Peer-to-Peer Networks

6. Analysis of DHT

Christian Schindelhauer
Technical Faculty
Computer-Networks and Telematics
University of Freiburg
Holes and Dense Areas
Theorem
- If \( n \) elements are randomly inserted into an array \([0,1]\) then with constant probability there is a „hole“ of size \( \Omega(\log n/n) \), i.e. an interval without elements.

Proof
- Consider an interval of size \( \log n / (4n) \)
- The chance not to hit such an interval is \( (1 - \log n/(4n)) \)
- The chance that \( n \) elements do not hit this interval is

\[
\left(1 - \frac{\log n}{4n}\right)^n = \left(1 - \frac{\log n}{4n}\right)^{\frac{4n}{\log n}} \frac{\log n}{4} \geq \left(1 - \frac{1}{4}\right) \frac{1}{4} \log n = \frac{1}{\sqrt{n}}
\]
- The expected number of such intervals is more than 1.
- Hence the probability for such an interval is at least constant.
Proof of Dense Areas

\[
\left(\frac{1}{4}\right)^{\frac{4}{4} \cdot \log n} = \left(\frac{1}{\log n}\right)^{\frac{4}{4}} \cdot \log \frac{1}{4^n} \\
= 2 \cdot \left(\frac{1}{2}\right) \cdot \log n \\
= 2 \cdot \frac{1}{2} \cdot \log n \\
= \frac{1}{\sqrt{n}} 
\]

Expectation:
\[
\frac{4\pi}{\log n} \cdot \frac{1}{\sqrt{n}} = \frac{4\sqrt{\pi}}{\log n} 
\]
Dense Spots

- **Theorem**
  - If \( n \) elements are randomly inserted into an array \([0,1]\) then with constant probability there is a dense interval of length \( 1/n \) with at least \( \Omega(\log n/ (\log \log n)) \) elements.

- **Proof**
  - The probability to place exactly \( i \) elements in to such an interval is
    \[
    \left( \frac{1}{n} \right)^i \left( 1 - \frac{1}{n} \right)^{n-i} \binom{n}{i}
    \]
  - for \( i = c \log n / (\log \log n) \) this probability is at least \( 1/n^k \) for an appropriately chosen \( c \) and \( k<1 \)
  - Then the expected number of intervals is at least \( 1 \)
Proof of Dense Areas

\[
i = \frac{c \cdot \log n}{\log \log n}
\]

Proof: If \(n \) balls are thrown into an interval of size \( \frac{1}{m} \), then

\[
P[i \text{ balls from } n \text{ balls fall into an interval of size } \frac{1}{m}] = \left(\frac{1}{m}\right)^i \left(1 - \frac{1}{m}\right)^{n-i} \left(\begin{array}{c} n \\ i \end{array}\right) \geq \frac{1}{m^i} \quad \text{for } i \leq \frac{n}{2}
\]
Proof of Dense Areas

\[ \frac{1}{q} \leq \left(1 - \frac{1}{m}\right)^m \leq \frac{1}{e} \]

\[ (1 - \frac{1}{m})^m = (1 - \frac{1}{en})^m \]

\[ \geq \left(\frac{1}{e}\right)^{1 - \frac{1}{n}} \]

\[ \geq \frac{1}{2} \]
Proof of Dense Areas

\[
\binom{m}{i} = \frac{m^i}{i! (m-i)!} = \frac{m(m-1)(m-2) \cdots (m-i)}{i!}
\]

\[
\geq \frac{\frac{m^i}{i!}}{\frac{i!}{i!}} = \frac{m^i}{i!} \geq \left(1 - \frac{1}{m}\right)^{m-i} \cdot \frac{m^i}{i!}
\]

\[
\left(1 - \frac{i-1}{m}\right)^{\frac{m^i}{i!}} \geq \left(\frac{1}{4}\right)^{\frac{m^i}{i!}} \left(1 - \frac{4}{m}\right)(i-1) \geq \left(\frac{1}{4}\right)^{\frac{1}{2}} = \left(\frac{1}{2}\right)^{\frac{1}{2}}
\]
Proof of Dense Areas

\[
\left( \frac{1}{2} \right) \cdot \frac{1}{\ln i} = 2 \\
1 - \frac{\ln i}{\ln n} \geq 2 \\
i + \ln i \leq \frac{c \cdot \log n}{\log \log n} \left( 1 + \ln c + \ln 2 \right) \log \log n \\
\leq \frac{c \cdot \log n}{\log \log n} \left( 1 + \ln c + \ln 2 \right) \log \log n \\
= c \left( 1 + \ln c + \ln 2 \right) \log n
\]
Averaging Effect

- **Theorem**

  - If $\Theta(n \log n)$ elements are randomly inserted into an array $[0,1]$ then with high probability in every interval of length $1/n$ there are $\Theta(\log n)$ elements.
Excursion

- Markov-Inequality
  - For random variable $X > 0$ with $\mathbb{E}[X] > 0$:
    \[
    P[X \geq k \cdot \mathbb{E}[X]] \leq \frac{1}{k}
    \]

- Chebyshev
  \[
  P[|X - \mathbb{E}[X]| \geq k] \leq \frac{\mathbb{V}[X]}{k^2}
  \]
  - for Variance
    \[
    \mathbb{V}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2
    \]

- Stronger bound: Chernoff
Theorem Chernoff Bound

- Let $x_1,...,x_n$ 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 $c > 0$
  $$P[S_n \geq (1 + c) \cdot E[S_n]] \leq e^{-\frac{1}{3} \min\{c, c^2\} pn}$$
- For $0 \leq c \leq 1$
  $$P[S_n \leq (1 - c) \cdot E[S_n]] \leq e^{-\frac{1}{2} c^2 pn}$$
Proof of 1st Chernoff Bound

- We show
  \[ P[S_n \geq (1 + c)\mathbb{E}[S_n]] \leq e^{-\frac{\min\{c, c^2\}}{3}pn} \]

- Für t>0:
  \[ P[S_n \geq (1 + c)pn] = P[e^{tS_n} \geq e^{t(1+c)pn}] \]

  \[ k = e^{t(1+c)pn}/\mathbb{E}[e^{t\cdot S_n}] \]

- Markov yields:
  \[ P \left[ e^{tS_n} \geq k\mathbb{E} \left[ e^{tS_n} \right] \right] \leq \frac{1}{k} \]

- To do: Choose t appropriately
Proof of 1st Chernoff Bound

We show
\[
\frac{1}{k'} \leq e^{-\frac{\min\{c,c^2\}}{3}pn}
\]

where
\[
k' = e^{t(1+c)pn} / \mathbb{E}[e^{t \cdot S_n}]
\]

Independence of random variables \(x_i\)

Next we show:
\[
e^{-t(1+c)pn} \cdot (1 + p(e^t - 1))^n \leq e^{-\frac{\min\{c,c^2\}}{3}pn}
\]
Proof of 1st Chernoff Bound

Show:
\[ e^{-t(1+c)pn} \cdot (1 + p(e^t - 1))^n \leq e^{-\min\{c, c^2\}pn} \]

where:
\[ t = \ln(1 + c) > 0 \]

\[
\begin{align*}
&\quad e^{-t(1+c)pn} \cdot (1 + p(e^t - 1))^n \\
&\leq e^{-t(1+c)pn} \cdot e^{pn(e^t-1)} \\
&= e^{-t(1+c)pn+pn(e^t-1)} \\
&= e^{-(1+c) \ln(1+c)pn+cpn} \\
&= e^{(c-(1+c) \ln(1+c))pn}
\end{align*}
\]

Next to show
\[ (1 + c) \ln(1 + c) \geq c + \frac{1}{3} \min\{c, c^2\} \]
Proof of 1st Chernoff Bound

To show for $c > 1$:

$$(1 + c) \ln(1 + c) \geq c + \frac{1}{3} c$$

For $c = 1$: $2 \ln(2) > 4/3$

Derivative:
- left side: $\ln(1+c)$
- right side: $4/3$

- For $c > 1$ the left side is larger than the right side since
  - $\ln(1+c) > \ln(2) > 4/3$

- Hence the inequality is true for $c > 0$. 
Proof of 1st Chernoff Bound

To show for $c < 1$:

$$(1 + c) \ln(1 + c) \geq c + \frac{1}{3}c^2$$

For $x > 0$:

$$\frac{d \ln(1 + x)}{dx} = \frac{1}{1 + x} = 1 - x + x^2 - x^3 + x^4 - \ldots$$

Hence

$$\ln(1 + x) = x - \frac{1}{2}x^2 + \frac{1}{3}x^3 - \frac{1}{4}x^4 + \ldots$$

By multiplication

$$(1 + x) \ln(1 + x) = x + \left(1 - \frac{1}{2}\right)x^2 - \left(\frac{1}{2} - \frac{1}{3}\right)x^3 + \left(\frac{1}{3} - \frac{1}{4}\right)x^4 - \ldots$$

Substitute $(1+c) \ln(1+c)$ which gives for $c \in (0,1)$:

$$(1 + c) \ln(1 + c) \geq c + \frac{1}{2}c^2 - \frac{1}{6}c^3 \geq c + \frac{1}{3}c^2$$
Theorem Chernoff Bound
- Let $x_1,...,x_n$ 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 $c>0$
  $$P[S_n \geq (1 + c) \cdot E[S_n]] \leq e^{-\frac{1}{3} \min\{c,c^2\}pn}$$
- For $0 \leq c \leq 1$
  $$P[S_n \leq (1 - c) \cdot E[S_n]] \leq e^{-\frac{1}{2} c^2 pn}$$
Proof of 2nd Chernoff Bound

- We show
  \[ P[S_n \leq (1 - c)\mathbb{E}[S_n]] \leq e^{-\frac{c^2}{2}pn}. \]

- For \( t < 0 \):
  \[ P[S_n \leq (1 - c)pn] = P[e^{tS_n} \geq e^{t(1-c)pn}] \]

\[ k = e^{t(1-c)pn} / \mathbb{E}[e^{tS_n}] \]

- Markov yields:
  \[ P \left[ e^{tS_n} \geq k \mathbb{E} \left[ e^{tS_n} \right] \right] \leq \frac{1}{k} \]

- To do: Choose \( t \) appropriately
Proof of 2nd Chernoff Bound

We show
\[
\frac{1}{k} \leq e^{-\frac{c^2}{2}pn}
\]

where
\[
k = e^{t(1-c)pn} / \mathbb{E}[e^{t \cdot S_n}]
\]
Independence of random variables $x_i$

Next we show:
\[
e^{-t(1-c)pn} \cdot (1 + p(e^t - 1))^n \leq e^{-\frac{c^2}{2}pn}
\]
Proof of 2nd Chernoff Bound

We show

$$e^{-t(1-c)pn} \cdot (1 + p(e^t - 1))^n \leq e^{-\frac{c^2}{2}pn}$$

where:

$$t = \ln(1 - c)$$

$$e^{-t(1-c)pn} \cdot (1 + p(e^t - 1))^n \leq e^{-t(1-c)pn} \cdot e^{pn(e^t-1)}$$

$$= e^{-t(1-c)pn + pn(e^t-1)}$$

$$= e^{-(1-c)\ln(1-c)pn - cpn}$$

Next to show

$$-c - (1 - c) \ln(1 - c) \leq -\frac{1}{2}c^2$$
Proof of 2nd Chernoff Bound

To prove:
\[-c - (1 - c) \ln(1 - c) \leq -\frac{1}{2} c^2\]

For $c=0$ we have equality

Derivative of left side: $\ln(1-c)$

Derivative of right side: $-c$

Now

\[\ln(1 + x) = x - \frac{1}{2} x^2 + \frac{1}{3} x^3 - \frac{1}{4} x^4 + \ldots\]

This implies

\[\ln(1 - c) = -c - \frac{1}{2} c^2 - \frac{1}{3} c^3 - \ldots < -c\]
Proof ctd.

\[-c - (A - c) \left( -c - \frac{A}{2} c^2 - \frac{A}{3} c^3 - \ldots \right) \]

\[-c + c + \frac{A}{2} c^2 + \frac{A}{3} c^3 + \frac{A}{4} c^4 + \frac{A}{5} c^5 + \ldots \]

\[= -\frac{A}{2} c^2 - \left( -\frac{A}{2} + \frac{1}{3} \right) c^3 - \left( -\frac{A}{3} + \frac{1}{4} \right) c^4 - \left( -\frac{A}{4} + \frac{1}{5} \right) c^5 - \ldots \]

\[-\frac{A}{2} c^2\]
Lemma

If \( m = k n \ln n \) Balls are randomly placed in \( n \) bins:

1. Then for all \( c > k \) the probability that more than \( c \ln n \) balls are in a bin is at most \( O(n^{-c'}) \) for a constant \( c' > 0 \).

2. Then for all \( c < k \) the probability that less than \( c \ln n \) balls are in a bin is at most \( O(n^{-c'}) \) for a constant \( c' > 0 \).

Proof:

Consider a bin and the Bernoulli experiment \( B(k n \ln n, 1/n) \) and expectation: \( \mu = m/n = k \ln n \)

1. Case: \( c > 2k \)

\[
P[X \geq c \ln n] = P[X \geq (1 + (c/k - 1))k \ln n] \\
\leq e^{-\frac{1}{3}(c/k - 1)k \ln n} \leq n^{-\frac{1}{3}(c-k)}
\]

2. Case: \( k < c < 2k \)

\[
P[X \geq c \ln n] = P[X \geq (1 + (c/k - 1))k \ln n] \\
\leq e^{-\frac{1}{3}(c/k - 1)^2k \ln n} \leq n^{-\frac{1}{3}(c-k)^2}.
\]

3. Case: \( c < k \)

\[
P[X \leq c \ln n] = P[X \leq (1 - (1 - c/k))k \ln n] \\
\leq e^{-\frac{1}{2}(1-c/k)^2k \ln n} \leq n^{-\frac{1}{2}(k-c)^2 / k}
\]
Peer-to-Peer Networks
6. Analysis of DHT

Christian Schindelhauer
Technical Faculty
Computer-Networks and Telematics
University of Freiburg