# Topology design and routing strategy of ToR (The Onion Router) Network

Siyuan Li

N16908535

sl64662

## 1. Introduction

Tor(The Onion Router), which is widely used for privacy purpose, is that users send their traffic first to one of the onion routers, and through the Tor network to their demand's destination.

Beyond that, considering the purpose of using ToR, the optimal routing problem is no longer just find the best path for packets transmission, but also to guarantee the security of users' traffic.

Therefore, we need to come up with a new model to solve these problems. My project mainly focus on finding a proper way for users to joining in existing ToR network, and allocating their traffic to destination.

## 2. Model

In network topology, onion routers can be expressed as a set of nodes of internal routers, which means that there are no demands using these node as source or destination.

Also, we can see ToR users as areas to be connected by routers(of ToR networks), therefore we can use "node location problem" as the model to add users to ToR network.

We have to mention that the goal of using ToR is to transport data in a secure way, so we don't want traffic to be transmitted in a certain path or by a certain router. If an end user, say user A, is only connected to the ToR network by certain router, say router 1, then the traffic of user A will be more likely to be attacked. So end users can be linked to multiple ToR routers, since for ToR network, end users may want to use different paths(different onion routers) in different connections, which provides better security performance.

So, we can define the model of topology as:

indices

| | |
|---|---|
| i=1,2,...,N | number of ToR users to be connected to the network |
| j=1,2,...,M | number of ToR routers for users to be connected to |

constants

| | |
|---|---|
| $\xi_{ij}$ | cost of connecting user i to router j |
| $K_i$ | number of routers user i needed |

variables

| | |
|---|---|
| $u_{ij}$ | =1, if user i is connected to router j; 0, otherwise |

objective

| | |
|---|---|
| minimize | $F = \sum_i \sum_j \xi_{ij} u_{ij}$ |

constraints

| | |
|---|---|
| $\sum_j u_{ij} = K_i$ | ,i=1,2,...,N |

Given the topology of the ToR network, we can use the above model to decided which router should an user to be connected to, with the optimal cost.

After adding all users to the network, we should begin to consider the next step, what we should do to make users' demand transmitted with both good efficiency and high security level.

Based on the theory of ToR network, every packet is encrypted using keys of every ToR router on it's path, this means that one of these links being successfully attacked won't lead to the total failure of the path(content of the packet is still safe). Only all links along that path are attacked will the content be read by attackers.

We can use the following model:(we call it security model 1)

indices

| | |
|---|---|
| d=1,2,...,D | demands between all ToR users |
| p=1,2,..., $P_d$ | candidate paths for demand d |
| e=1,2,...,E | links in the whole network |

constants

| | |
|---|---|
| $h_d$ | volume of demand d |
| $k_d$ | diversity request for demand d |
| $l_d$ | "length" request for demand d{paths used have to go through $l_d$ links) |
| $q_e$ | probability of link e is under attack |
| $\delta_{edp}$ | =1, if link e is on path p realizing demand d; 0, otherwise |
| $\xi_e$ | unit cost of link e |

variables

| | |
|---|---|
| $x_{dp}$ | flow of demand d on path p |
| $y_e$ | capacity of link e |
| $u_{dp}$ | binary variable corresponding to traffic on path p of demand d |

objective

minimize $F = \sum_e \xi_e y_e$

constraints

$\sum_p x_{dp} = h_d$ ,d=1,2,...,D

$\sum_d \sum_p \delta_{edp} x_{dp} \le y_e$ ,e=1,2,...,E

$x_{dp} \le u_{dp} h_d / k_d$ ,d=1,2,...,D; p=1,2,..., $P_d$

$\sum_e \delta_{edp} \ge l_d u_{dp}$ ,d=1,2,...,D; p=1,2,..., $P_d$

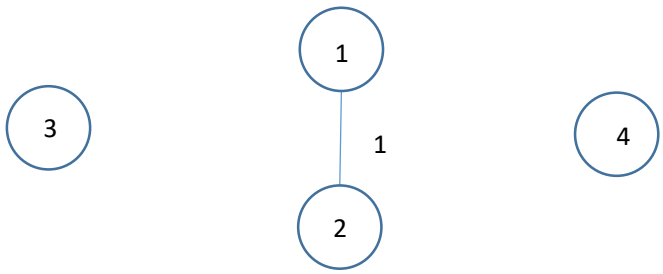
In the above model, we use a diversity factor to set an upper bound on each link to make sure that there will not be a lot of demand d on a certain link. Beyond that, we introduce a new "length" constraint that make all candidate paths realizing demand d go through more than $l_d$

links. These two measures help us make user's demand safer.

3. Simulation

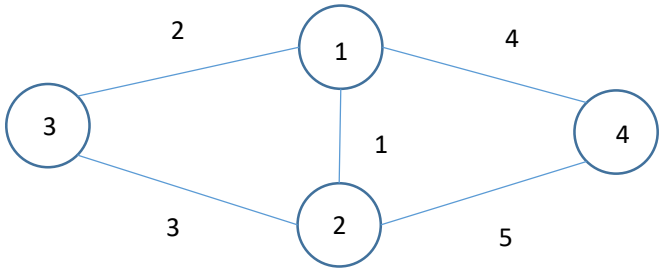
Let's first start with the simplest topology, a ToR network with just two routers: router1 and router2, represented by node 1,2, connected by link 1.

We also need to add some users for ToR network, here we have user1 and user2, represented by node 3,4.




Set $K_i = 2$, for i=1,2.(see Q1.mod)

By using model of topology, we can add node 3 and 4 to the network, connected by link 2,3,4,5.




Next we set a demand d: node3 to node 4, h=4;

There are 4 candidate paths: 3-1-4,(link 2 and 4)

3-2-4,(link 3 and 5)

3-1-2-4,(link 2, 1 and 5)

3-2-1-4,(ink 3, 1 and 4)

With unit link cost: link1: 1

link2: 2

link3: 2

link4: 2

link5: 2

Diversity request k=2; length request l=3; $q_e = 0.1$ for all links

Then we can set data file properly to do the simulation.

Use the routing model to see how traffic is allocated on this network.

(see S1.mod and S1.dat)

```
ampl: display F;
F = 20

ampl: display x;
x :=
1 1   0
1 2   0
1 3   2
1 4   2
;
```

```
ampl: display q;
q [*] :=
1  0.1
2  0.1
3  0.1
4  0.1
5  0.1
;
```

From the running result, traffic of demand 1 is allocated on path 3 and path 4, where go through 3 links on. And the corresponding network cost is 20. Meanwhile, we can calculate the probability of user's demand being successfully attacked, using the above model, is 0.1%.

Next we will make a comparison between this new model and the shortest path model in the security aspect.

For the new model, as mentioned, in this special case, the network cost is 20, and the probability of being attacked is 0.1%.

For the shortest path model, the network cost will be 16, and the probability of being attacked will be 1%.

This means that by using the new model instead of the shortest path model, we get better performance in keeping the path safety, but we have to sacrifice the total cost of the transmission.

Generally speaking, the new model tries to keep it's traffic safe by setting an extra constraint: all paths carrying traffic have to be longer than certain length l. Beyond that, we ask demand's volume to be equally split among k paths satisfying previous constraint(k paths don't have to be of same length).

4. Improvement

In the previous example, we use security model 1 to solve a very basic ToR network problem. In the result we see that demand 1 is realized by taking path 3 and 4, which both use link 1, then if link 1 has a high probability of being attacked, this strategy will result in a bad performance of protecting data.

Therefore we make an improvement to security model 1. Since now we are mainly focusing on the security issue, and in security model 1, we have seen that better security performance will lead to the increase of routing cost, so now we just focus on the security. Instead of using routing cost as the objective function, we will use the sum of probability of links' being attacked weighted by the load of the link.

Then we get a new model:(call it security model 2)

indices

| | |
|---|---|
| d=1,2,...,D | demands between all ToR users |
| p=1,2,..., $P_d$ | candidate paths for demand d |
| e=1,2,...,E | links in the whole network |

constants

| | |
|---|---|
| $h_d$ | volume of demand d |
| $k_d$ | diversity request for demand d |
| $l_d$ | "length" request for demand d{paths used have to go through $l_d$ links) |
| $q_e$ | probability of link e is under attack |
| $\delta_{edp}$ | =1, if link e is on path p realizing demand d; 0, otherwise |
| $\xi_e$ | unit cost of link e |

variables

$x_{dp}$ flow of demand d on path p

$y_e$ capacity of link e

$u_{dp}$ binary variable corresponding to traffic on path p of demand d

objective

minimize $F = \sum_e q_e y_e$

constraints

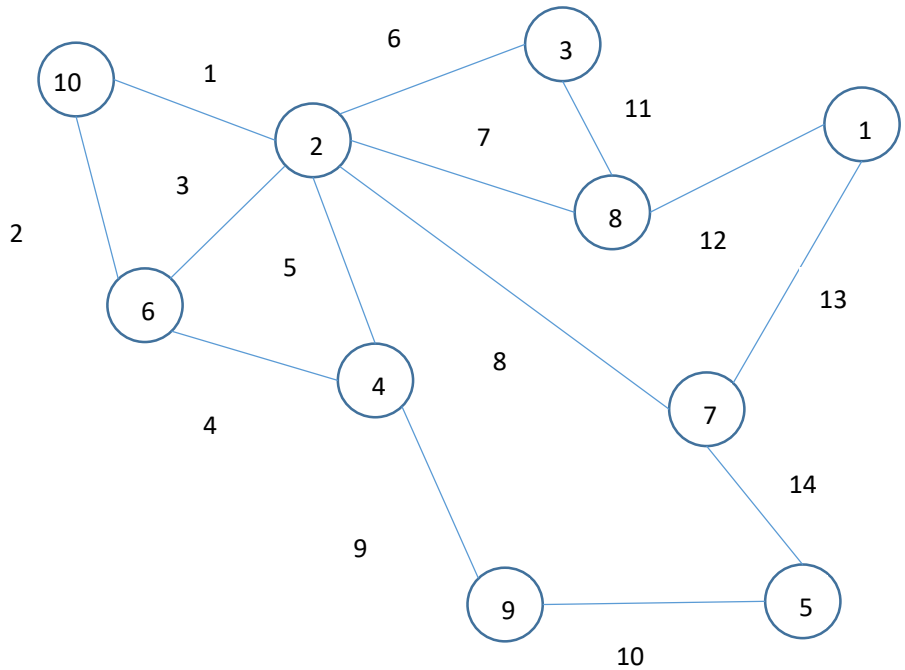
$\sum_p x_{dp} = h_d$ ,d=1,2,...,D

$\sum_d \sum_p \delta_{edp} x_{dp} \le y_e$ ,e=1,2,...,E

$x_{dp} \le u_{dp} h_d / k_d$ ,d=1,2,...,D; p=1,2,..., $P_d$

$\sum_e \delta_{edp} \ge l_d u_{dp}$ ,d=1,2,...,D; p=1,2,..., $P_d$

Then we will use this model to solve problems in a more complex network.



With the given topology, we can simply define all nodes and links, then we need to set all needed parameter to do the simulation.

Demands:

1: node 1 to node 6 with volume 12;

The shortest path is: 12-7-3

2: node 3 to node 7 with volume 10;

The shortest path is: 6-8

3: node 4 to node 8 with volume 16;

The shortest path is: 5-7

4: node 2 to node 9 with volume 18;

The shortest path is: 5-9

Diversity request:

k=4

Length request:

l=4

(k and l can be set differently for different demands)

Probability of link e is under attack:

q=0.1 for e=1,2,3,4;

q=0.2 for e=5,6,7,8,9;

q=0.3 for e=10,11,12,13,14;

Unit cost of link e:

```
param cost :=
 1 3
 2 6
 3 5
 4 1
 5 4
 6 2
 7 3
 8 5
 9 6
 10 2
 11 4
 12 5
 13 6
 14 5;
```

$\delta_{edp}$ (see in S2.dat)

The running result(of security model 2) is showed in the graph below.

```
ampl: display x;
x [*,*] (tr)
:    1    2    3    4     :=
1    0   0     4   4.5
2    3   2.5   4   0
3    3   0     4   0
4    0   2.5   0   4.5
5    3   2.5   0   4.5
6    0   2.5   0   4.5
7    0   0     0    .
8    3   0     0    .
9    0   0     4    .
10   0   0     .    .
;
```

The running result of security model 1 with the same data is:

```
ampl: display x;
x [*,*] (tr)
:    1    2    3    4     :=
1    0    0    4    4.5
2    3    2.5  4    0
3    0    0    4    0
4    0    2.5  0    4.5
5    3    2.5  0    4.5
6    3    2.5  0    4.5
7    0    0    0    .
8    3    0    0    .
9    0    0    4    .
10   0    0    .    .
;

ampl:
```

5. Comparison

So far we have use three different models to solve the ToR network routing problem, they are shortest path routing model, security model 1, security model 2. Now we will analyse the difference between them.

We will use the expectation of probability that traffic is attacked to evaluate these three model, let's call it $r$.

$$r = \frac{1}{D}\sum_d r_d$$

$$r_d = \sum_p \frac{1}{k_d}\prod_e u_{dp}\delta_{edp}q_e$$

For shortest path routing, $r_d$ is a constant if the network topology and configuration is given. In the previous topology, we can have:

$$r_1 = 6\times10^{-3}$$

$$r_2 = 4\times10^{-2}$$

$$r_3 = 4\times10^{-2}$$

$$r_4 = 4\times10^{-2}$$

$$r = 3.15 \times 10^{-2}$$

For security model 1, we have:

$$r_1 = 1.53 \times 10^{-3}$$

$$r_2 = 2.004 \times 10^{-3}$$

$$r_3 = 3.95 \times 10^{-4}$$

$$r_4 = 3.263 \times 10^{-4}$$

$$r = 1.063825 \times 10^{-3}$$

For security model 2, we have:

$$r_1 = 1.5 \times 10^{-3}$$

$$r_2 = 2.004 \times 10^{-3}$$

$$r_3 = 3.95 \times 10^{-4}$$

$$r_4 = 3.263 \times 10^{-4}$$

$$r = 1.05625 \times 10^{-4}$$

Similarly we can calculate r for different k value(just simply change the param k in the S2.dat).

For k =3, for security model 1: $r = 1.148 \times 10^{-4}$

for security model 2: $r = 1.044 \times 10^{-4}$

For k =2, for security model 1: $r = 1.4 \times 10^{-3}$

for security model 2: $r = 1.366 \times 10^{-3}$

(r for shortest path model is too large to be shown in this figure)


From the comparison result we can get that:

Using security model 1 and 2 is much safer than simply using shortest path routing in the ToR network;

By increasing the path diversity factor k, we can improve the security performance;

Using security model 2 is a little bit better than model 1in the probability of traffic being successfully attacked.