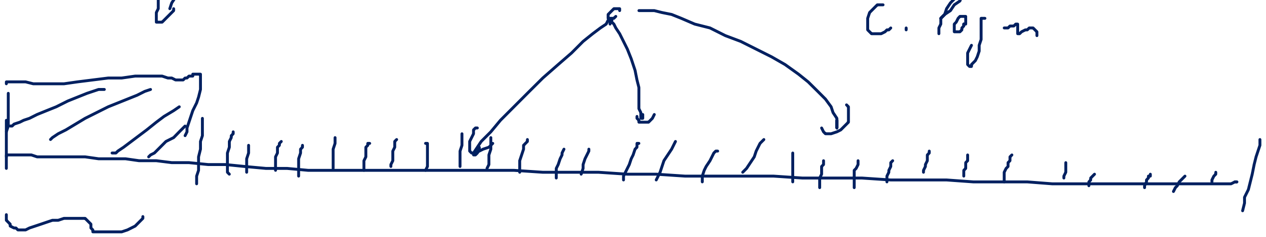
$$\frac{m}{2n} \le \frac{n}{2n} \le \frac{1}{2}$$



$$\le \frac{1}{2}$$

$$\underbrace{\frac{1}{2}\cdot\frac{1}{2}\cdots\frac{1}{2}}_{c\cdot\log n} = 2^{-c\cdot\log n} = \frac{1}{n^c}$$

kill the large interval



$c \cdot \log n$

$\frac{1}{n} \cdot 2$

$n \rightarrow 2n$

n

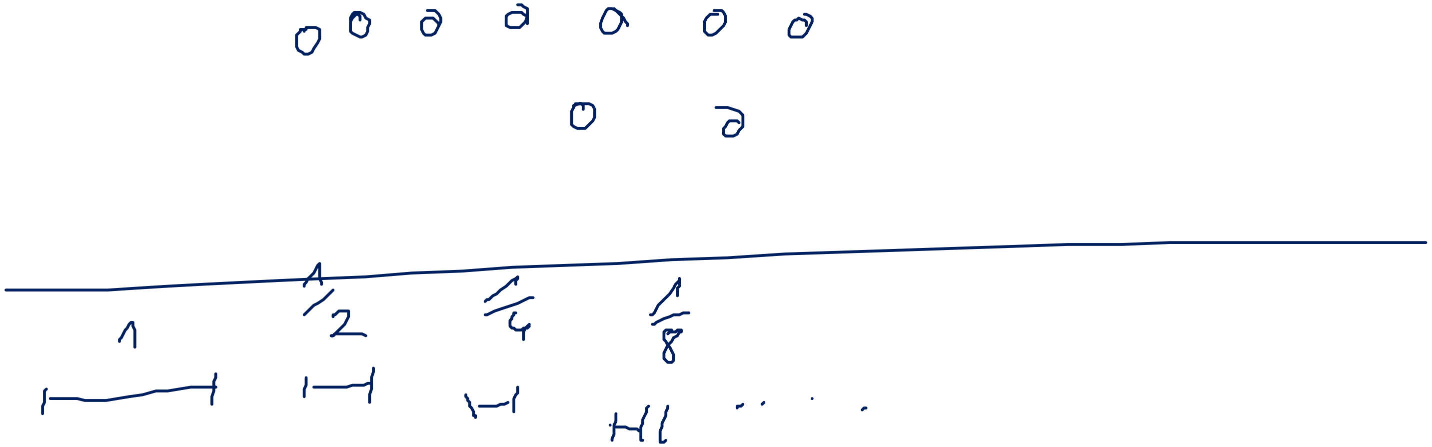
$\boxed{c \cdot n \cdot \log n}$ tries

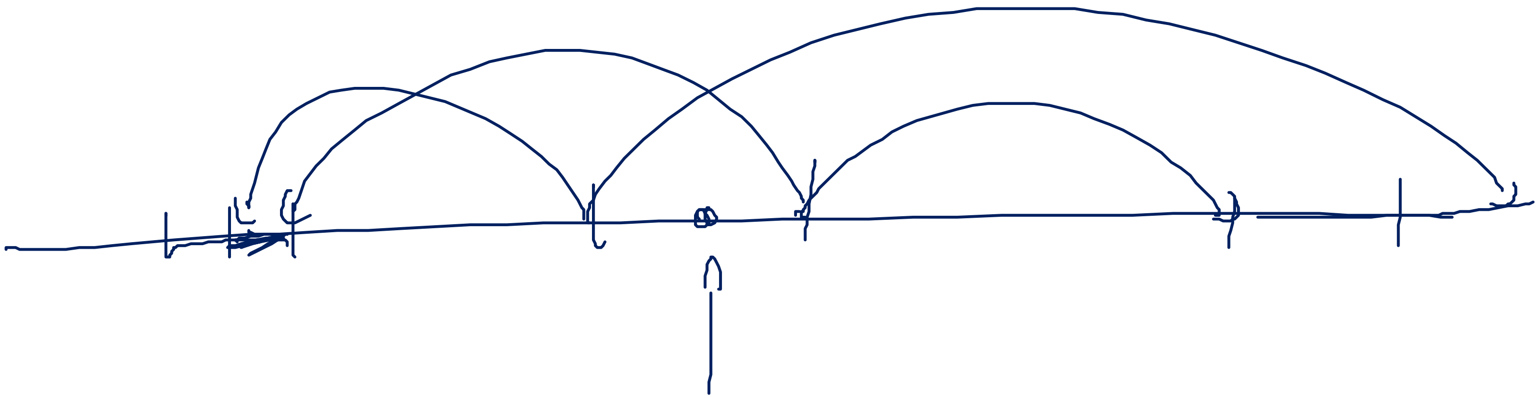
p

k.

$\frac{1}{n}$

$$\left(1 - \frac{1}{n}\right)^n = \frac{1}{e}$$


1
1/2
1/4
1/8



Distance Halving

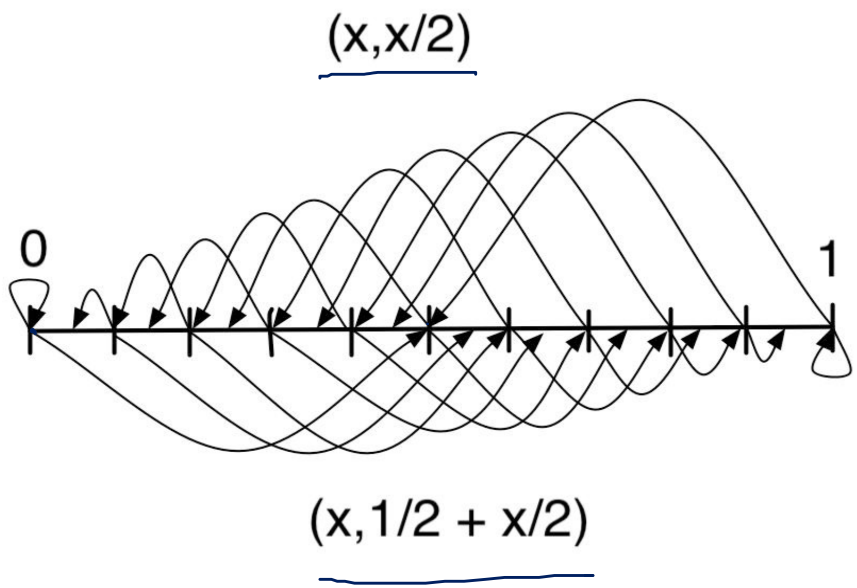
Moni Naor, Udi Wieder

2003

Continuous Graphs

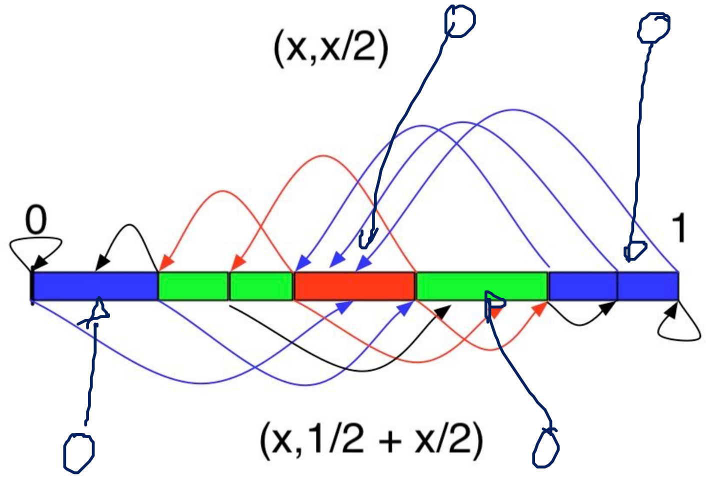
- are infinite graphs with continuous node sets and edge sets

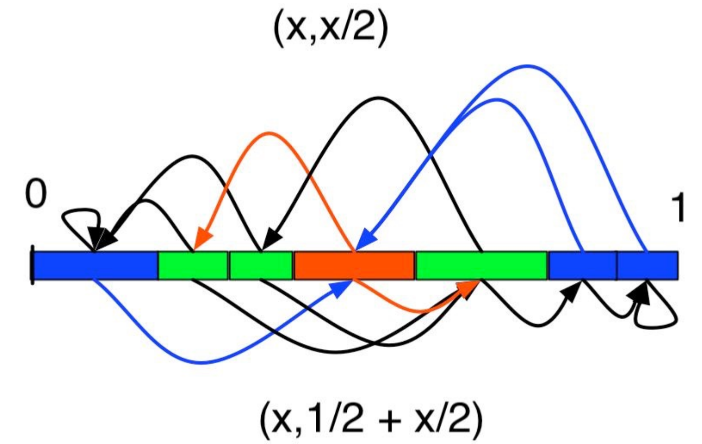
- The underlying graph
  - $x \in [0,1)$
  - Edges:
    - $(x, x/2)$, *left* edges
    - $(x, 1+x/2)$, *right* edges
  - plus revers edges.
    - $(x/2, x)$
    - $(1+x/2, x)$



The Transition from Continuous to Discrete Graphs

- Consider discrete intervals resulting from a partition of the continuous space
- Insert edge between interval A and B
  - if there exists $x \in A$ and $y \in B$ such that edge (x,y) exists in the continuous graph
- Intervals result from successive partitioning (halving) of existing intervals
- Therefore the degree is constant if
  - the ratio between the size of the largest and smallest interval is constant
- This can be guarranteed by the principle of multiple choice
  - which we present later on





Principle of Multiple Choice

- **Before inserted check c log n positions**
- **For position p(j) check the distance a(j) between potential left and right neighbor**
- **Insert element at position p(j) in the middle between left and right neighbor, where a(j) was the maximum choice**
- **Lemma**
  - After inserting n elements with high probability only intervals of size 1/(2n), 1/n und 2/n occur.

Proof of Lemma

1st Part: With high probability there is no interval of size larger than 2/n

follows from this Lemma

Lemma*

Let c/n be the largest interval. After inserting 2n/c peers all intervals are smaller than c/(2n) with high probability

From applying this lemma for c=n/2,n/4, ...,4 the first lemma follows.

from Chernoff!

Proof

- **2nd part: No intervals smaller than 1/(2n) occur**
  - The overall length of intervals of size 1/(2n) before inserting is at most 1/2
  - Such an area is hit with probability at most 1/2
  - The probability to hit this area more than c log n times is at least $2^{-c \log n} = n^{-c}$
  - Then for c>1 such an interval will not further be divided with probability into an interval of size 1/(4m).

Chernoff-Bound

- Theorem Chernoff Bound
  - Let $x_1,\ldots,x_n$ independent Bernoulli experiments with
    - $P[x_i = 1] = p$
    - $P[x_i = 0] = 1-p$
  - Let $S_n = \sum_{i=1}^{n} x_i$
  - Then for all c>0

$$\mathbf{P}[S_n \geq (1+c) \cdot \mathbf{E}[S_n]] \quad \leq \quad e^{-\frac{1}{3}\min\{c,c^2\}pn}$$

  - For $0 \leq c \leq 1$

$$\mathbf{P}[S_n \leq (1-c) \cdot \mathbf{E}[S_n]] \quad \leq \quad e^{-\frac{1}{2}c^2 pn}$$

Proof of Lemma*

- Consider the longest interval of size c/n. Then after inserting 2n/c peers all intervals are smaller than c/(2n) with high probability.
- Consider an interval of length c/n
- With probability c/n such an interval will be hit
- Assume, each peer considers t log n intervals
- The expected number of hits is therefore

$$E[X] = \frac{c}{n} \cdot \frac{2n}{c} \cdot t \log n = 2t \log n$$

- From the Chernoff bound it follows

$$P[X \leq (1-\delta)E[X]] \leq n^{-\delta^2 t}$$

- If $\delta^2 t \geq 2$ then this interval will be hit at least $2(1-\delta)t \log n$ times

n^{-2}

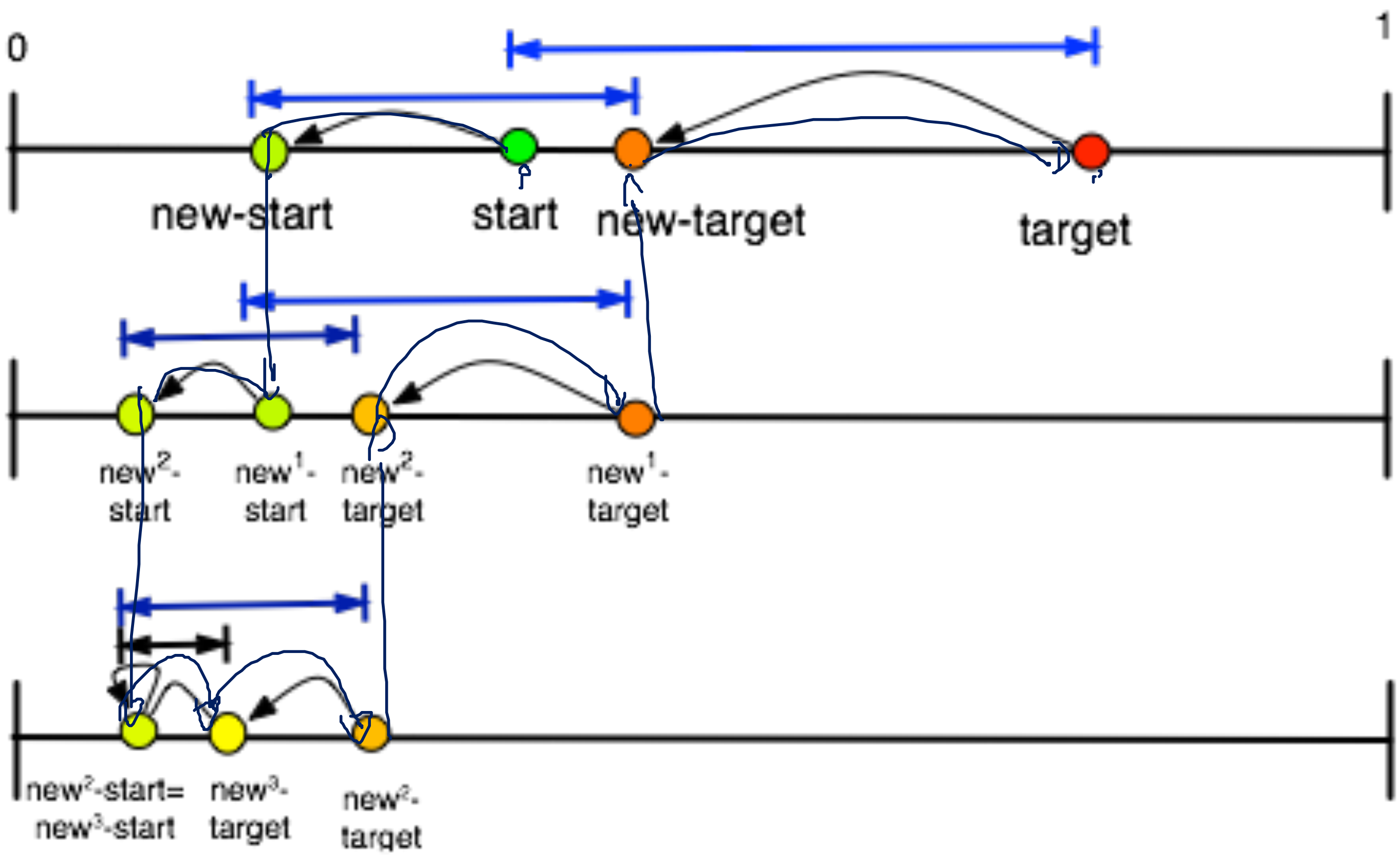
- Choose $2(1-\delta) \geq 1$

$$\delta \geq \frac{1}{2} \qquad t \leq \frac{1}{2}\delta^2$$

- Then, every interval is partitioned w.h.p.

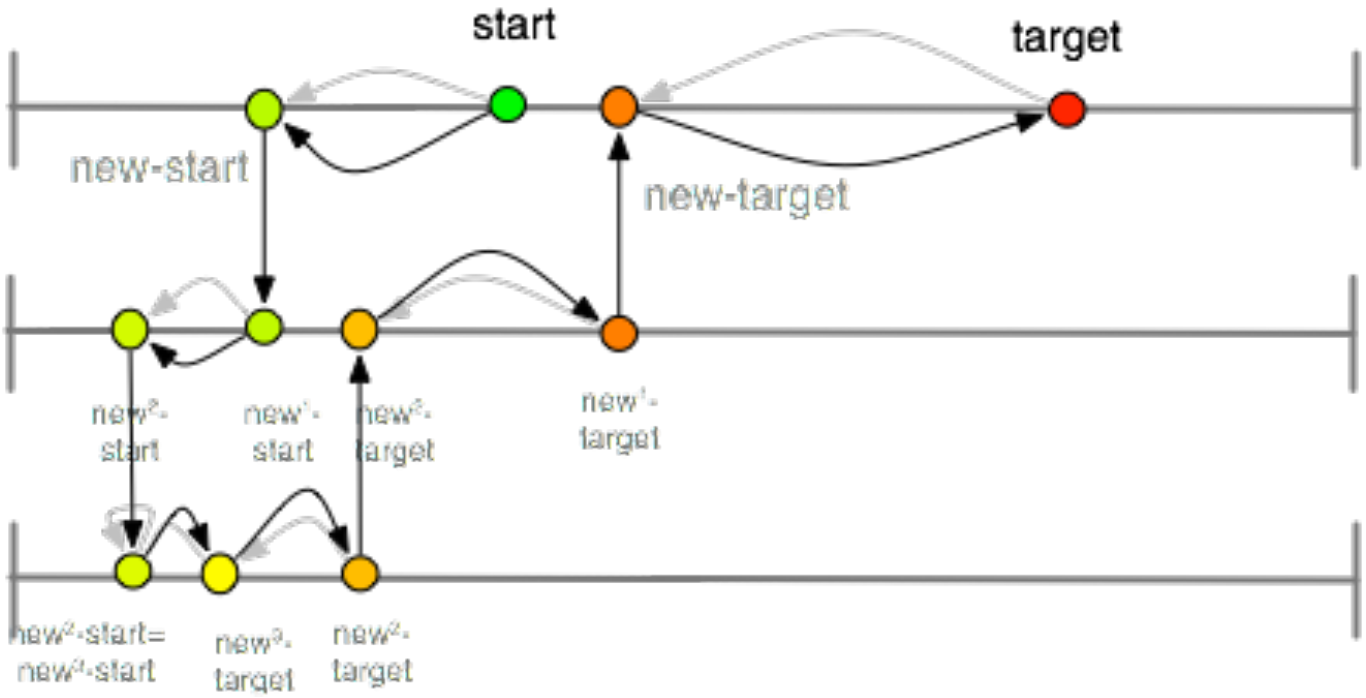
Lookup in Distance-Halving

- Map start/target to new-start/ target by using left edges
- Follow all left edges for 2+ log n steps
- Then, the new-new...-new-start and the new-new-...new-target are neighbored.



Lookup in Distance-Halving

- Follow all left edges for 2+ log n steps
- Use neighbor edge to go from new*-start to new*-target
- Then follow the reverse left edges from new^{m+1}-target to new^{m}-target

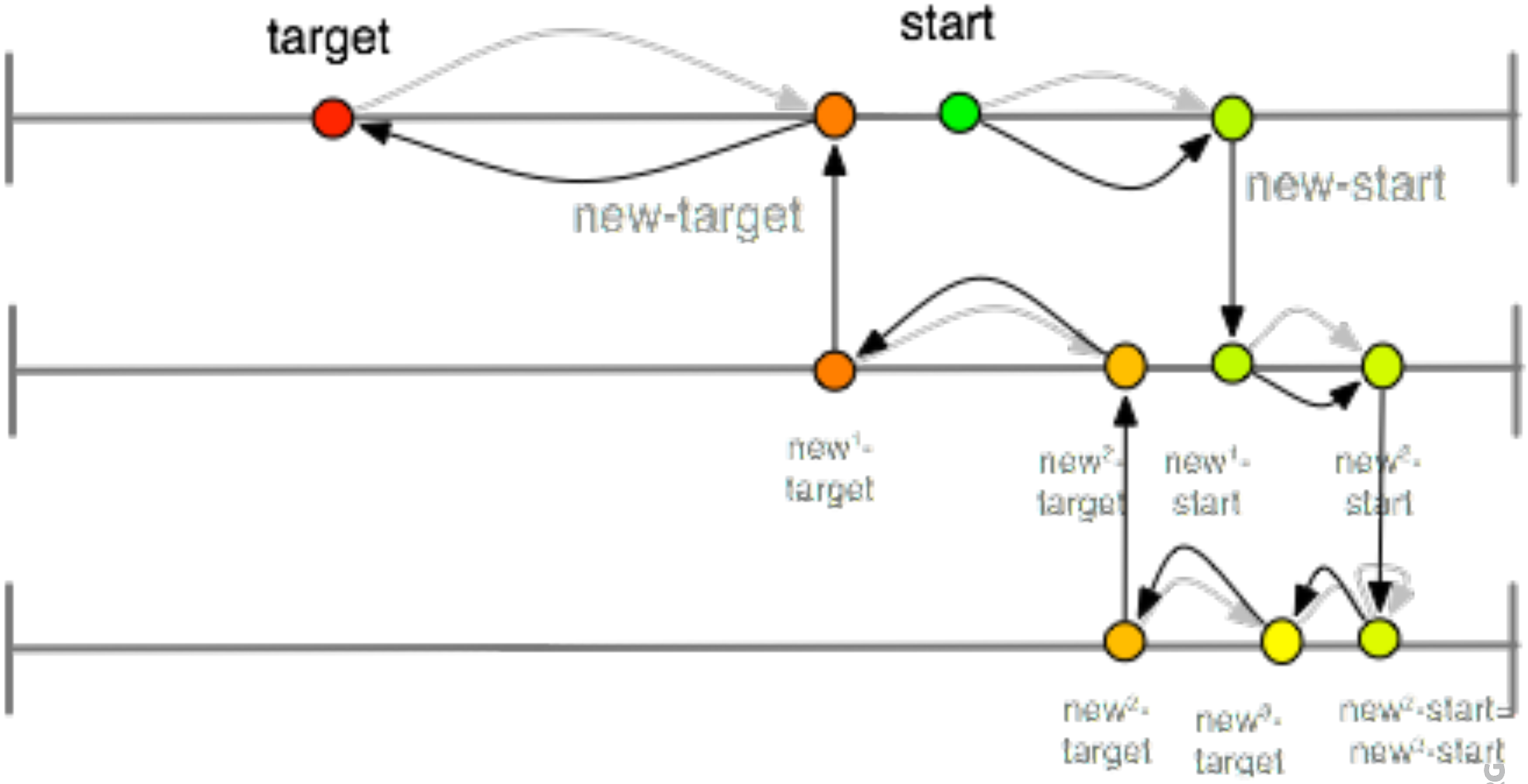


Structure of Distance-Halving

- Peers are mapped to the intervals
  - uses DHT for data
- Additional neighbored intervals are connected by pointers
- The largest interval has size 2/n w.h.p.
  - i.e. probability $1-n^{-c}$ for some constant c
- The smallest interval size 1/(2n) w.h.p.
- Then the indegree and outdegree is constant
- Diameter is O(log n)
  - which follows from the routing

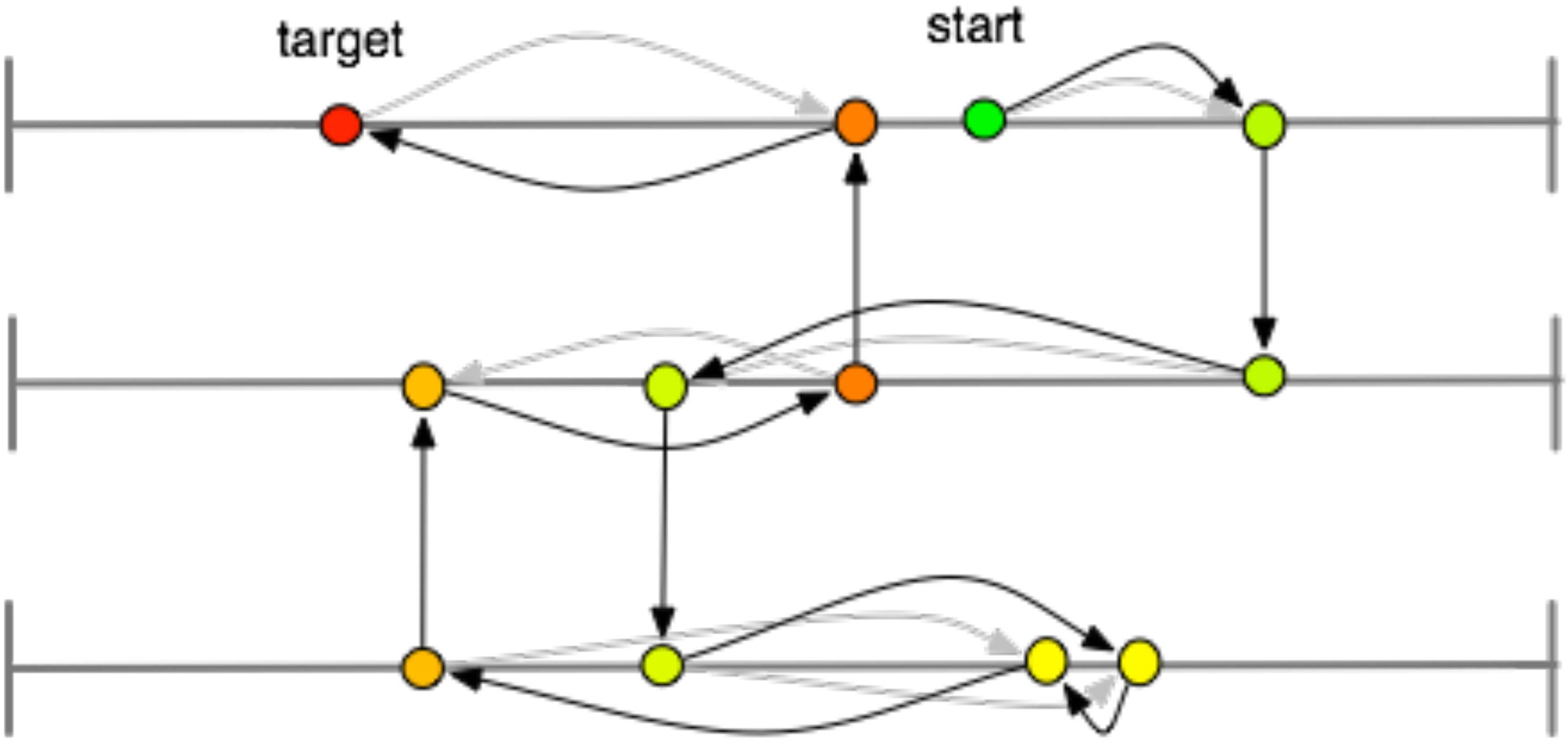
Lookup in Distance-Halving

- This works also using only right edges



Lookup in Distance-Halving

- This works also using a mixture of right and left edges



Congestion Avoidance during Lookup

- Left and right-edges can be used in any ordering
  - if one stores the combination for the reverse edges

- Analog to Valiant‘s routing result for the hyper-cube one can show

- The congestion ist at most O(log n),
  - i.e. every peer transports at most a factor of O(log n) more packets than any optimal network would need

- The same result holds for the ~~Viceroy~~ network

Inserting peers in Distance-Halving

1. Perform multiple choice principle
   - i.e. c log n queries for random intervals
   - Choose largest interval
   - halve this interval
2. Update ring edges
3. Update left and right edges
   - by using left and right edges of the neighbors

Lemma

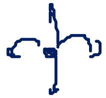
Inserting peers in Distance Halving needs at most $O(\log^2 n)$ time and messages.

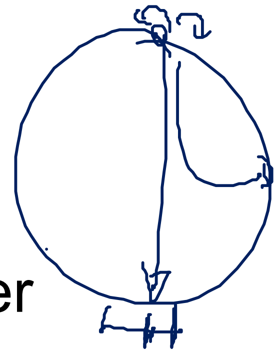
Summary Distance-Halving

≙ Hyper-Cube

- Simple and efficient peer-to-peer network
  - degree O(1)
  - diameter O(log n)
  - load balancing
  - traffic balancing
  - lookup complexity O(log n)
  - insert O($\log^2$n)

$(x, x+\varepsilon)$





- We already have seen continuous graphs in other approaches
  - Chord
  - CAN
  - Koorde
  - ViceRoy