Mixes on unstructured topologies

Shishir Nagaraja
Computer Lab
University of Cambridge
what does anonymity mean?

- Unlinkability: Hide the connection between the senders and the recipients.
- Untraceability: Hide the connection between actions of the same sender.
- Unobservability: Hide the fact that the user is talking.
Mix-network

User a
User b
User c
User 1
User 2
User 3
User n
Attacker
Mix-network topology and mix-route

User a

User b

User c

User 1

User 2

User 3

User n

Attacker
Desirable properties of the mix-network

- ‘High’ traffic analysis resistance.
- Size: The larger the network, greater the anonymity set or maximal anonymity.
- Traffic: More traffic means better anonymity.
- Robustness:
  - Liability management in anonymous communication
  - Clear incentives for carrying traffic under legal pressure
Evaluation framework

- Is the given topology any good?
  - Figure out the efficiency of the mixing process.
  - Analyze the traffic-analysis resistance of the mix-network.

- Modeling mix network operation
  - Markovian random walks

- What we are not interested in:
  - Side channel analysis
  - Variation in protocol behaviour across topologies
Measuring anonymity

- Number of bits the attacker is missing to uniquely link an actor to an action – (Serjantov and Danezis, PET 2002).

\[ A = \mathcal{E}[\alpha_i] = - \sum_i Pr[\alpha_i] \log_2 Pr[\alpha_i] \]
Evaluation recap

Under conditions of maximal anonymity:

- Minimum mix-route length required.
- Amount of traffic needed to prevent intersection attacks – traffic load patterns.
- Resistance to corrupt nodes.
Theory

1. Walk of length $t$

\[ \Delta(t) = \max_i \left| \frac{q_i^t - \pi_i}{\pi_i} \right| \]

Convergence

\[ \Delta(t) \leq n^{1.5} (\lambda_2)^t \]

Second eigenvalue of the transition matrix

\[ T_{ij} = \left( \frac{1}{k_i} \right) A_{ij} \]

Infinite length Walk

Adjacency matrix

\( k_i \) Degree of node i
2. \( q^t = q^{t-1}T \)

See the paper for a proof of why the second eigenvalue is a constant for varying network size (n).
Structured graph topologies

- Optimal mixing properties are obtained in expander graphs such as Ramanujan graphs.
- For N=5000 nodes, we have, $\lambda_2 \geq 0.5527$
- Hence, we can calculate mix-route length as approximately 4 hops.

(source: www.ams.org)
Unstructured network topologies

- Erdős-Rényi random graph topology

We chose $p$ such that the $|\text{biggest component}| \sim |V|=5000$ nodes

\[
P(k) = e^{-\lambda} \frac{\lambda^k}{k!}
\]

\[
\lambda = \binom{N-1}{k} P_{ER}^k (1 - P_{ER})^{N-1-k}
\]
Mixing efficiency of ER graph topology

Convergence of random walks on ER networks

Entropy

Route length

ER $<d> = 14$
Graph topologies continued...

- Scale-free topologies
- Power law
  - Heavy-tailed distribution
    - \( P(X > x) \sim x^{-a}, \ 0 < a < 2 \)
  - Zipf distribution / Zeta distribution
    - \( P(k) = Ck^{-(a+1)} \)
  - Pareto distribution
    - \( f(x) = ab^ax^{-(a+1)} \)
Mixing efficiency of SFR graph topology

SFR models the massive AT&T call graph

Convergence of random walks on scale-free random networks

<table>
<thead>
<tr>
<th>Network (N = 5000)</th>
<th>&lt;d&gt;</th>
<th>orD</th>
<th>t</th>
<th>A_{network}</th>
</tr>
</thead>
<tbody>
<tr>
<td>SFR</td>
<td>2</td>
<td>8</td>
<td>11.4383</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>7</td>
<td>11.5626</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>6</td>
<td>11.5958</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>6</td>
<td>11.6135</td>
<td></td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>5</td>
<td>11.6351</td>
<td></td>
</tr>
<tr>
<td>ER</td>
<td>14</td>
<td>7</td>
<td>12.2339</td>
<td></td>
</tr>
<tr>
<td>Expander</td>
<td>14</td>
<td>4</td>
<td>12.2877</td>
<td></td>
</tr>
</tbody>
</table>
Mixing efficiency of KWS topology (weak and strong ties)

$q=1, r=2$
Mixing efficiency of SF-BA graph topology

- **Growth**
  - Start with $m_0$ nodes, and then add a node with $m$ edges at every time step.
  - $m = m_0$
- **Preferential attachment**
  \[
  \prod (k_i) = \frac{k_i}{\sum_j k_j}
  \]
- It is a simple model but...
  - Fixed exponent = 3

Convergence of random walks on scale-free and ER networks
Mixing efficiency of LiveJournal topology

Convergence of random walks on LiveJournal [N=3800000]

<table>
<thead>
<tr>
<th>Network</th>
<th>m</th>
<th>t</th>
<th>$A_{network}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>SF</td>
<td>2</td>
<td>15</td>
<td>11.5852</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>10</td>
<td>11.6961</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>6</td>
<td>11.7293</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>6</td>
<td>11.7687</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>6</td>
<td>11.7953</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>6</td>
<td>11.8090</td>
</tr>
<tr>
<td>KWS</td>
<td>$q=2, p=1$</td>
<td>11</td>
<td>12.2945</td>
</tr>
<tr>
<td></td>
<td>$q=10, p=1$</td>
<td>5</td>
<td>12.2939</td>
</tr>
<tr>
<td></td>
<td>$q=2, p=1$</td>
<td>63</td>
<td>11.6440</td>
</tr>
<tr>
<td></td>
<td>$q=10, p=4$</td>
<td>63</td>
<td>11.6380</td>
</tr>
<tr>
<td>ER</td>
<td>$\langle d \rangle = 14$</td>
<td>7</td>
<td>12.2339</td>
</tr>
<tr>
<td>Expander</td>
<td>$D = 14$</td>
<td>4</td>
<td>12.2877</td>
</tr>
</tbody>
</table>
Corrupt nodes

- User a
- User b
- User c
- Attacker
- Mix2
- Mix10
- Mix3
- Mix13
- Mix9
- User 1
- User 2
- User 3
- User n

Network

Anonymity

PET 2007
Intersection attacks

- $\Pr[\text{any unused link}] \sim 0$
- Attacker: is the traffic from Alice proceeding along (i,j)?
  - Mean volume vs observed volume of traffic
  - $L$ (confidence parameter) – number of standard deviations from the mean.
  - $b$ – batch size, $p_i = 1/d_i$, $k = 2$ mixing rounds.

\[ k > 4l^2 \frac{p_i}{1-p_i} (b - 1) \]
<table>
<thead>
<tr>
<th>Network</th>
<th>$\langle d' \rangle$</th>
<th>$p_{\min}$</th>
<th>Batch size</th>
</tr>
</thead>
<tbody>
<tr>
<td>SFR</td>
<td>2</td>
<td>0.0344</td>
<td>10.08</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.0222</td>
<td>15.84</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.0243</td>
<td>14.4</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.0192</td>
<td>18.36</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>0.0135</td>
<td>26.28</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>0.0125</td>
<td>28.44</td>
</tr>
<tr>
<td>KWS</td>
<td>27 ($q = 1, r = 1$)</td>
<td>0.0294</td>
<td>11.88</td>
</tr>
<tr>
<td></td>
<td>43 ($q = 10, r = 1$)</td>
<td>0.0169</td>
<td>20.88</td>
</tr>
<tr>
<td></td>
<td>26 ($q = 1, r = 4$)</td>
<td>0.0333</td>
<td>10.44</td>
</tr>
<tr>
<td></td>
<td>28 ($q = 10, r = 4$)</td>
<td>0.0294</td>
<td>11.88</td>
</tr>
<tr>
<td>SF-linear</td>
<td>4</td>
<td>0.0048</td>
<td>74.16</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>0.0048</td>
<td>74.16</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>0.0041</td>
<td>86.04</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>0.0038</td>
<td>93.6</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>0.0037</td>
<td>96.12</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>0.0031</td>
<td>112.32</td>
</tr>
<tr>
<td>LJ</td>
<td>7.3221</td>
<td>0.00857</td>
<td>41.64</td>
</tr>
<tr>
<td>ER</td>
<td>14</td>
<td>0.0333</td>
<td>10.44</td>
</tr>
<tr>
<td>Expander</td>
<td>14</td>
<td>0.0714</td>
<td>4.68</td>
</tr>
</tbody>
</table>

Table 1: Batch sizes required to prevent intersection attacks
structured vs unstructured...

- Traffic analysis resistance – Comparable
- Maximal anonymity – Comparable
- Topological robustness in the face of litigation pressure – this depends on the social capital in the network.
  - Friends process each other’s traffic.
  - Processing “3rd party” traffic - indirect reciprocity - encourages a diverse user base which brings its benefits (Anonymity loves company, Dingledine and Mathewson 2006).
Conclusions

A successful mix-network design needs to consider the issues of liability management.

- Tapping social capital in a network to enhance topological robustness is an attractive proposal,

- And we have established the essential technical feasibility of this if this means using an unstructured mixnet topology

- Specifically:
  - Mix-route length is not a problem
  - Corrupt hubs are not a problem either
  - Batch sizes are however a challenge. Where do we bring 8 times the amount of dummy traffic from when compared to expanders? – will social chatter suffice?