Peer-to-Peer Networks

6. Analysis of DHT

Christian Ortolf
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

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

- 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

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

\[ = 2^{\frac{4}{\sqrt{n}} \cdot \log n} \]

\[ = 2 \cdot \frac{4}{\sqrt{n}} \cdot \log n \]

\[ = n^{-\frac{4}{\sqrt{n}}} \]

\[ = \frac{4 \sqrt{n}}{\log n} \]

Expectation:
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:** Let \( n \) balls be placed uniformly at random in an interval of length \( \frac{1}{m} \).

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

\[
\frac{1}{\frac{\lambda}{\epsilon}} \leq (1 - \frac{\lambda}{m})^m \leq \frac{1}{\epsilon}
\]

\[
(1 - \frac{\lambda}{n})^n = (1 - \frac{1}{\frac{n}{\lambda}})^n \leq \frac{n - 1}{n} \leq \frac{1}{\frac{n}{\lambda}}
\]

\[
\geq \left( \frac{\frac{\lambda}{\epsilon}}{\epsilon} \right)^{\frac{1}{\frac{n}{\lambda}}} \geq \frac{1}{\frac{n}{\lambda}}
\]
Proof of Dense Areas

\[
\binom{n}{i} = \frac{n!}{i!(n-i)!} = \frac{n \cdot (n-1) \cdot (n-2) \cdots (n-i+1)}{i!} \\
\geq \left(1 - \frac{i-1}{n}\right)^{n-i} \cdot \frac{n}{i} \\
\left(1 - \frac{i-1}{n}\right)^{\frac{n}{i}} \cdot \frac{n}{i-i} \geq \left(\frac{4}{3}\right)^{n} \left(1 - \frac{i}{n}\right) \left(n-i\right) \\
\geq \left(\frac{4}{3}\right)^{\frac{n}{i}} \left(1 - \frac{i}{n}\right) \left(n-i\right) \\
\geq \left(\frac{4}{3}\right)^{\frac{n}{i}} \left(\frac{4}{3}\right)^{\frac{n}{2}} = \left(\frac{4}{3}\right)^{1} = \frac{4}{3}
\]
Proof of Dense Areas

\[
\left(\frac{1}{2}\right) \cdot \frac{1}{i+1} = 2
\]

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

\[
= c (1 + \log c + \log 2) \cdot \log n
\]
## 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 $E[X] > 0$:
    \[
    P[X \geq k \cdot E[X]] \leq \frac{1}{k}
    \]

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

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

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

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

- Markov yields:
  \[ \Pr[e^{tS_n} \geq k \cdot \mathbb{E}[e^{tS_n}]] \leq \frac{1}{k} \]

- 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

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

Independence of random variables \( x_i \)

Next we show:

\[
\begin{align*}
\mathbb{E}[e^{tS_n}] &= \mathbb{E} \left[ e^{t \sum_{i=1}^{n} x_i} \right] \\
&= \mathbb{E} \left[ \prod_{i=1}^{n} e^{tx_i} \right] \\
&= \prod_{i=1}^{n} \mathbb{E} \left[ e^{tx_i} \right] \\
&= \prod_{i=1}^{n} (e^{0}(1 - p) + e^{t}p) \\
&= (1 - p + e^{t}p)^n \\
&= (1 + (e^{t} - 1)p)^n
\end{align*}
\]
**Proof of 1st Chernoff Bound**

Show:

\[
e^{-t(1+c)pn} \cdot (1 + p(e^t - 1))^n \leq e^{-\frac{\min\{c, c^2\}}{3} pn}
\]

where:

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

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

\[= e^{-t(1+c)pn + pn(e^t-1)}
\]

\[= e^{-(1+c) \ln(1+c)pn + c pn}
\]

\[= e^{(c-(1+c) \ln(1+c))pn}
\]

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 $x_1, \ldots, x_n$ 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 $c > 0$
  \[
P[S_n \geq (1 + c) \cdot E[S_n]] \leq e^{-\frac{1}{3} \min\{c, c^2\}pn}
  \]

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

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

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

To do: Choose t appropriately
Proof of 2nd Chernoff Bound

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

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

Independence of random variables \(x_i\)

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

\[
\mathbb{E}[e^{tS_n}] = \mathbb{E} \left[ e^{t \sum_{i=1}^{n} x_i} \right]
\]
\[
= \mathbb{E} \left[ \prod_{i=1}^{n} e^{tx_i} \right]
\]
\[
= \prod_{i=1}^{n} \mathbb{E} \left[ e^{tx_i} \right]
\]
\[
= \prod_{i=1}^{n} \left( e^{0(1 - p) + e^t p} \right)
\]
\[
= (1 - p + e^t p)^n
\]
\[
= (1 + (e^t - 1)p)^n
\]
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:

\[ t = \ln(1 - c) \]

\[ e^{-t(1-c)pn} \cdot (1 + p(e^t - 1))^n \leq e^{-t(1-c)pn} \cdot e^{pn(e^t-1)} \]
\[ = e^{-t(1-c)pn + pn(e^t-1)} \]
\[ = e^{-(1-c)\ln(1-c)pn - cpn} \]

1 + x ≤ e^x

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.

\[-c - (A - c) \left( -c \frac{1}{2} c^2 - \frac{A}{3} c^3 - \cdots \right) \]

\[-c + c + \frac{1}{2} c^2 + \frac{A}{3} c^3 + \frac{1}{4} c^4 + \frac{1}{5} c^5 + \cdots \]

\[-c^2 - \frac{1}{2} c^3 - \frac{1}{3} c^4 - \frac{1}{4} c^5 - \cdots \]

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

\[\leq -\frac{1}{2} c^2\]
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 Ortolf
Technical Faculty
Computer-Networks and Telematics
University of Freiburg