An Analysis of Onion-Based Anonymous Routing in Delay Tolerant Networks

Kazuya Sakai, Tokyo Metropolitan University
Min-Te Sun, National Central University
Wei-Shinn Ku, Auburn University
Jie Wu, Temple University
Faisal S. Analazi, The Ohio State University
June 27-30, 2016
Outline

1. Introduction
2. Preliminary and Related Works
3. Abstract Onion-Based Anonymous Routing
4. Analyses
5. Simulations
6. Conclusions
1. Introduction

- Delay tolerant networks (DTNs)
  - Intermittently disconnected, **store-and-carry** forwarding
- The network model of a DTN
  - A graph representation is **contact-based**
  - The link weight between two nodes is defined by **contact frequency**, $\lambda_{i,j}$

\[
\begin{align*}
\lambda_{1,2} &= 1/80 \\
\lambda_{2,4} &= 1/100 \\
\lambda_{4,5} &= 1/120 \\
\lambda_{1,3} &= 1/60 \\
\lambda_{3,4} &= 1/70
\end{align*}
\]
Introduction (Cont.)

• Anonymous communications
  • Protect the privacy of end hosts
  • Prevent from traffic analyses
  • Intermediate nodes never know where a packet comes from and goes to

• Applications
  • Critical communications, e.g., battle fields
2. Preliminary and Related Works

- Onion Routing, e.g., Tor
  - Layered encryptions are applied to a message
Related Works

• Many anonymous routing protocols have been proposed for ad hoc networks
  • Onion-based, zone-based, and so on
• But, only a few protocols have been designed for DTNs
  • e.g., onion-based [1] and threshold-based
  • No protocol with multi-copy forwarding has been designed
  • No theoretical work has been done

3. Abstract Onion-Based Anonymous Routing

- Abstract onion-based anonymous routing protocol
  - Introduction of onion groups
    - A set of nodes consists of an onion group
    - Anycast-like forwarding
  - Single-copy and multi-copy forwarding
Multi-Copy Forwarding

• System Initialization:
  • The nodes in the system are divided into \( \left\lfloor \frac{n}{g} \right\rfloor \) groups, where \( g \) is the group size
  • Public/private keys initializations

• Input parameters:
  • The number of intermediate onion routers: \( K \)
  • The number of message copies: \( L \)
  • The message deadline: \( T \)

• Multi-Copy Forwarding \((v_s, v_d, m, K, L, T)\)
  • Selects a set of onion groups, \( R_1, R_2, ..., R_K \)
  • Generates an onion
  • Sets \( v_s.ticket = L \)
Multi-Copy Forwarding (Cont.)

• $v_i$ does the following at contact with $v_j$ at the $k$-th hop
  • $v_i$ and $v_j$ establishes a secure link
  • If $v_j \in R_k$ and Forward(.)
    • $v_i$ sends $m$ to $v_j$
    • $v_i$ decrements $v_i.ticket$ by 1
    • If $v_i.ticket = 0$, then $v_i$ deletes $m$ from the buffer
    • $v_j$ sets $v_j.ticket = 1$
  • if $v_d$ receives $m$, routing succeeds
  • If $m$ is not delivered to $v_d$ within $T$, routing fails

Note: the definition of Forward(.) is left to application designers e.g., intermediate onion relays are allow to have one copy
4. Analyses

- Analyses
  - Performance: delivery rate and message overhead
  - Security: traceable rate and path anonymity

- The attack model
  - The compromise attack: a node is physically compromised, and the transmission of a message is monitored
  - Compromised nodes are randomly chosen by the uniform distribution
Delivery Rate

• One-hop delivery
  - The inter-contact time between $v_i$ and $v_j$ is defined by $1/\lambda_{i,j}$
  - $\lambda_{i,j}$ is assumed to be exponentially distributed
  - $\Pr[v_i \text{ contacts with } v_j \text{ within } T] = \int_0^T \lambda_{i,j} e^{-\lambda_{i,j} t} \, dt = 1 - e^{-\lambda_{i,j} T}$

• Multi-hop delivery
  - An opportunistic path is modeled by the hypoexponential distribution [ICDCS’11]

• For the proposed protocol
  - We will define an opportunistic onion path
  - Which incorporates anycast forwarding

• The contact frequency for the $k$-th hop: $\lambda_k$

$$\lambda_k = \begin{cases} 
\sum_{j=1}^{g} \lambda_{s,r,k,1} & \text{for } k = 1 \\
\frac{1}{g} \sum_{i=1}^{g} \sum_{j=1}^{g} \lambda_{r_{k-1,i},r_{k,j}} & \text{for } 2 \leq k \leq K \\
\sum_{j=1}^{g} \lambda_{r_{k-1,j,d}} & \text{for } k = K + 1 
\end{cases}$$

• The delivery rate: $P_{\text{delivery}}(T)$

• The coefficient: $A_k^{(\eta)} = \prod_{j=1,j \neq k}^{g} \left( \frac{\lambda_j}{\lambda_j - \lambda_k} \right)$

$$P_{\text{delivery}}(T) = \sum_{k=1}^{\eta} A_k^{(\eta)} \left( 1 - e^{-\lambda_k T} \right)$$

$$P_{\text{delivery}}(T, L) = \sum_{k=1}^{\eta} A_k^{(\eta)} \left( 1 - e^{-\lambda_k L T} \right)$$
Message Forwarding Cost

- The spray-and-wait between two nodes
  - For a connected pair of $v_i$ and $v_d$, a message is eventually delivered, when $T \to \infty$
  - At most $2L - 1$ message transmissions
- Onion-based routing with $L = 1$
  - Num of msg tx is $K + 1$
- Onion-based routing with $L \geq 2$
  - The first hop: $2L - 1$
  - The $k$-th hop ($2 \leq k \leq K$): $KL$
  - Num of msg tx is $(K + 2)L - 1$

Fig 1. Source spray-and-wait
$2(L - 1) + 1 = 2L - 1$

Fig 2. Msg TX btwn 2 groups ($L \leq g$ holds)
The Traceable Rate

• The attack model
  
  • For path $v_1 \rightarrow v_2 \rightarrow \ldots \rightarrow v_k$, should $v_i$ be compromised, link $v_i \rightarrow v_{i+1}$ will be disclosed

• The traceable rate
  
  • Quantifies how much portion of a path is disclosed
  
  • $P_{trace} = \frac{1}{\eta^2} \sum_{i=1}^{c_{seg}} (c_{seg,i})^2$, where $\eta$ is the path length

\[ P_{trace} = \frac{2^2 + 1^2}{4^2} = \frac{5}{16} \]

\[ P_{trace} = \frac{3^2}{4^2} = \frac{9}{16} \]
Our approach

- The binary rep. of a path is defined by $b = \{b_1, b_2, ..., \}$
  - $b_i = 1$ indicates the compromised link
- The problem is reduced to estimating the run length of 1’s
- The run is the same consecutive 0’s or 1’s

$$v_1 \rightarrow v_2 \rightarrow v_3 \rightarrow v_4 \rightarrow v_5 \rightarrow v_6 \quad b = 01101$$

$$P_{trace} = \frac{2^2 + 1^2}{5^2}$$
The Traceable Rate (Cont.)

• This can be modeled by the geometric distribution
  • Let $X_i$ be the random variable that represents the run length of the first segment starting from $b_i$
  • $c$ : # of compromised nodes, $n$ : # of nodes
  • $E[X_i] = \sum_{k=[E[X_i]]+1}^{\eta} k^2 \left(\frac{c}{n}\right)^k \left(1 - \frac{c}{n}\right)$

• The traceable rate : $P_{trace}(c)$
  • We have $C_{seg} \leq [\eta/2]$
  • $P_{trace}(c) = \frac{1}{\eta^2} \sum_{i=1}^{[\eta/2]} E[X_i^2]$
Path Anonymity

• Anonymity
  • An entropy-based metric, \(-\sum_{i \in \phi} p_i \log(p_i)\)
  • Application-dependent
  • Example: a bit string 01XX
• Path anonymity for DTNs
  • Path anonymity : \(D(\phi') = \frac{H(\phi')}{H_{max}}\)
  • Note: the copies of a message can be correlated

\(\{0100, 0101, 0110, 0111\}\)

The \(k\)-th onion router in path 1 must be one of the nodes in \(R_k\)
Path Anonymity (Cont.)

• The maximal entropy of the system
  • There are \( \binom{n}{\eta} \) possible paths, when \( c = 0 \)
  • \( H_{\text{max}} = - \sum_{\forall \text{path} \in \phi} \frac{(n-\eta)!}{n!} \log \left( \frac{(n-\eta)!}{n!} \right) \)
  • \( \sum_{\forall \text{path} \in \phi} \frac{(n-\eta)!}{\eta!} \) equals to 1, since every path in an anonymous set is identified with the same probability
Path Anonymity (Cont.)

• The entropy of the system
  • The joint probability that a node is selected as an onion router and is compromised: \( \frac{1}{g} \cdot \frac{cg}{n} = \frac{c}{n} \)
  • Let \( Y \) be the random variable that represents the number of compromised nodes on a path
    • This follows the Binomial dist.: \( E[Y] = \sum_{i} i \left( \eta \right) \left( \frac{c}{n} \right)^{i} \left( 1 - \frac{c}{n} \right)^{\eta - i} \)
    • The probability that an adversary can guess the next node:
      • \( P_{\text{guess}}(v_{i}, n, g, k) = \begin{cases} \frac{1}{g} & \text{if } v_{i} \text{ is compromised} \\ \frac{1}{n-k} & \text{if otherwise} \end{cases} \)
Path Anonymity (Cont.)

• Let $c_0 = E[Y]$

• The probability of successfully guessing path $i$:
  \[
  p_i = \frac{(n-K+c_0)!}{n!} \cdot \frac{1}{g^{c_0}}
  \]

• The entropy of the system:
  \[
  H(\phi') = -\sum_{\forall \text{path} \in \phi'} \frac{(n-\eta+c_0)!}{g^{c_0}n!} \log \left( \frac{(n-\eta+c_0)!}{g^{c_0}n!} \right)
  \]

• Path anonymity $D(\phi')$
  \[
  D(\phi') = \frac{H(\phi')}{H_{\text{max}}} = \frac{\eta-c_0)(\ln(n)-1)+c_0 \ln(g)}{\eta(\ln(n)-1)}
  \]
5. Simulations

- Comparisons between analysis and simulation
  - Delivery rate, message overhead, traceable rate, and path anonymity

- Parameters
  - Group size $g$, Num of onion routers $K$, Num of copies $L$

- Two scenarios
  - Randomly generated graphs
  - Real traces with CRAWDAD dataset
Simulations with Random Graphs

- A contact graphs are randomly generated
- 1000 simulations experiments are conducted

Table. Simulation parameters.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value (default value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>The number of nodes, $n$</td>
<td>1000</td>
</tr>
<tr>
<td>The inter-contact time, $\lambda$</td>
<td>0 to 360 minutes</td>
</tr>
<tr>
<td>The group size, $g$</td>
<td>1 – 10 (5)</td>
</tr>
<tr>
<td>The number of onion routers, $K$</td>
<td>1 – 10 (3)</td>
</tr>
<tr>
<td>The number of copies, $L$</td>
<td>1 - 5</td>
</tr>
<tr>
<td>The message deadline, $T$</td>
<td>60 to 1800 minutes</td>
</tr>
<tr>
<td>The % of compromised nodes, $c/n$</td>
<td>0% - 50% (10%)</td>
</tr>
</tbody>
</table>
The delivery rate increases as the value of $L$ increases.

The # of msg forwarding increases, as the value of $L$ increases.
Traceable Rate and Anonymity

• Note: the traceable rate is independent from the value of $L$
• The traceable rate decreases, as the onion length increases
• The path anonymity decreases, as the value of $L$ increases
Results with Real Traces

• There exist on and off-business hours

Fig. the delivery rate
Cambridge traces (a small and dense network with 12 iMotes)

Fig. the delivery rate
Infocom’15 traces (a medium size network with 41 iMotes)
6. Conclusions

• We emphasize the theoretical aspects of onion-based anonymous routing in DTNs
  • An abstract of onion-based anonymous routing protocols
  • The analysis of the delivery rate, message overhead, traceable rate, and path anonymity
  • The analyses provide the closed-form solutions to approximate the simulation results
Thank you
Observations

• Three parameters, $g$, $K$, and $L$
  • Group size $g$: determined by the administrator
  • The onion length $K$: flexible setting is not possible
  • The number of copies $L$: tunable by each message transmission

• Rule of thumb
  • Larger $L$ improve delivery rate, but reduces path anonymity