# Anonymity on Quicksand

## Using BGP to compromise Tor


Laurent Vanbever

ETH Zürich


Columbia University

May, 8 2015

Joint work with


Yixin Sun, Annie Edmundson, Oscar Li, Jennifer Rexford, Mung Chiang, Prateek Mittal

WE ARE
ANONYMOUS

WE ARE NOT
ANONYMOUS FOR LONG.

WE ARE NOT
ANONYMOUS FOR LONG.
COURTESY OF BGP.

# Internet communications are *not* anonymous

Looking at an Internet communication, one can

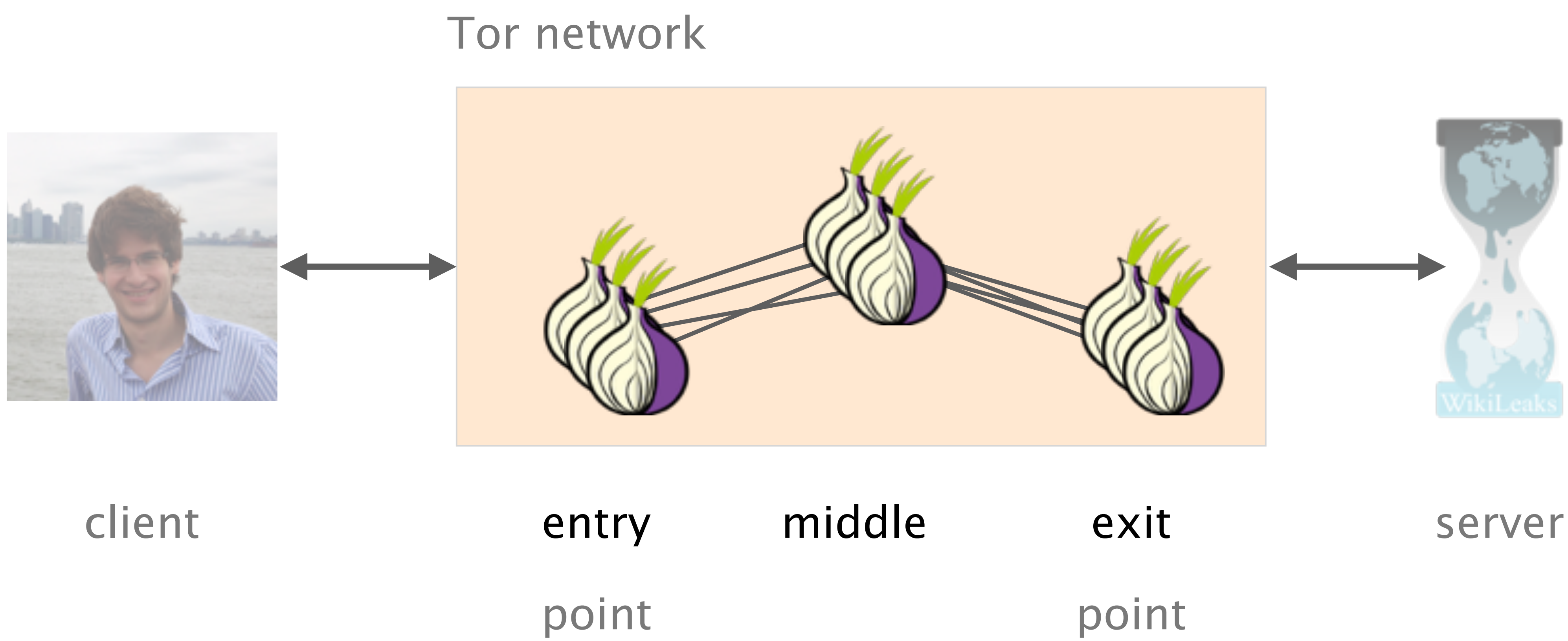
- infer who is talking to whom
- infer physical locations
- use that to track behavior and interests

even if the communication is encrypted

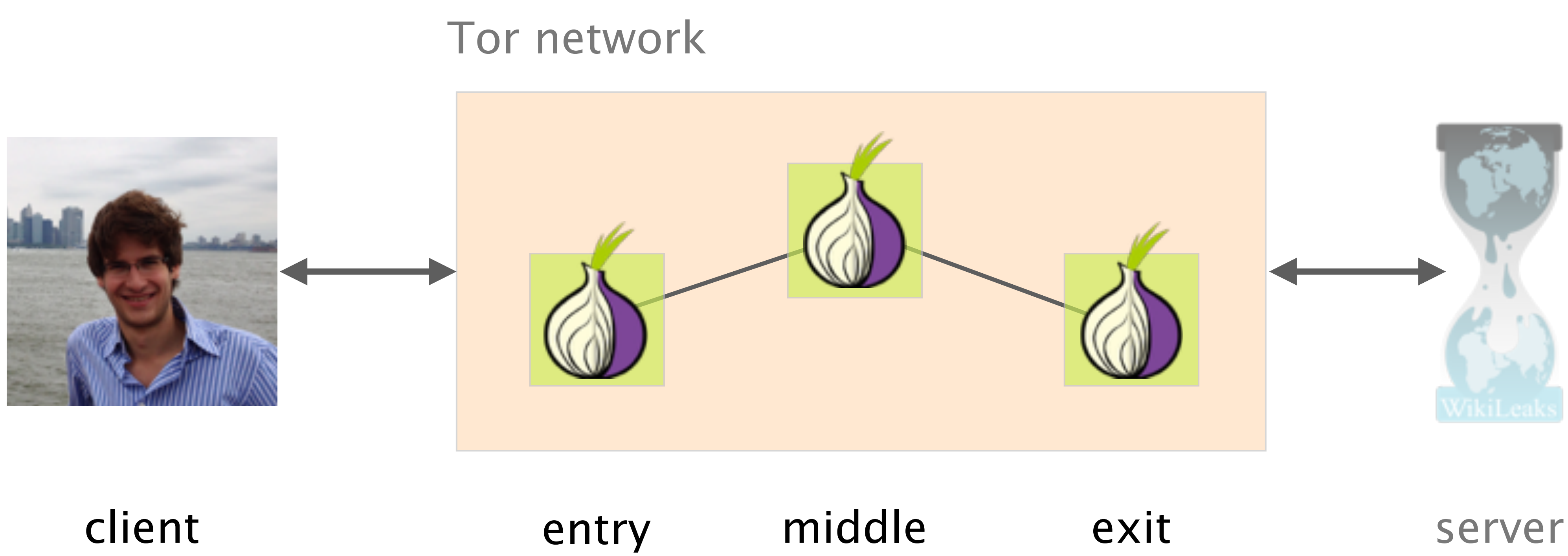
# Tor aims at preventing adversaries to follow packets between a sender and a receiver

To do that,

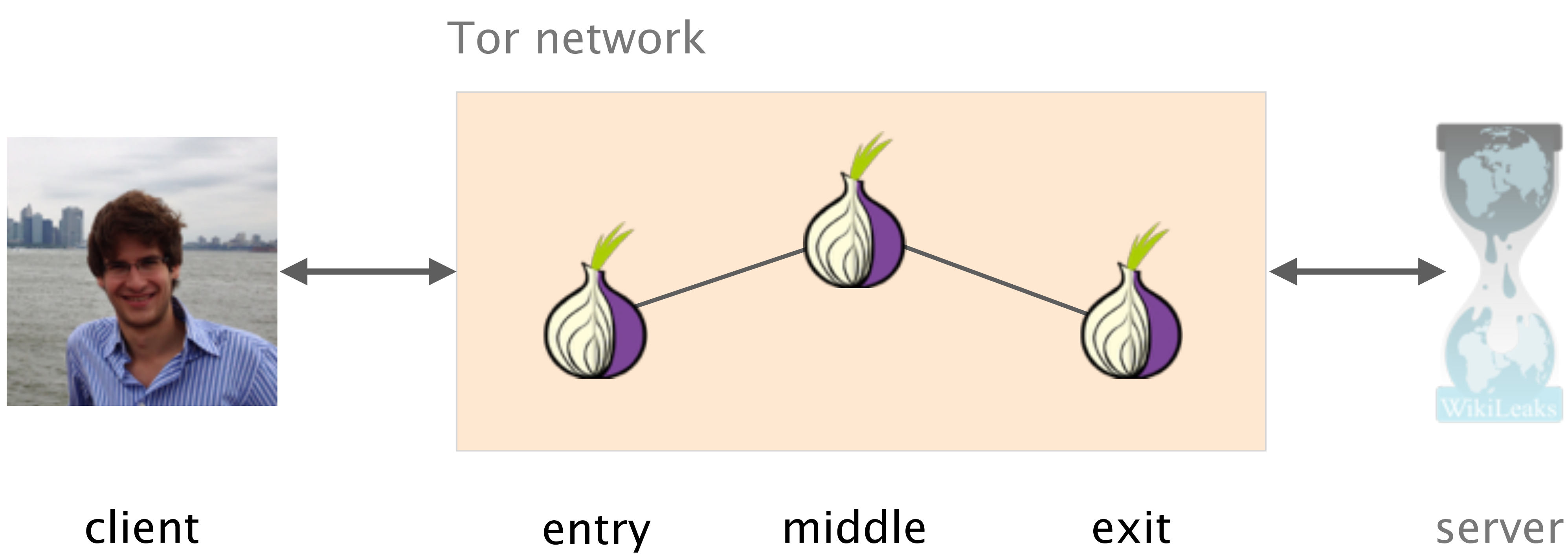
Tor bounces traffic around a network of relays

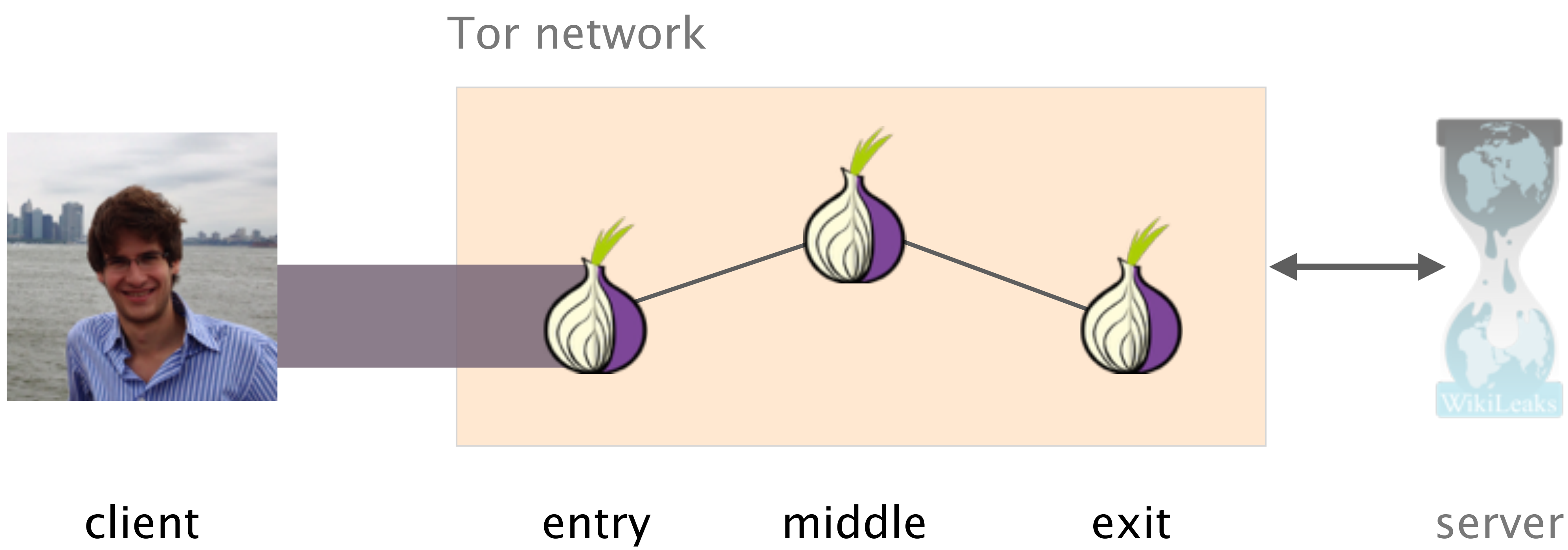


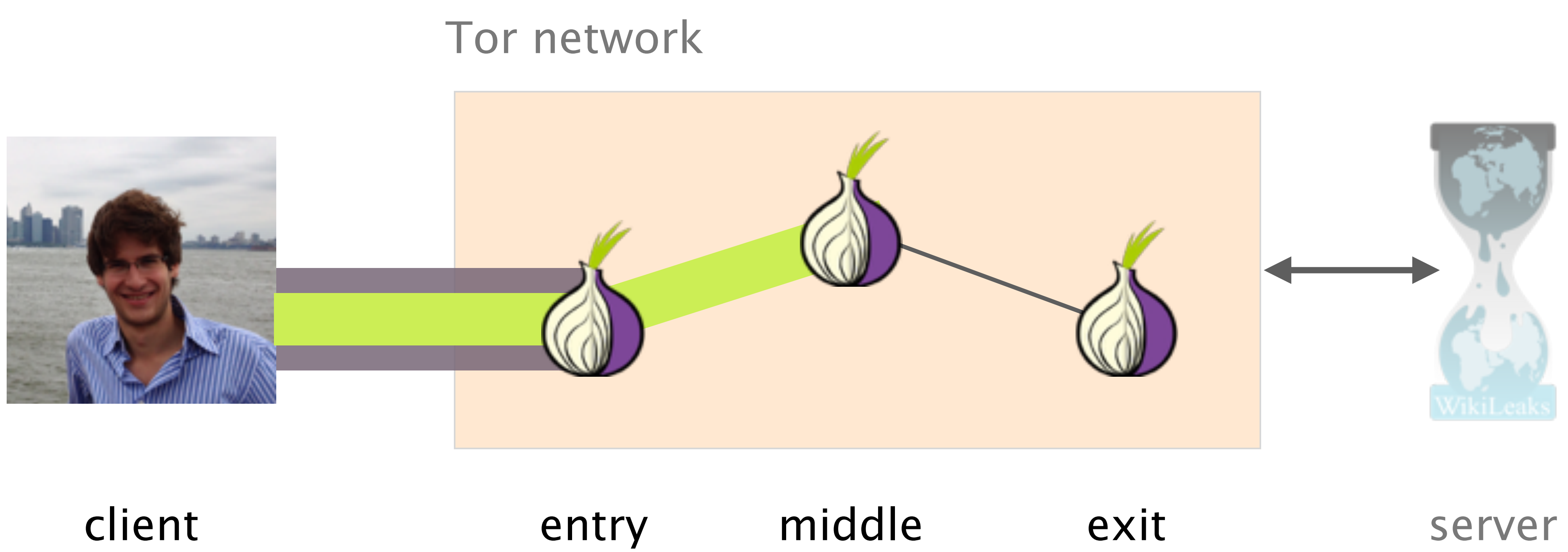
# Tor clients start by selecting 3 relays, one of each type



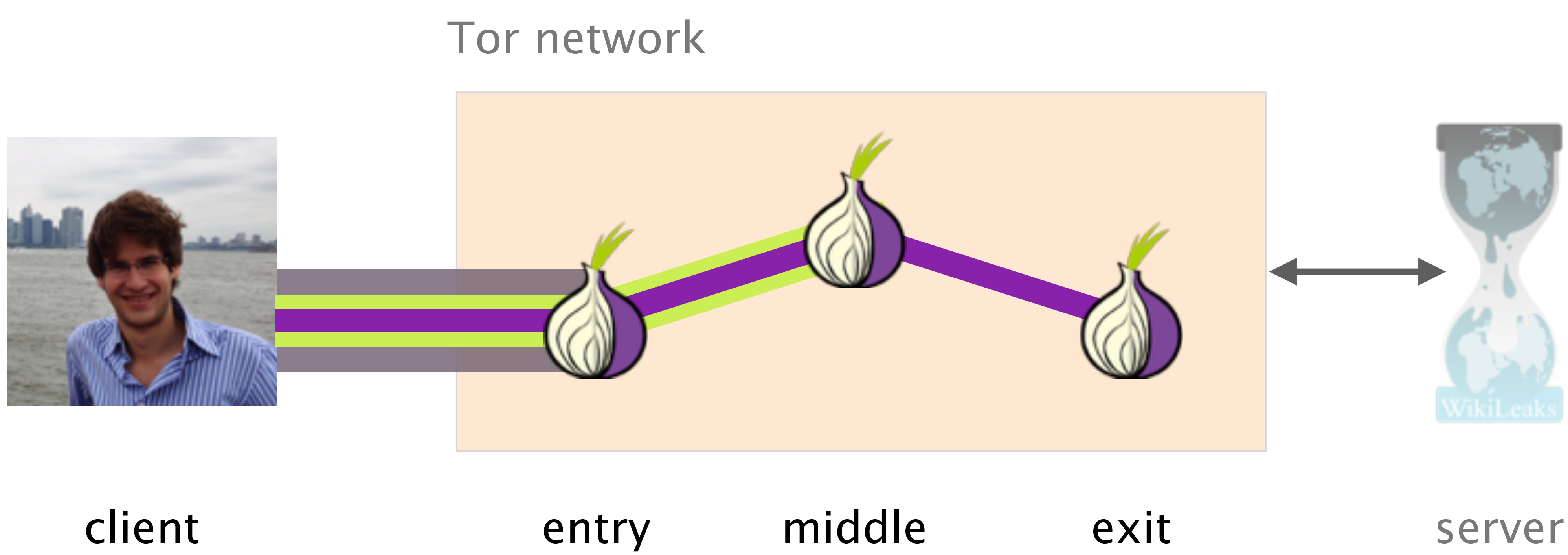
# Tor clients then incrementally build encrypted circuits through them



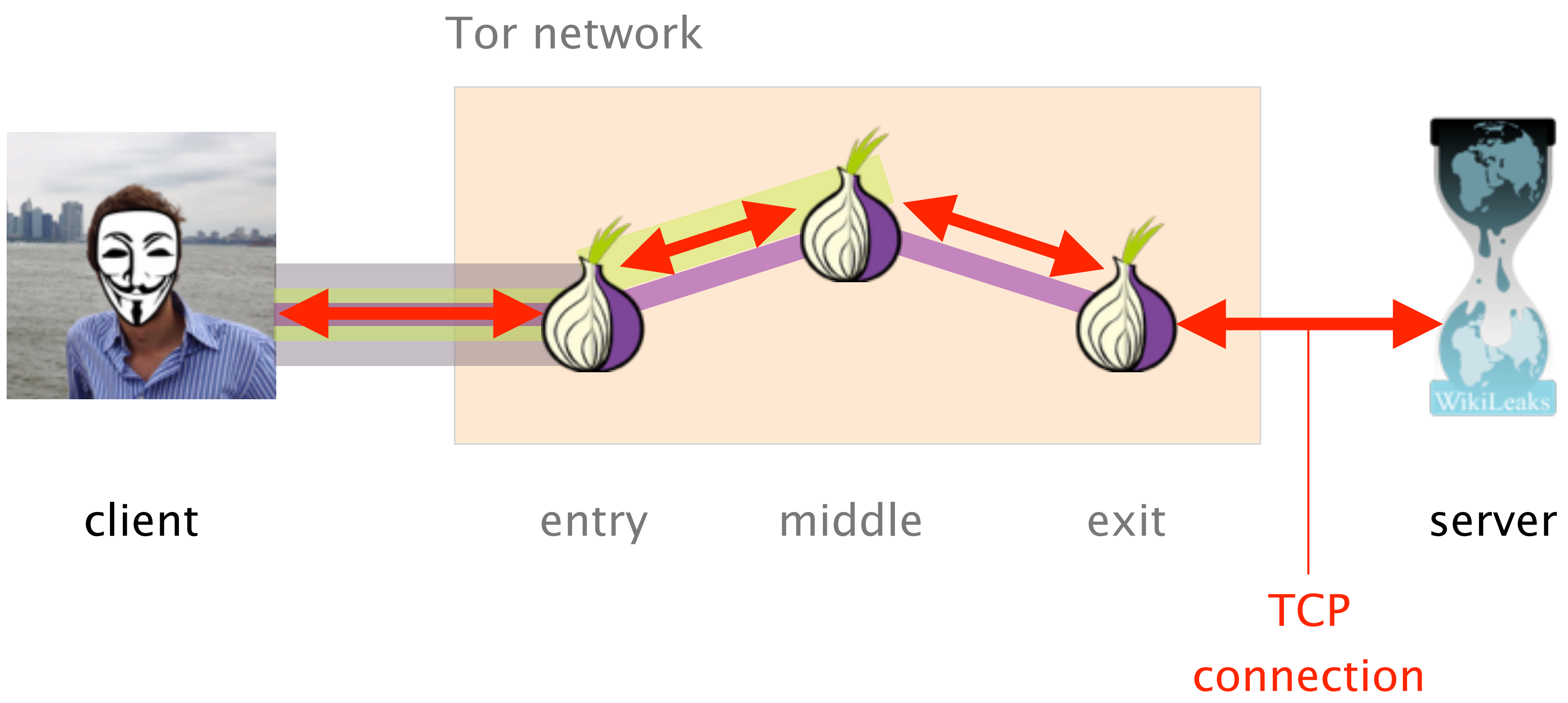

Tor network
WikiLeaks
client
entry
middle
exit
server



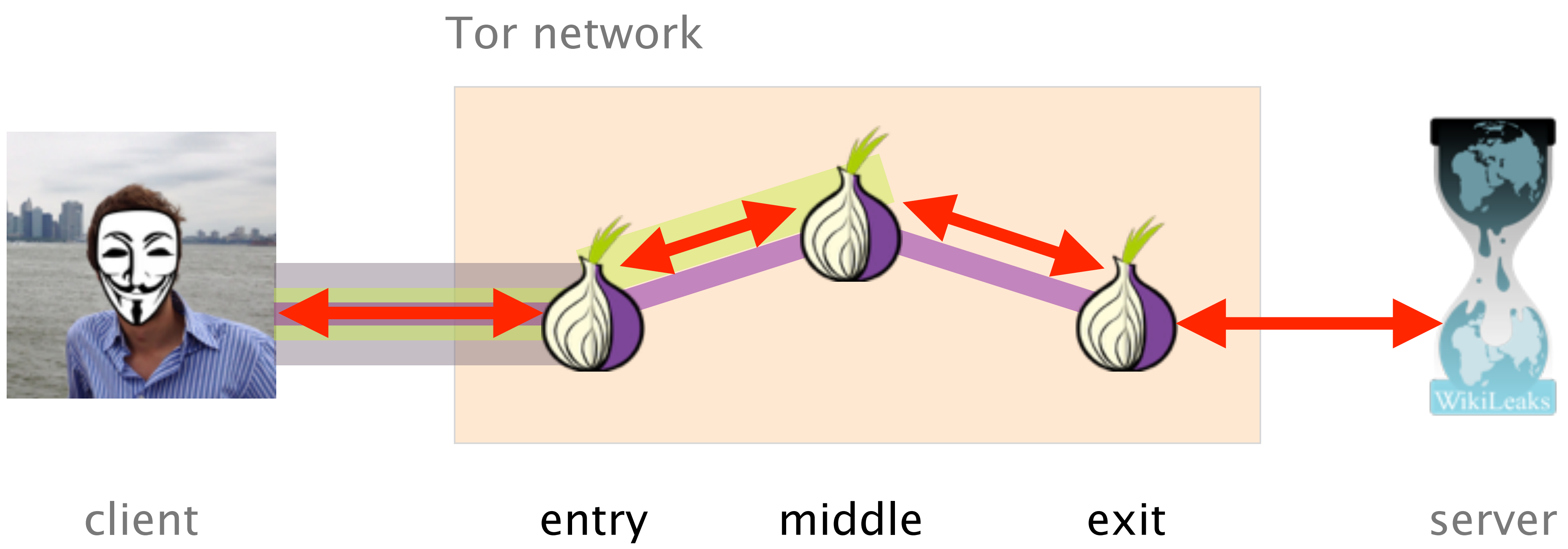
Tor network
WikiLeaks
client
entry
middle
exit
server

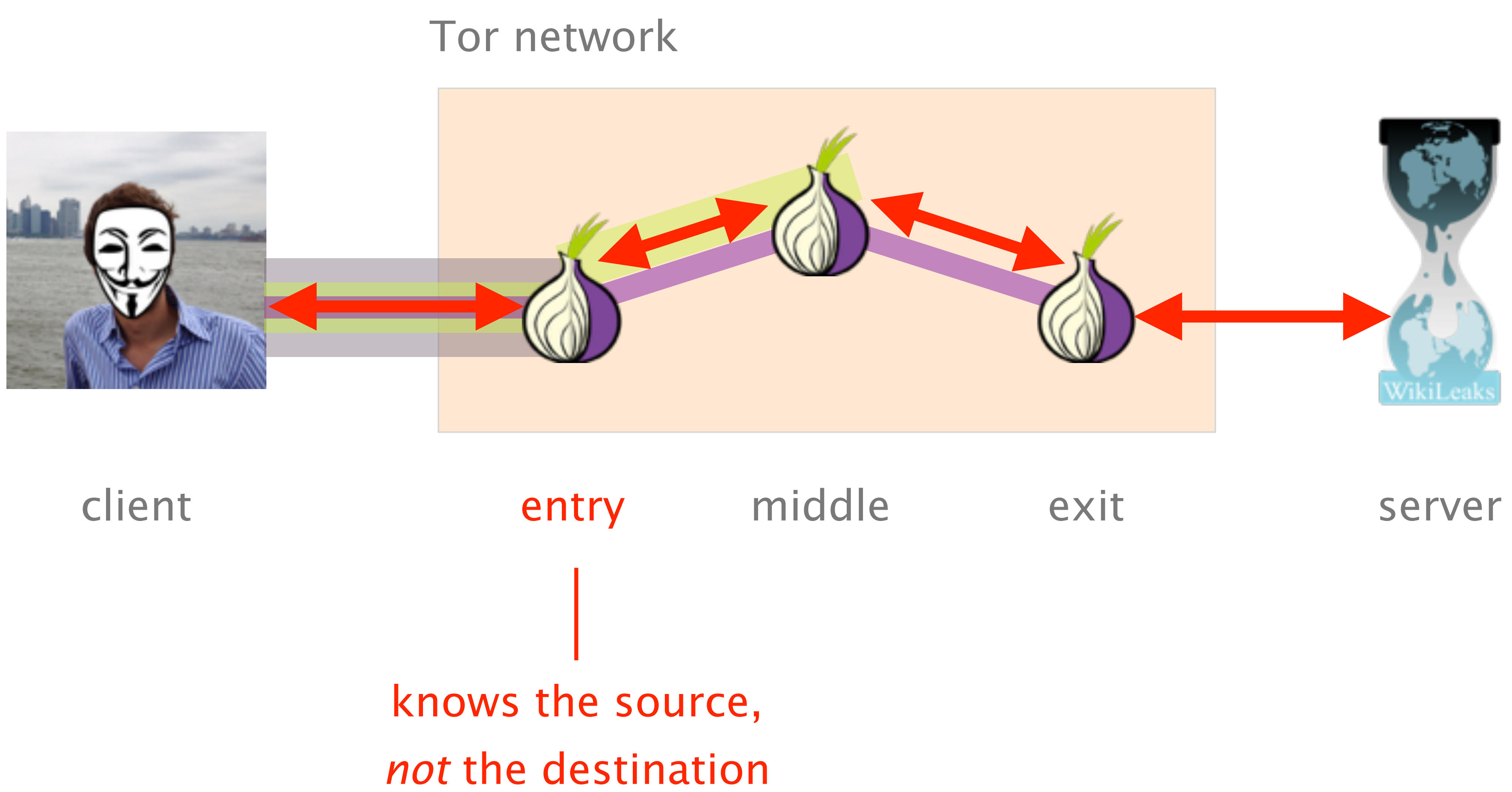

Tor network
client
entry
middle
exit
server
WikiLeaks

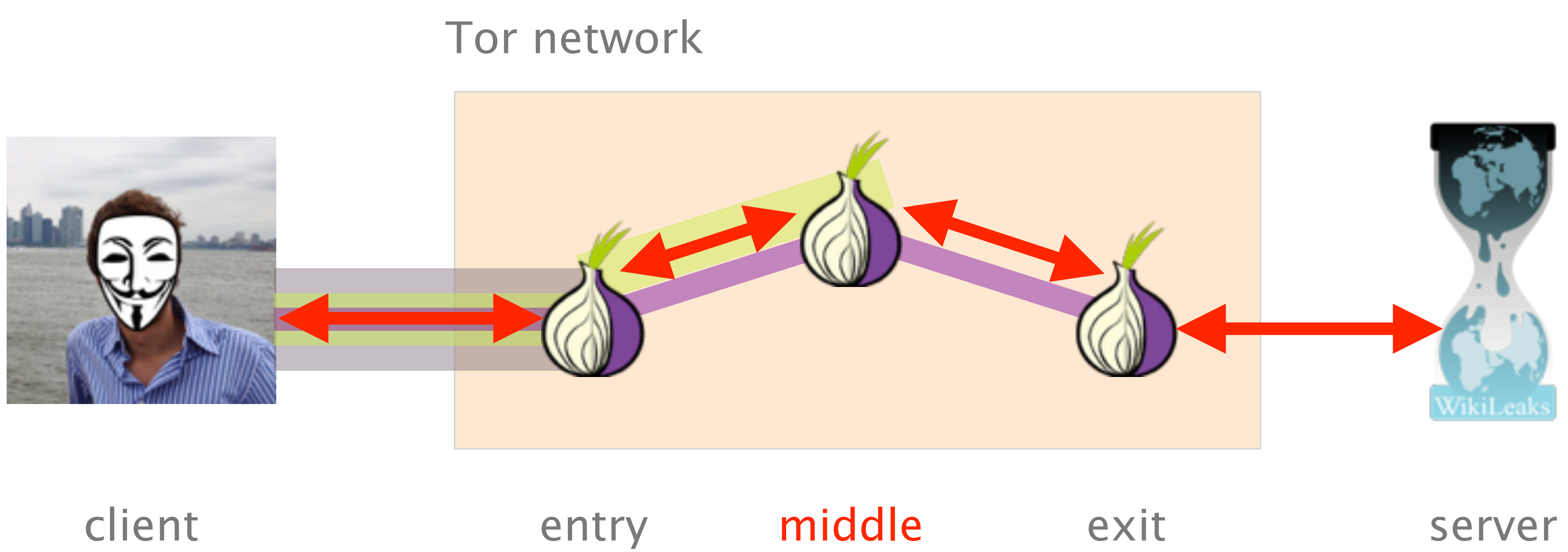
# Anonymous communication takes place by forwarding across consecutive tunnels



# Not a single Tor entity knows the association (client, server)

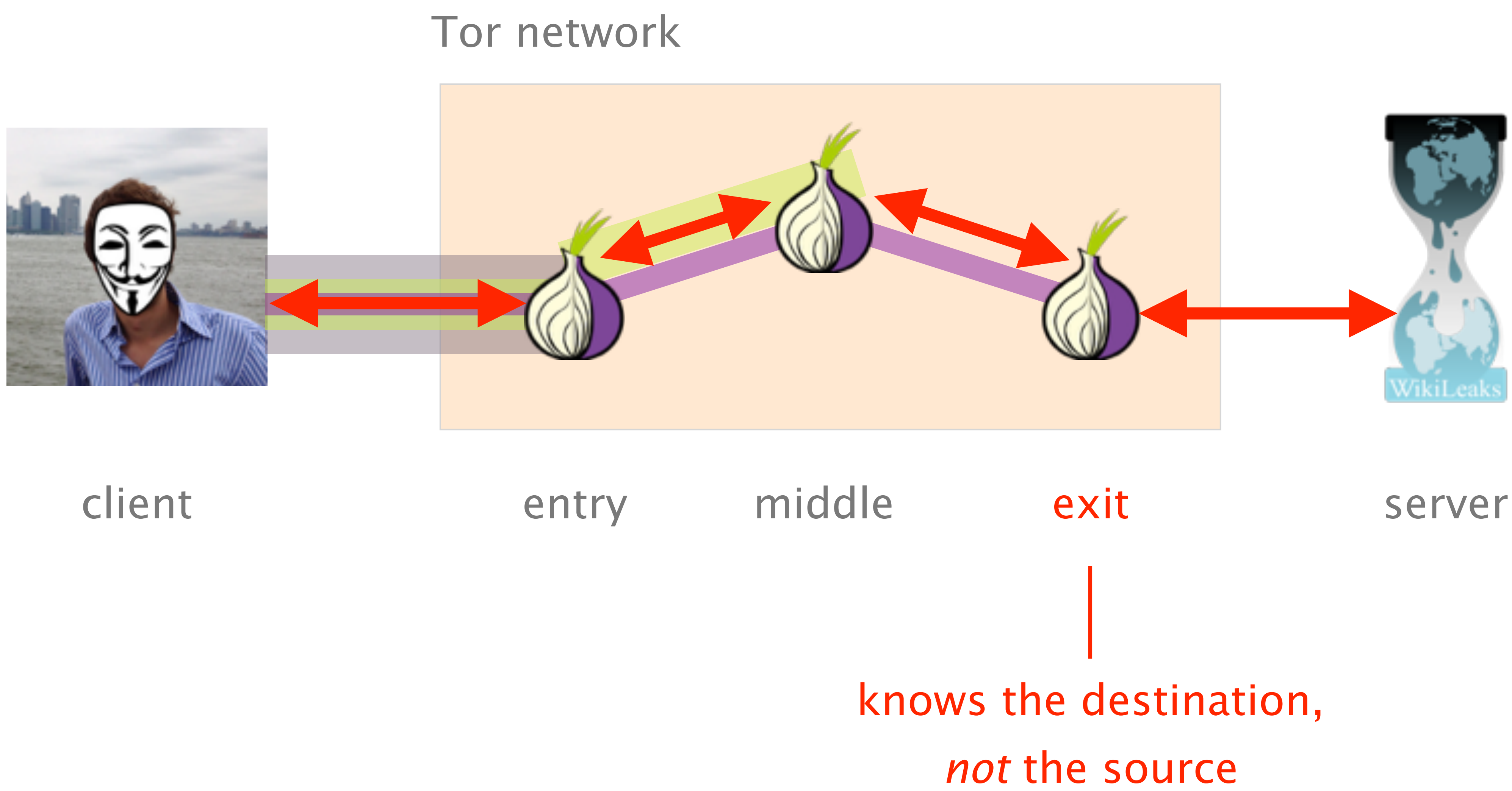



Tor network
WikiLeaks
client
entry
middle
exit
server
knows the source,
*not* the destination

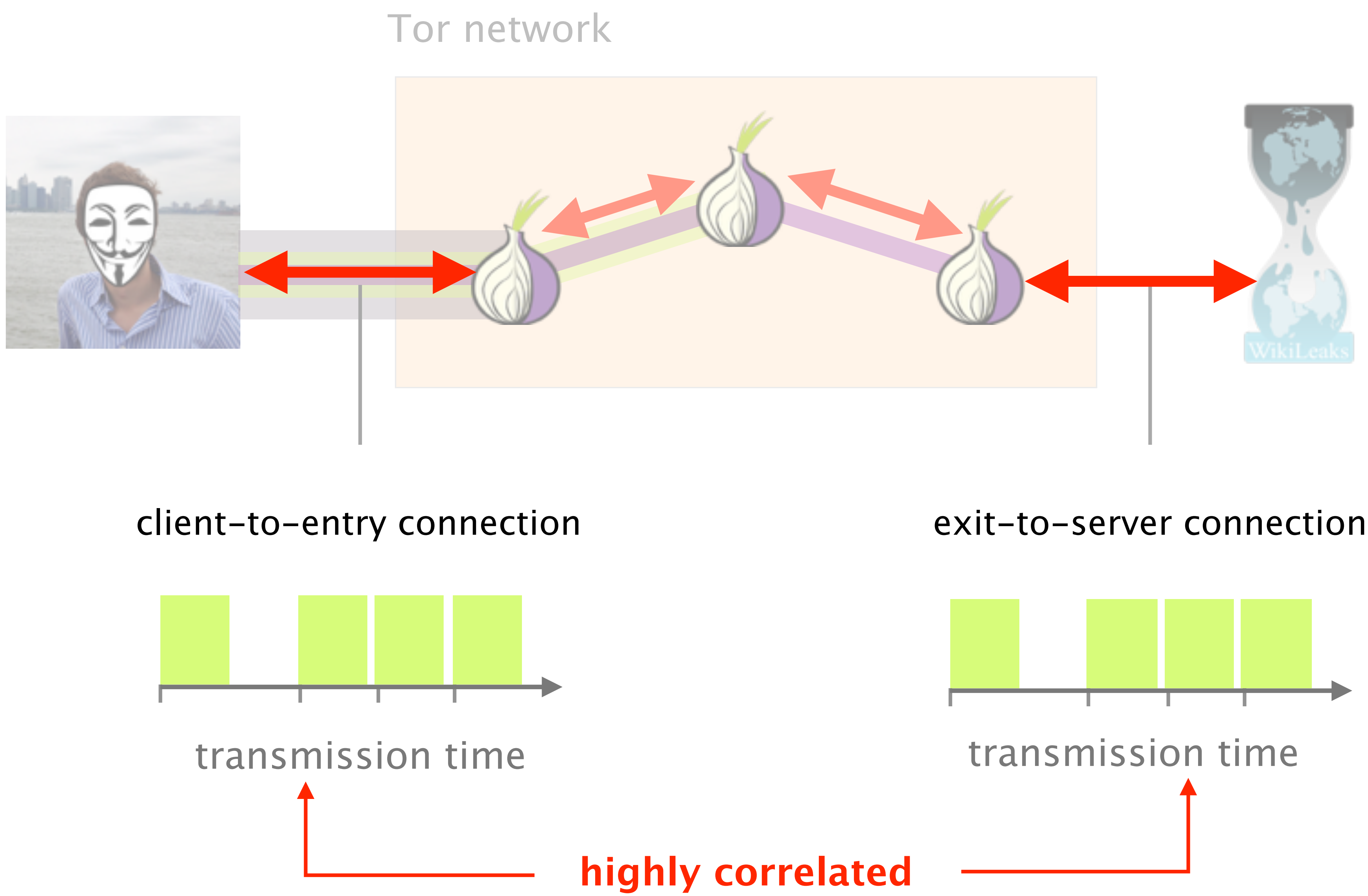



knows neither the source,

*nor* the destination


Tor network
WikiLeaks
client
entry
middle
exit
server
knows the destination,
*not* the source

# Traffic entering and leaving Tor is highly correlated



By correlating client-to-entry & exit-to-server flows, one can effectively deanonymize Tor users

Traffic correlation attacks require to see client-to-entry and exit-to-server traffic

Traffic correlation attacks require to see client-to-entry and exit-to-server traffic

How?

# Two ways

**Manipulate Tor**

malicious relays

**Manipulate routing**

malicious networks

Two ways

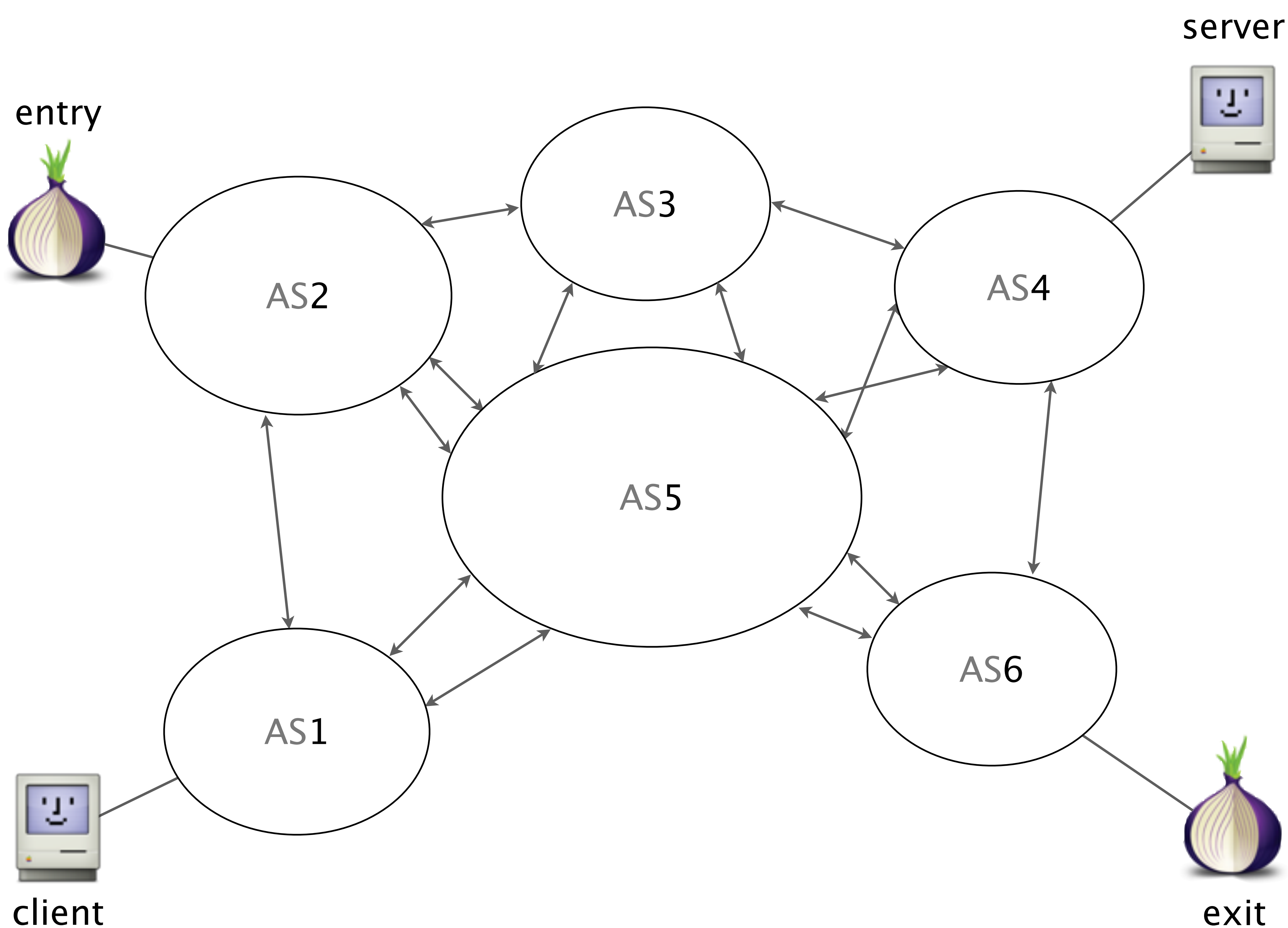
Manipulate Tor

malicious relays

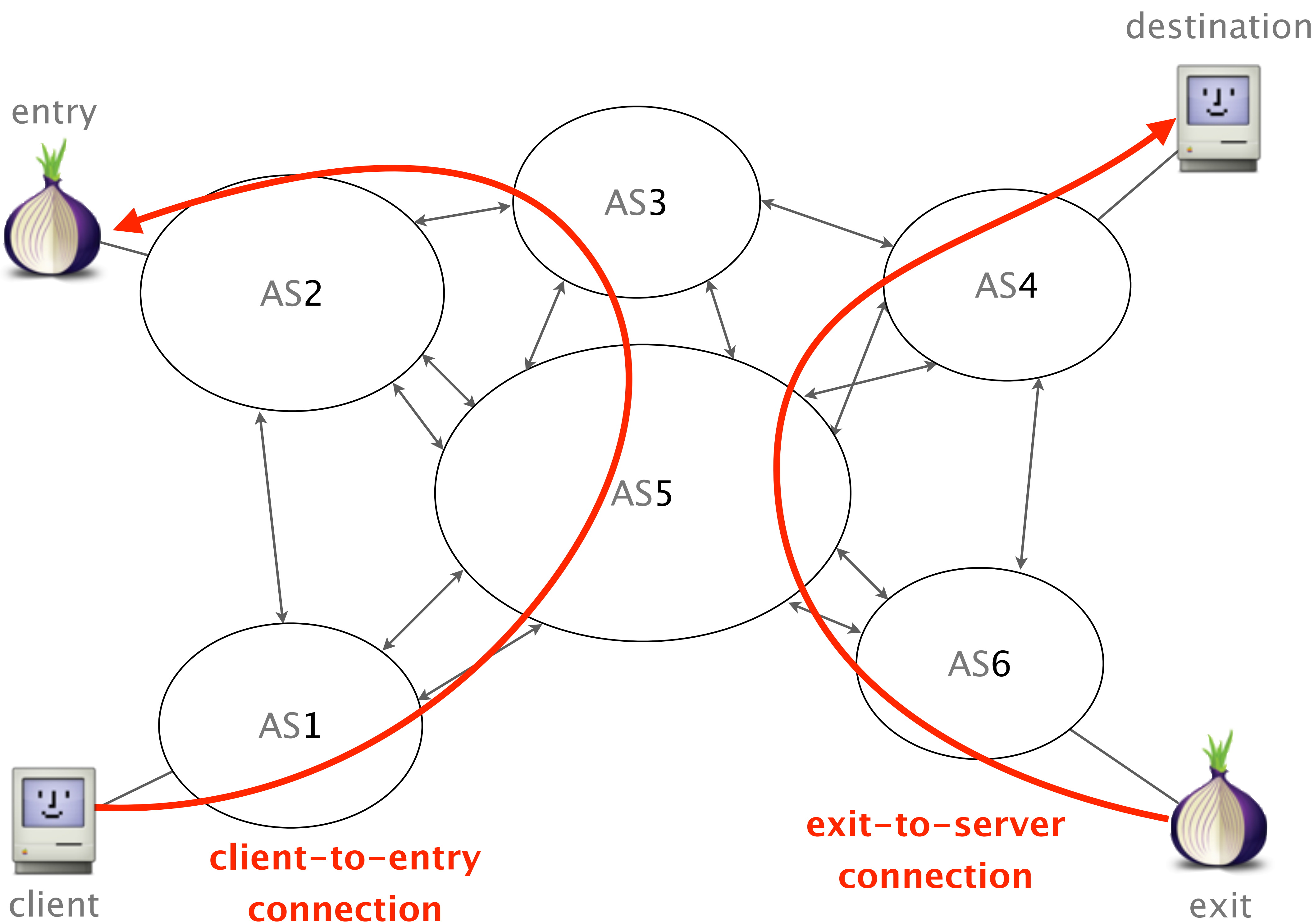
Manipulate routing

malicious networks

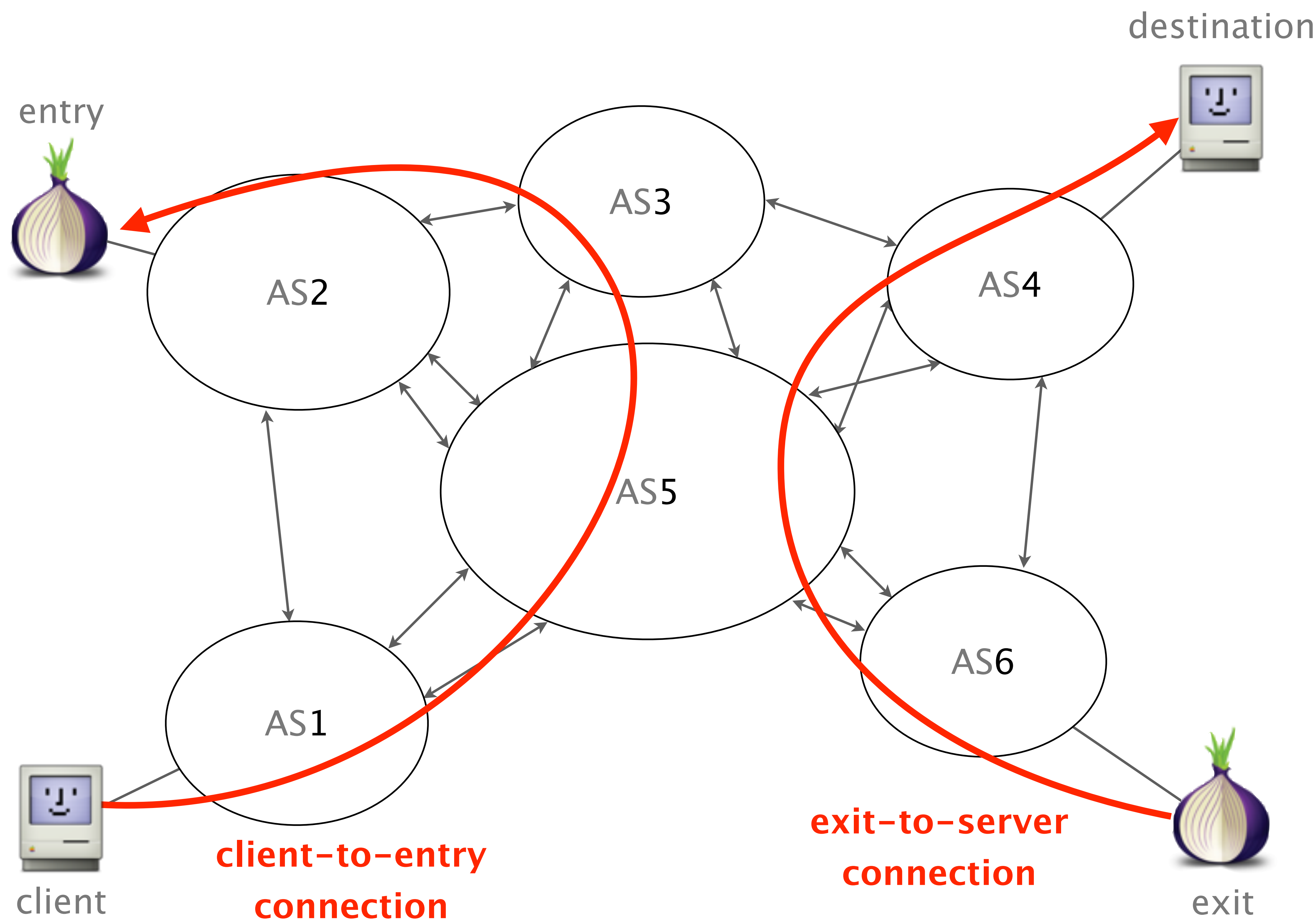
This talk

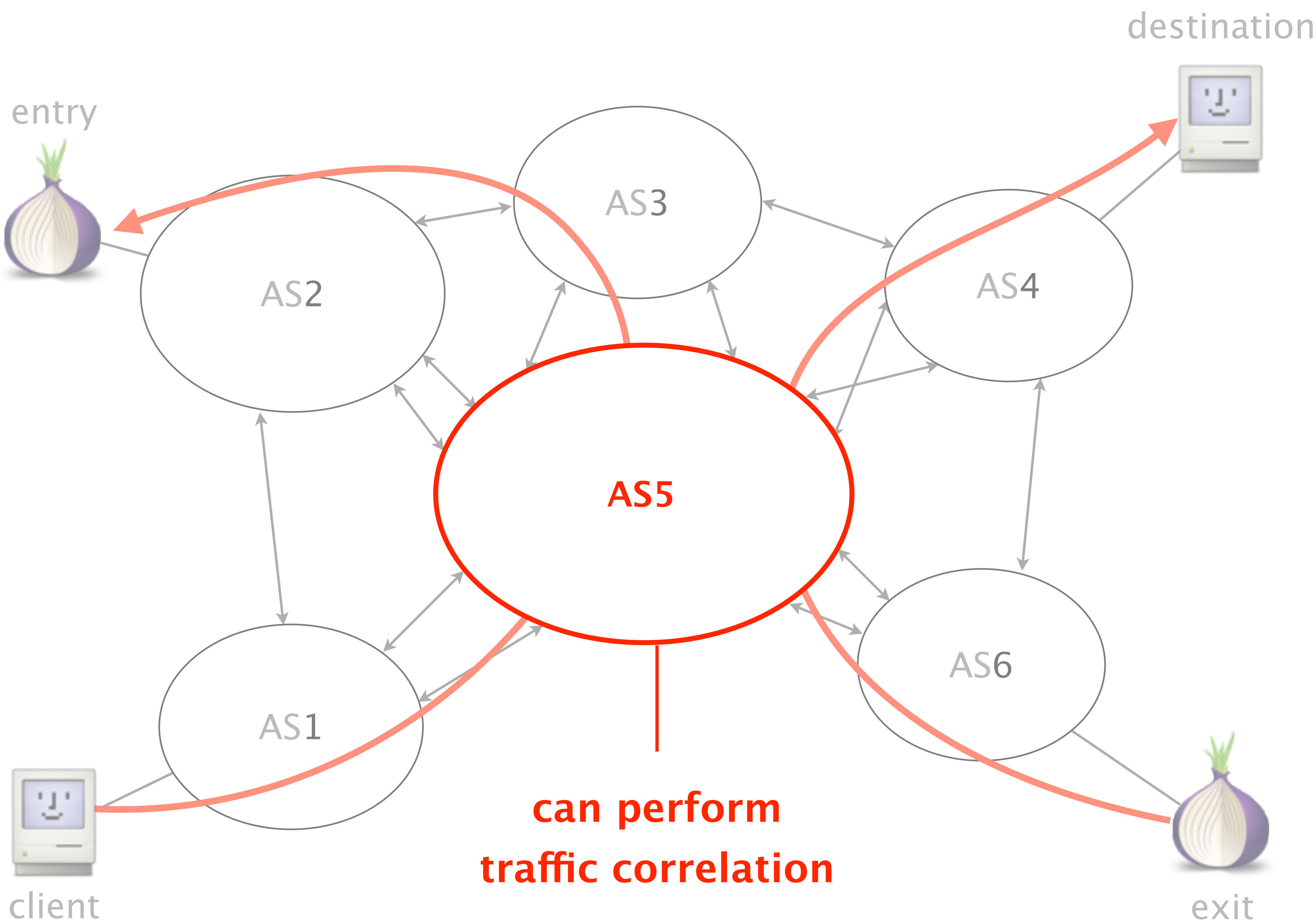

server
entry
AS3
AS2
AS4
AS5
AS6
AS1
client
exit

# Tor connections get routed according to BGP



Traffic correlation attacks require to see client-to-entry *and* exit-to-server traffic




destination
entry
AS3
AS2
AS4
AS5
AS6
AS1
can perform
traffic correlation
client
exit

# Network-level adversaries are known

| | Related work | |
|---|---|---|
| 2004 | Location diversity in anonymity networks | Feamster and Dingledine |
| 2007 | Sampled traffic analysis by Internet-exchange-level adversaries | Murdoch and Zieliński |
| 2009 | AS-awareness in Tor Path Selection | Edman and Syverson |
| 2013 | Traffic correlation on Tor by realistic adversaries | Johnson *et al.* |

However, these works assume

that the Internet is static

However, these works assume

that the Internet is static

… which is not the case

However, these works assume

that the Internet is static

… which is not the case

Contribution

What's the impact on Tor?

User anonymity decreases over time due to BGP dynamics

# User anonymity decreases over time due to BGP dynamics

BGP-induced causes

**Asymmetric routing**

path from A to B != from B to A

**Natural BGP convergence**

policy changes, failures, etc.

**Active BGP manipulation**

IP prefix hijack, interception (MITM)…

# Anonymity on Quicksand

## Using BGP to compromise Tor

1 **Attacks**

All your traffic belongs to me

2 **Results**

Eyes wide open

3 **Countermeasures**

Close the curtains

# Anonymity on Quicksand

## Using BGP to compromise Tor

1 **Attacks**

All your traffic belongs to me

**Results**

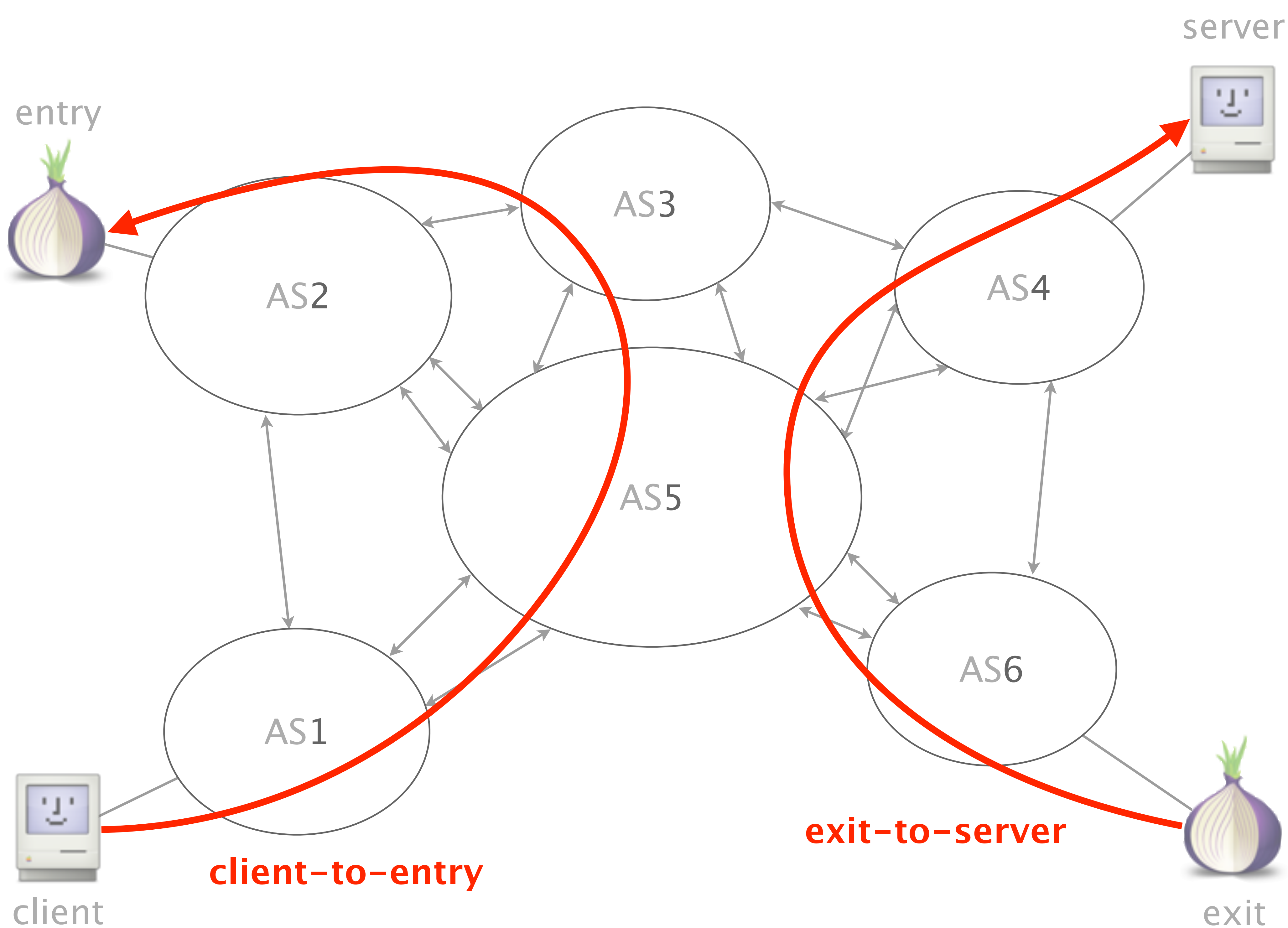
Eyes wide open

**Countermeasures**

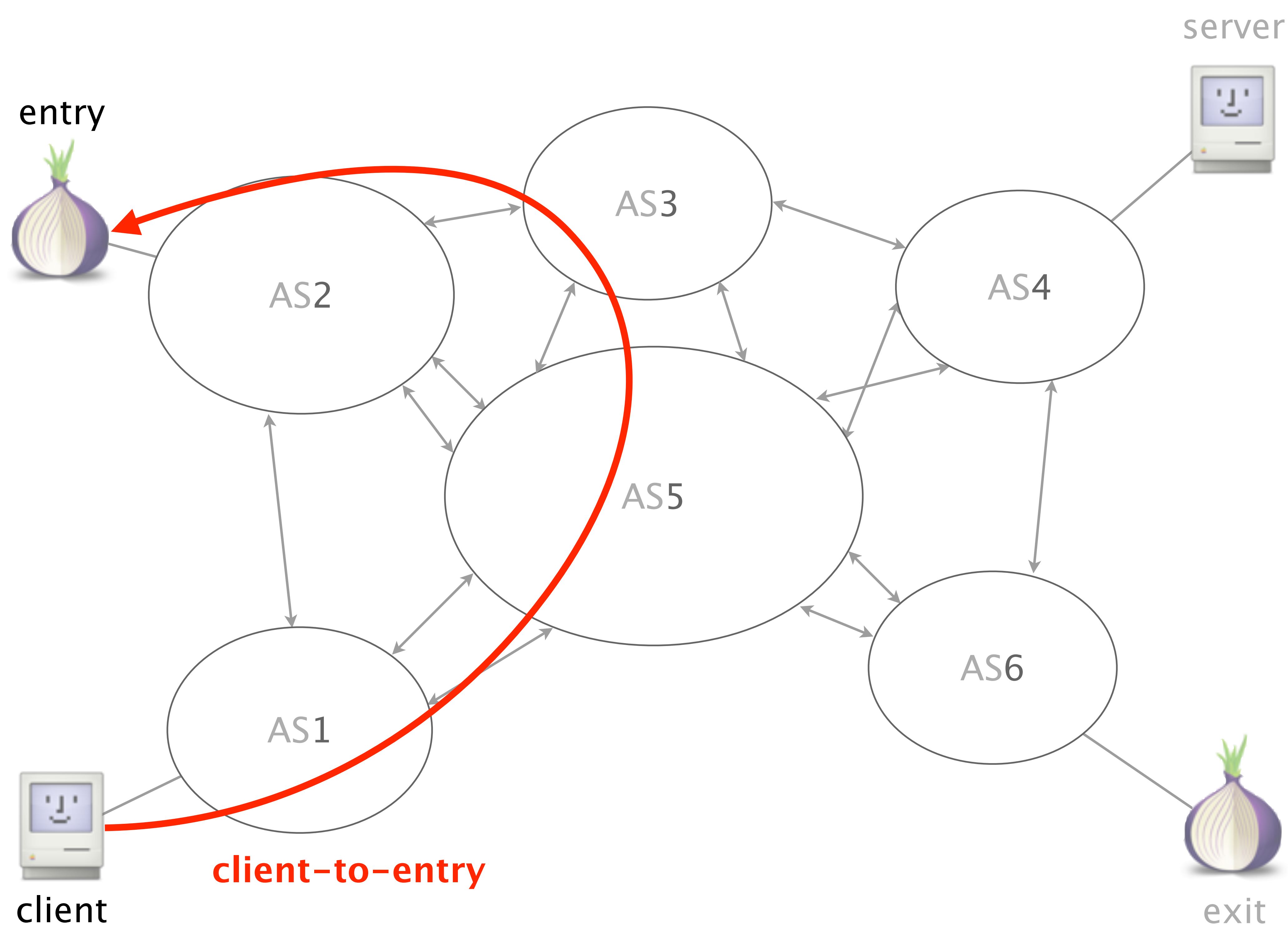
Close the curtains

#1. Asymmetric routing increases the numbers of AS-level adversaries

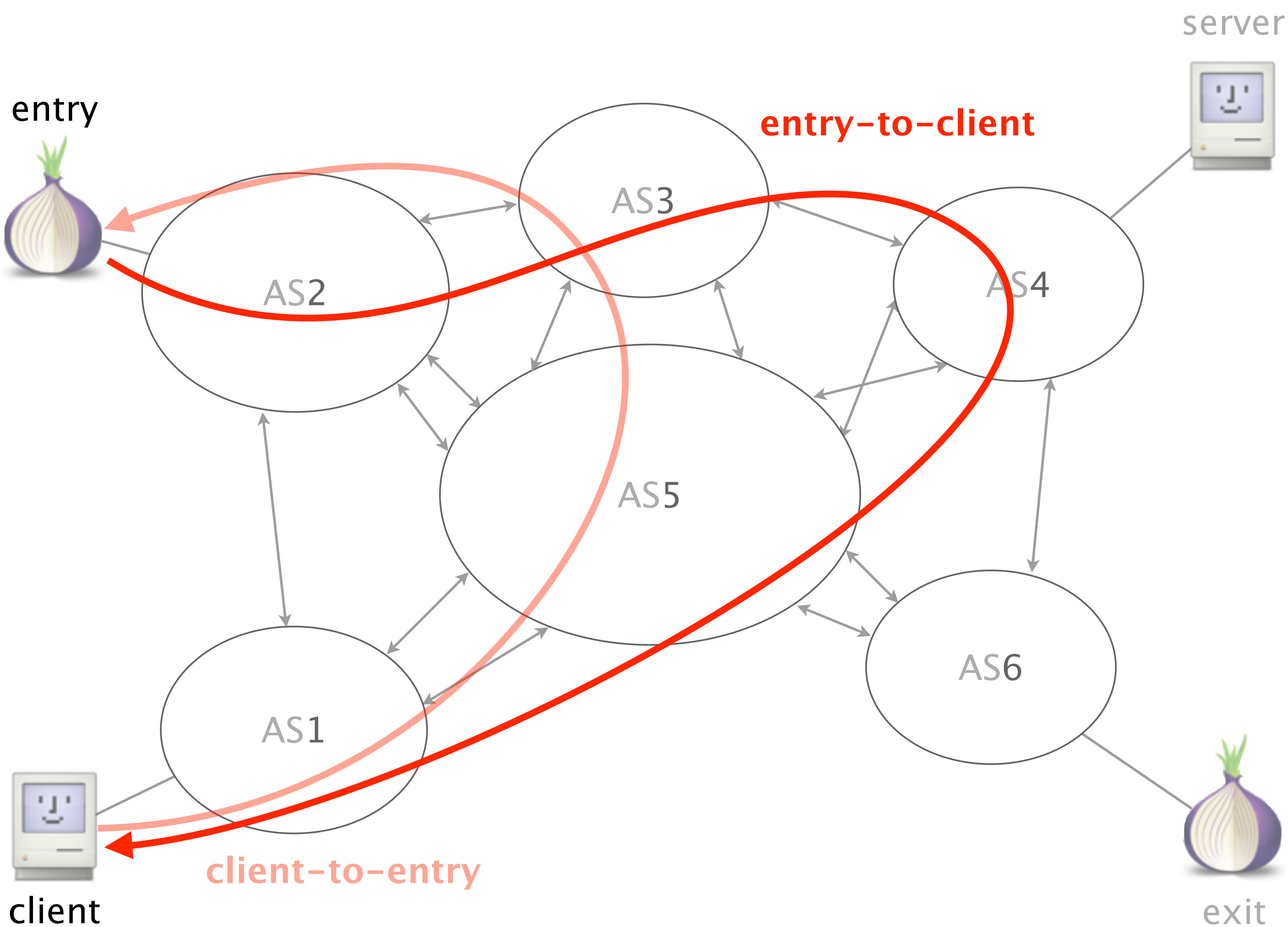
So far, we have considered one side of Tor traffic: client–to–entry and exit–to–server



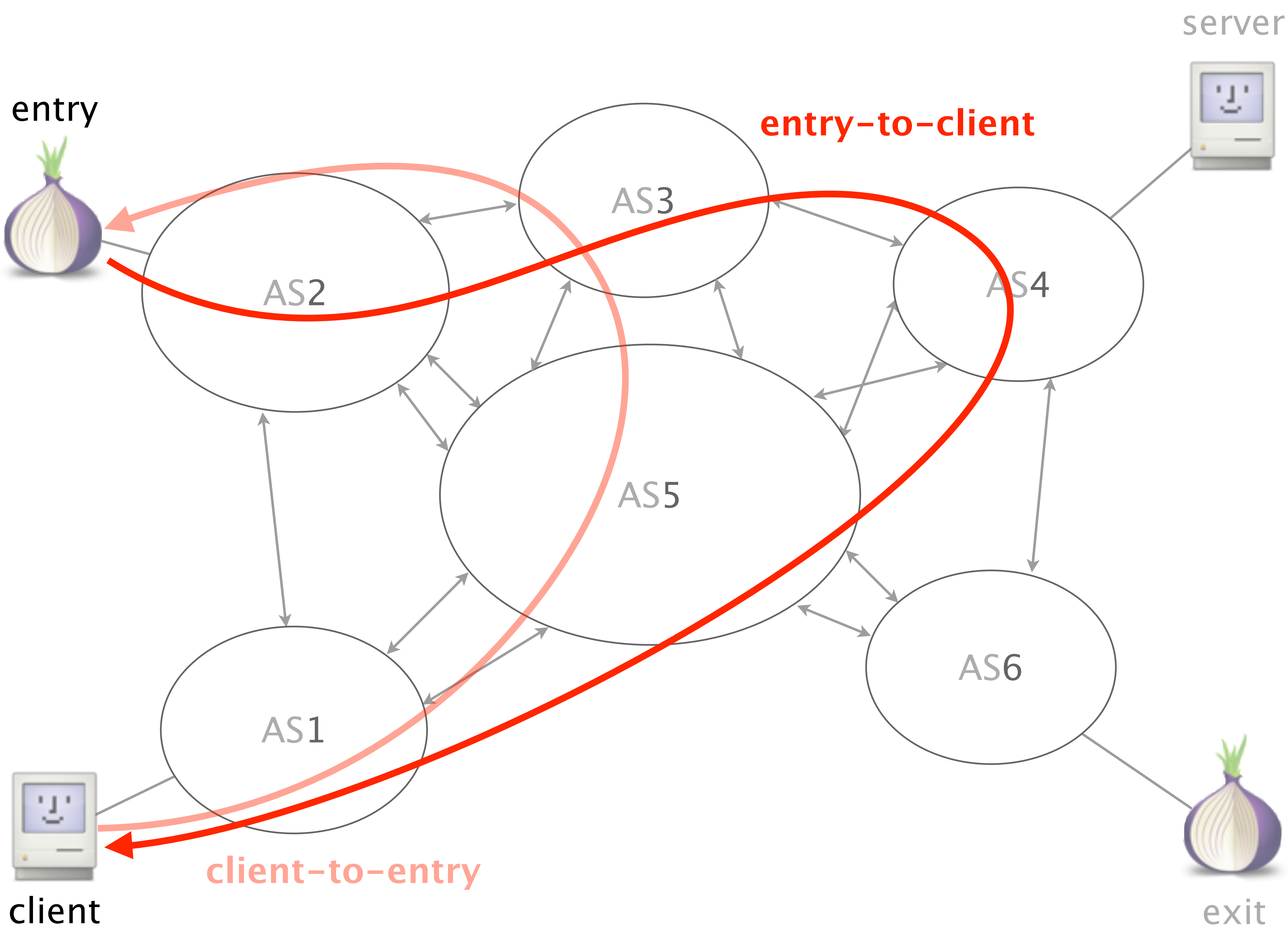
# However, because of policies, routing is often *asymmetric*



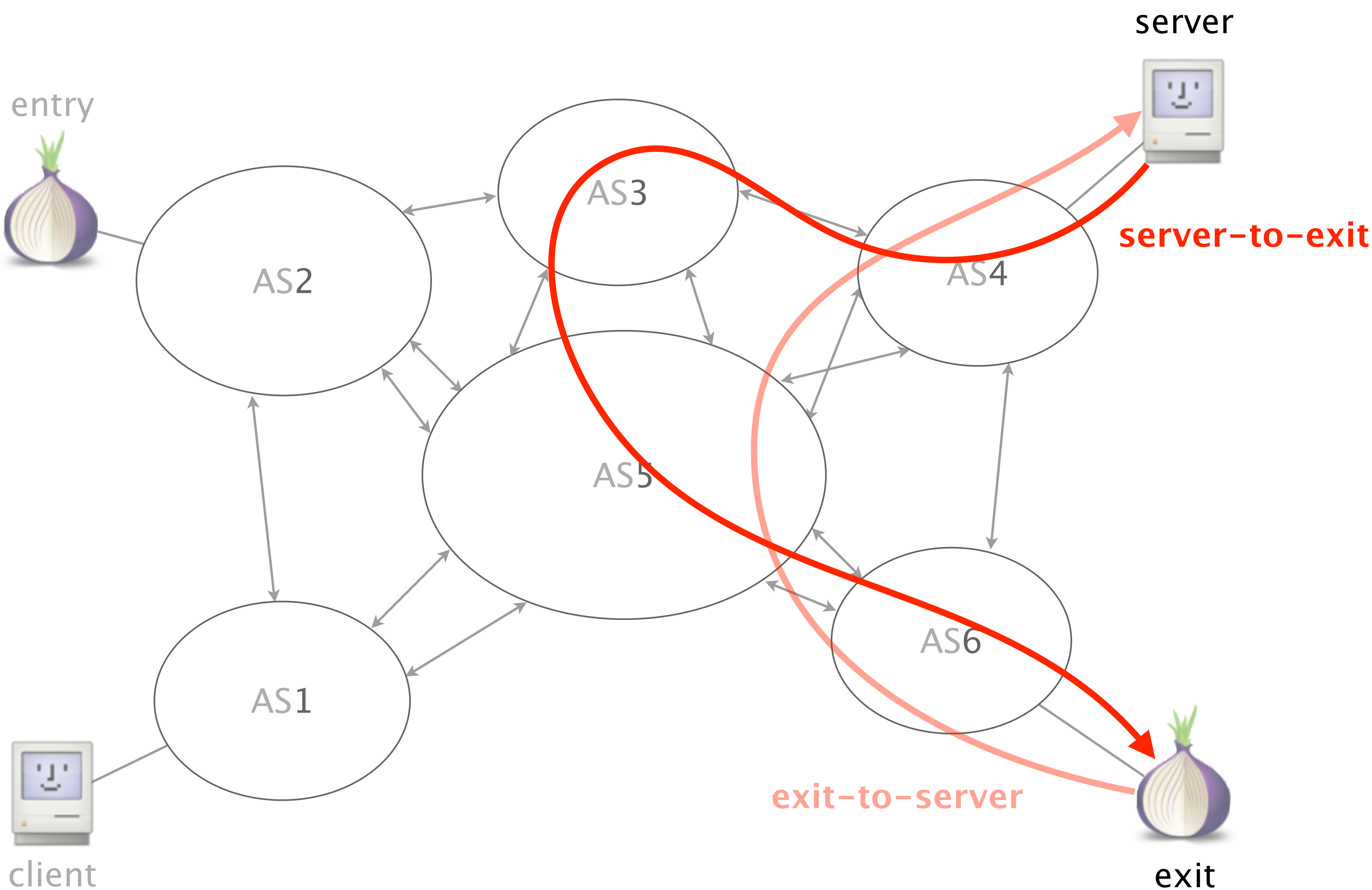
However, because of policies, routing is often *asymmetric*



While AS4 does not see client–to–entry traffic, it sees entry–to–client traffic



# The same applies to server-to-exit traffic



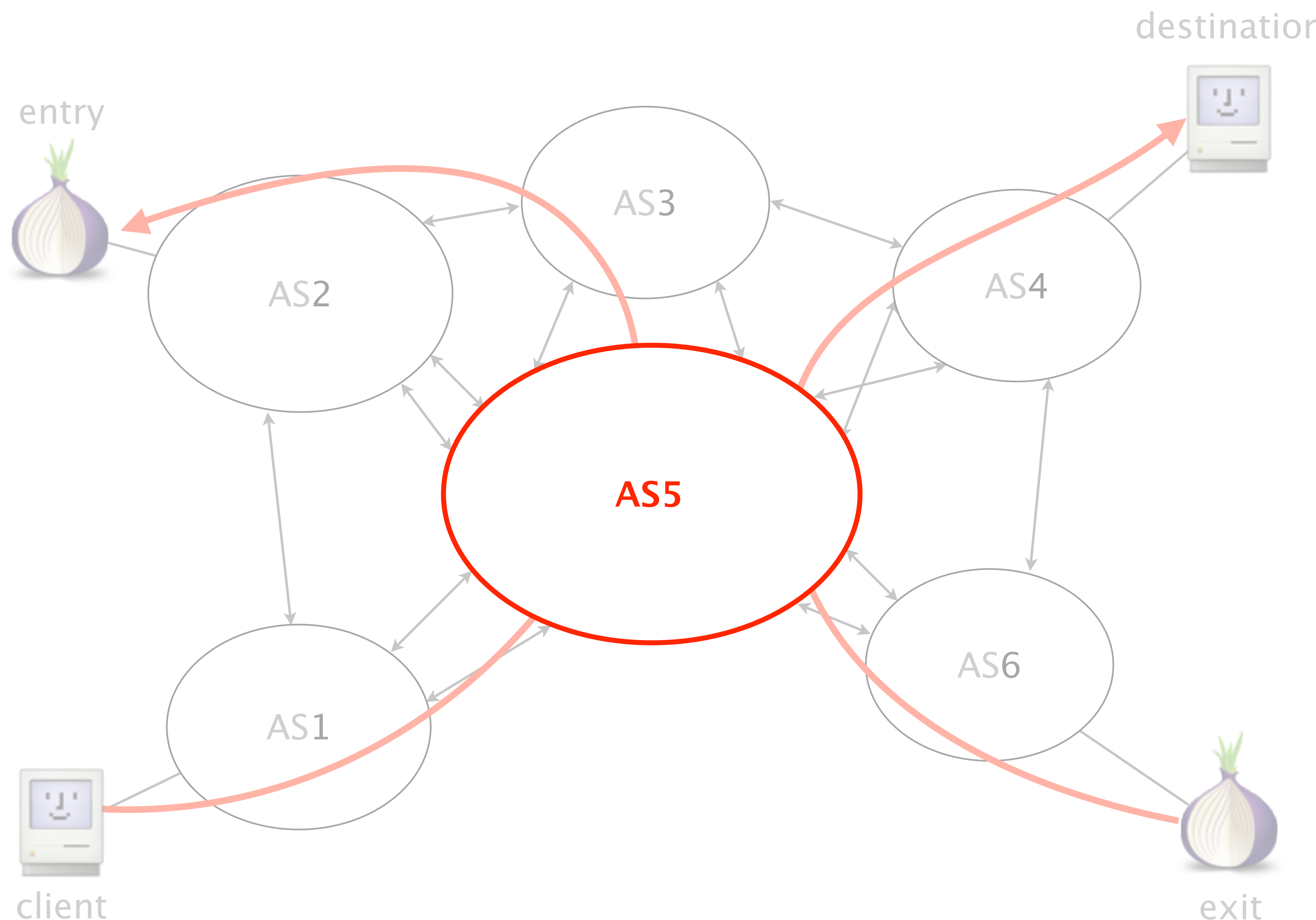
In terms of timing properties,
both sides of a TCP connection are highly correlated

# In terms of timing properties, both sides of a TCP connection are highly correlated

When collecting TCP timing information,

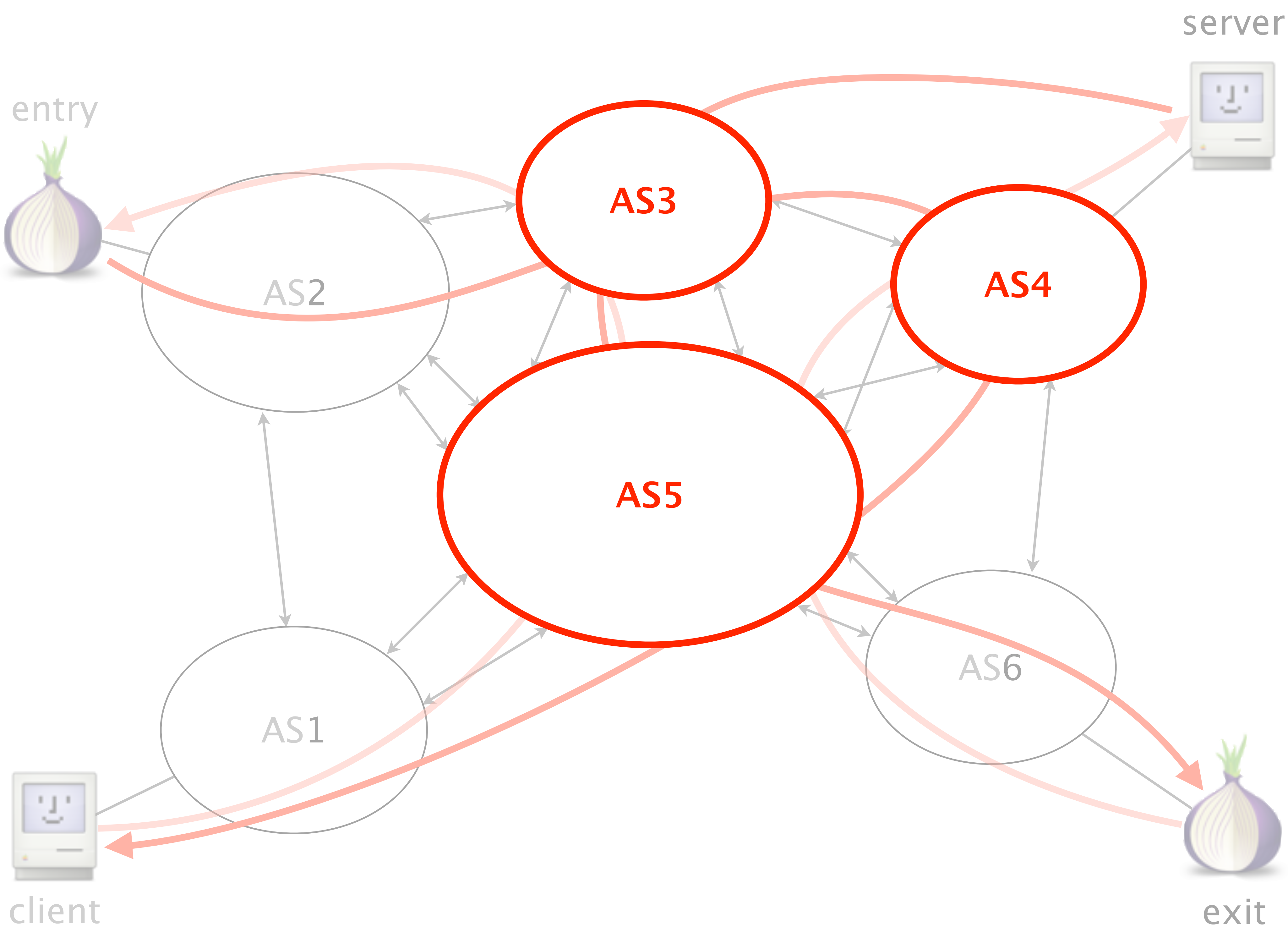
seeing one direction (*e.g.*, data packets)
is almost equivalent to
seeing two directions (ACKs & data packets)

Considering only one direction,
only AS5 is potentially compromising



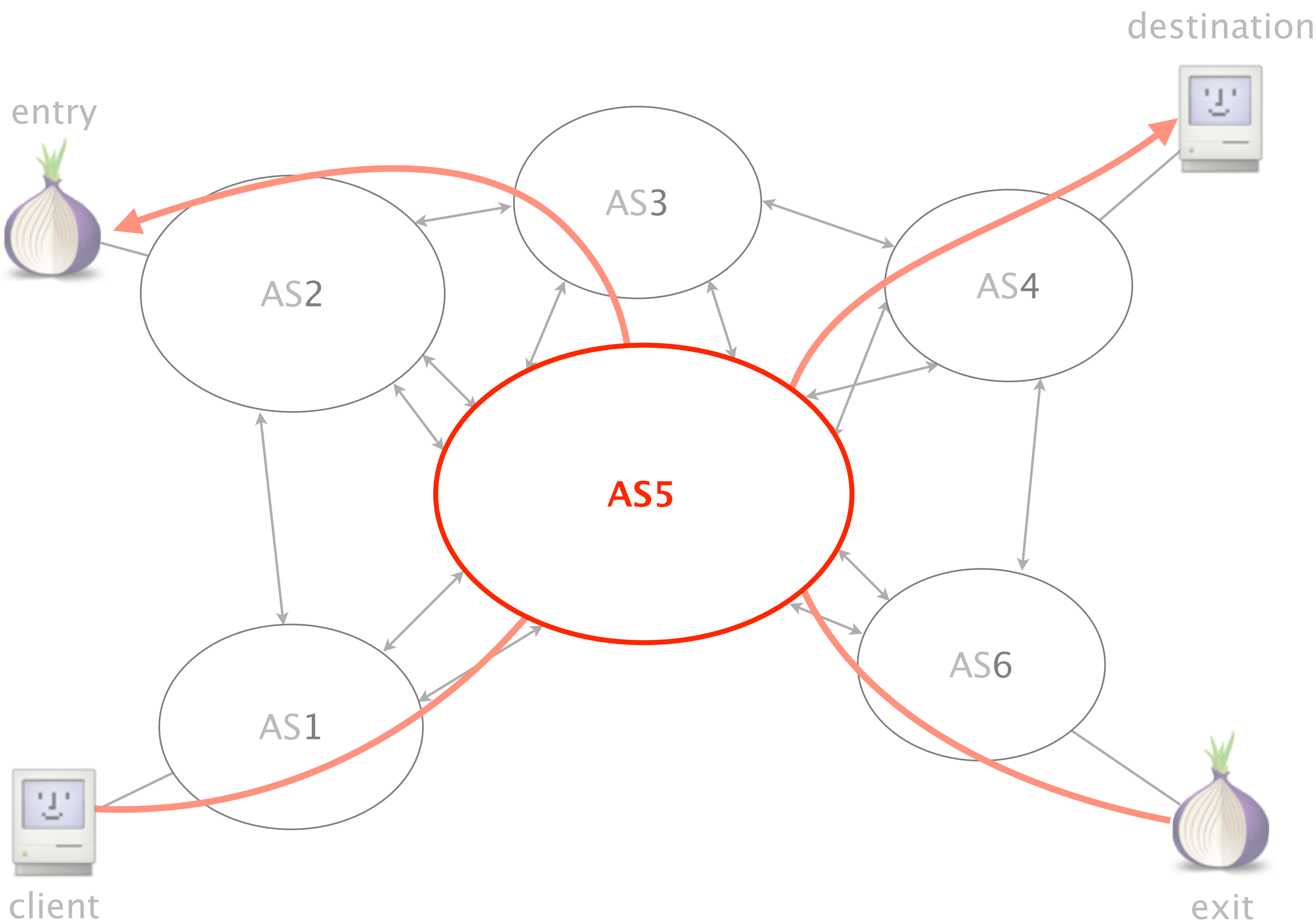
Considering both directions,

AS3, AS4 and AS5 are potentially compromising

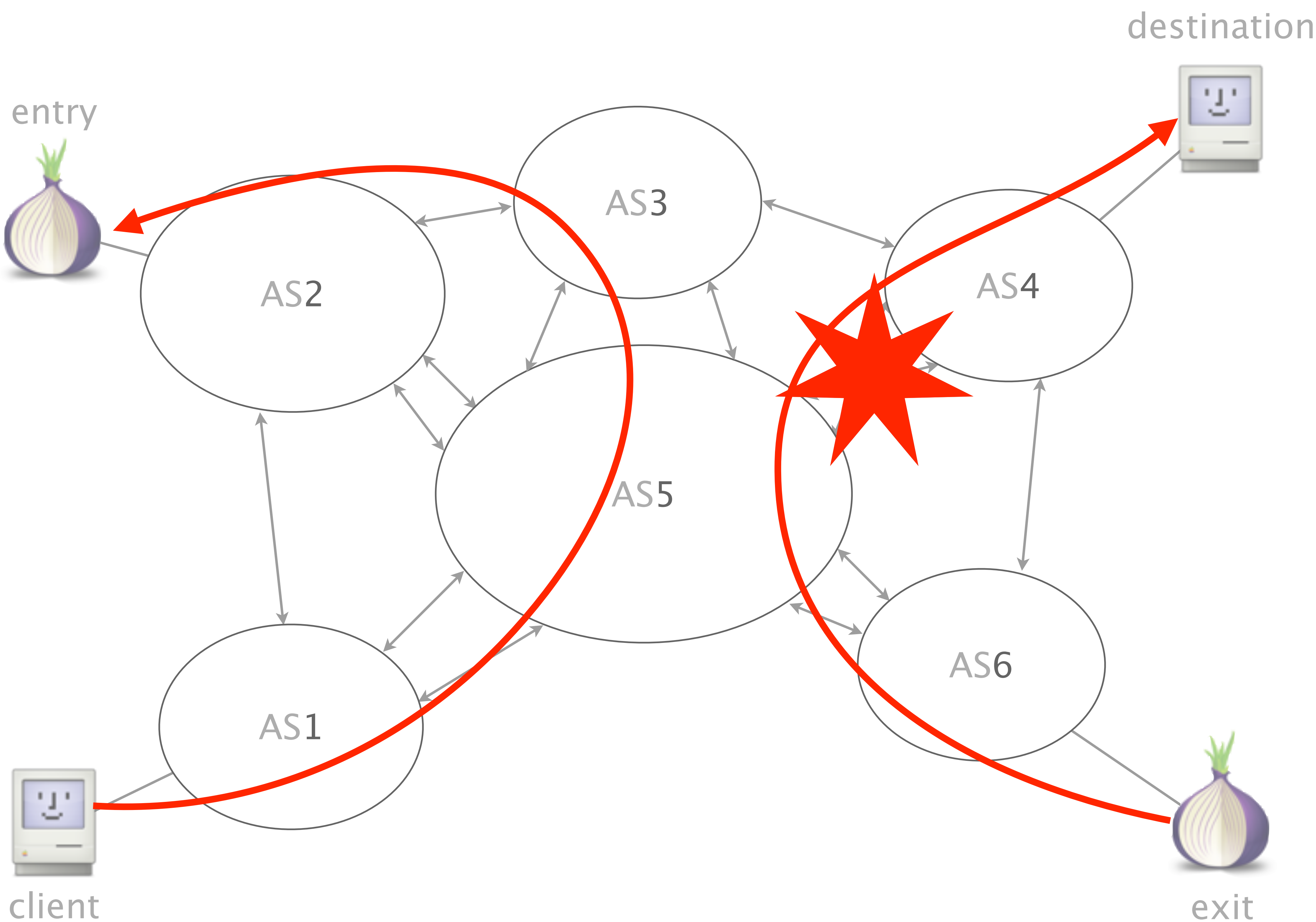


#2. Natural BGP dynamics increases the number of AS-level adversaries

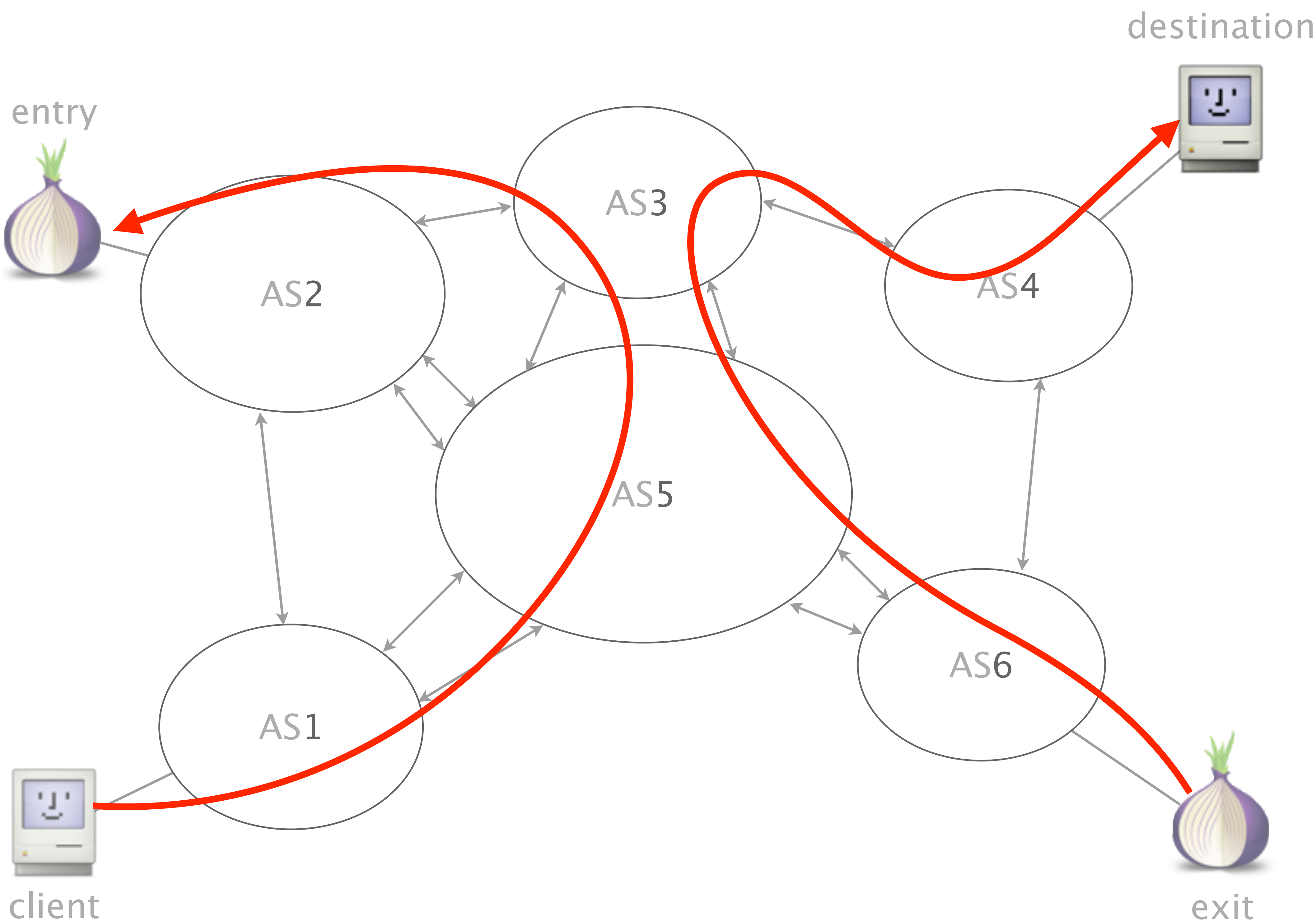
# Initially, only AS5 is compromising



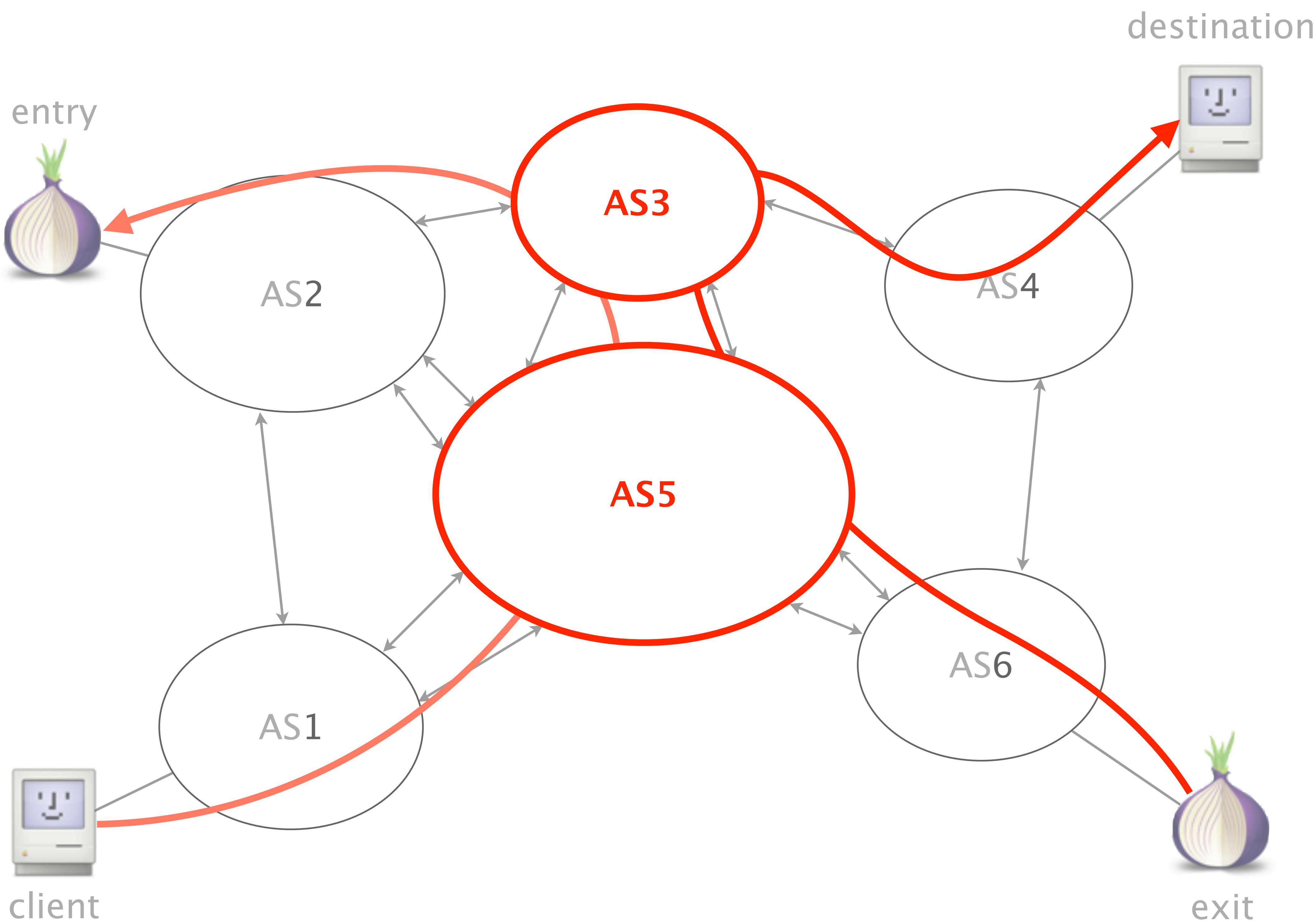
# Assume that the link between AS4 and AS5 fails



# Traffic gets rerouted via AS3

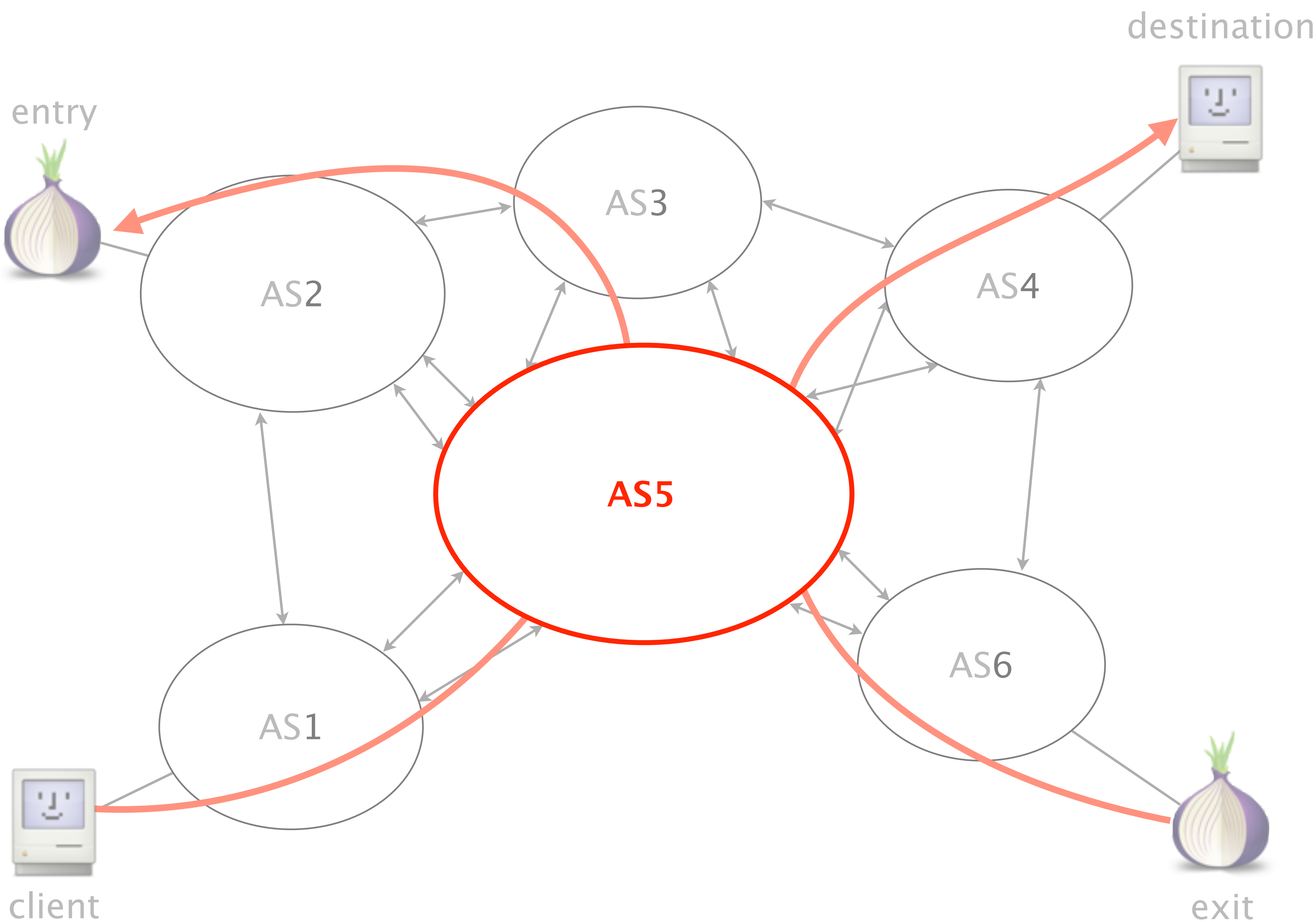


# Now, both AS3 and AS5 are seeing client-to-entry and exit-to-server traffic

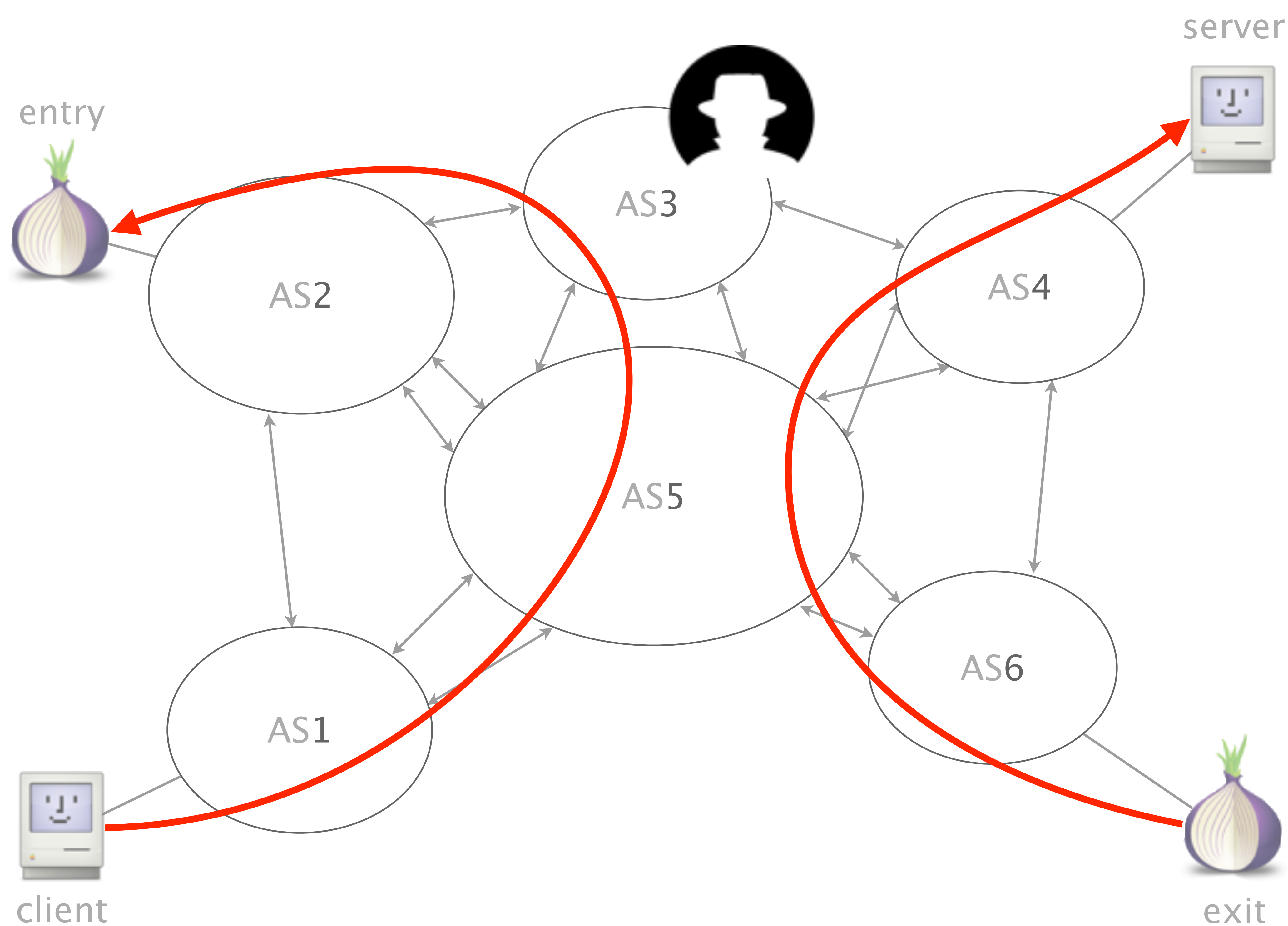


#3. BGP hijacking attacks enable on-demand, fine-grained Tor attacks

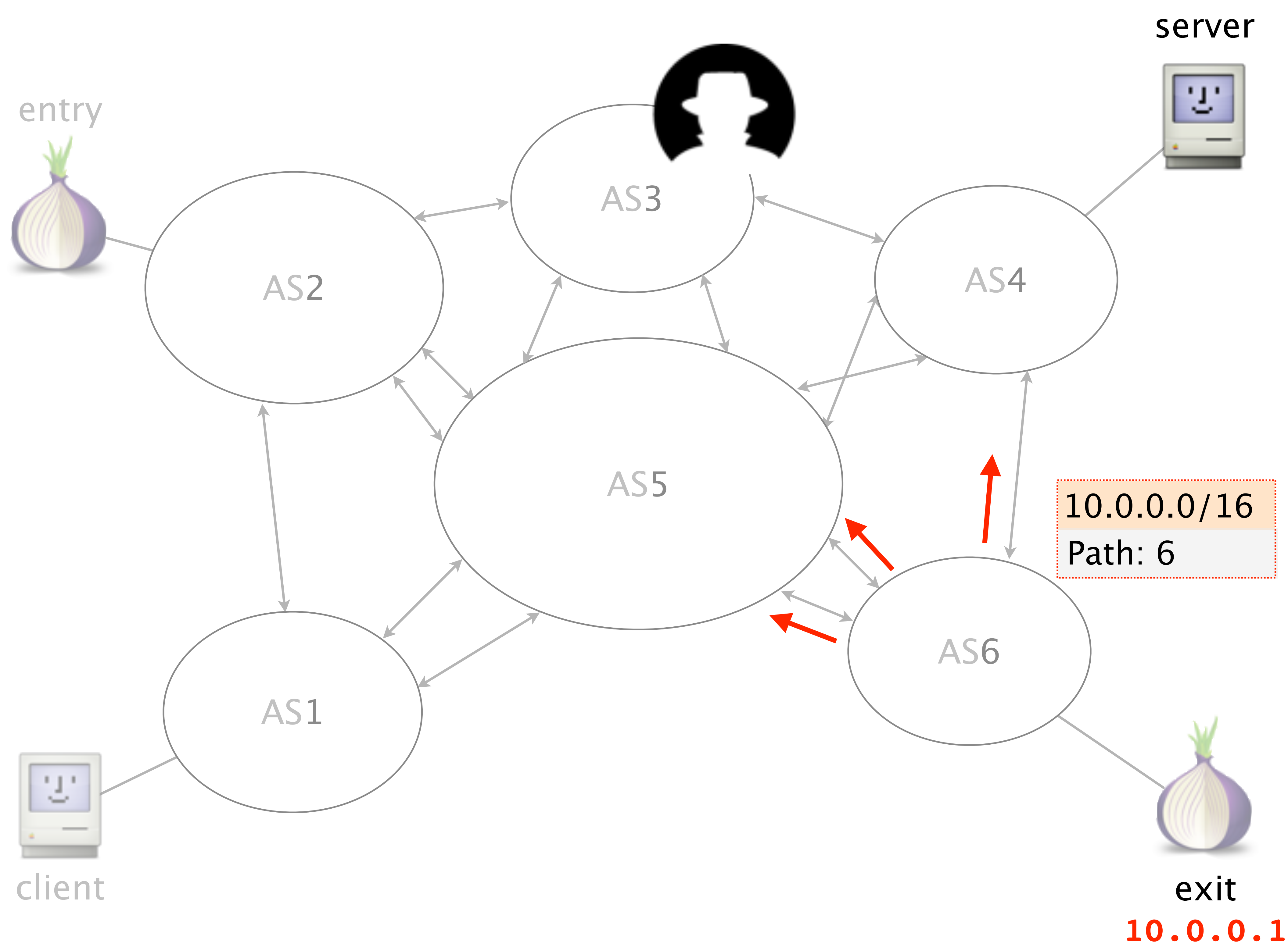
# Initially, only AS5 is compromising



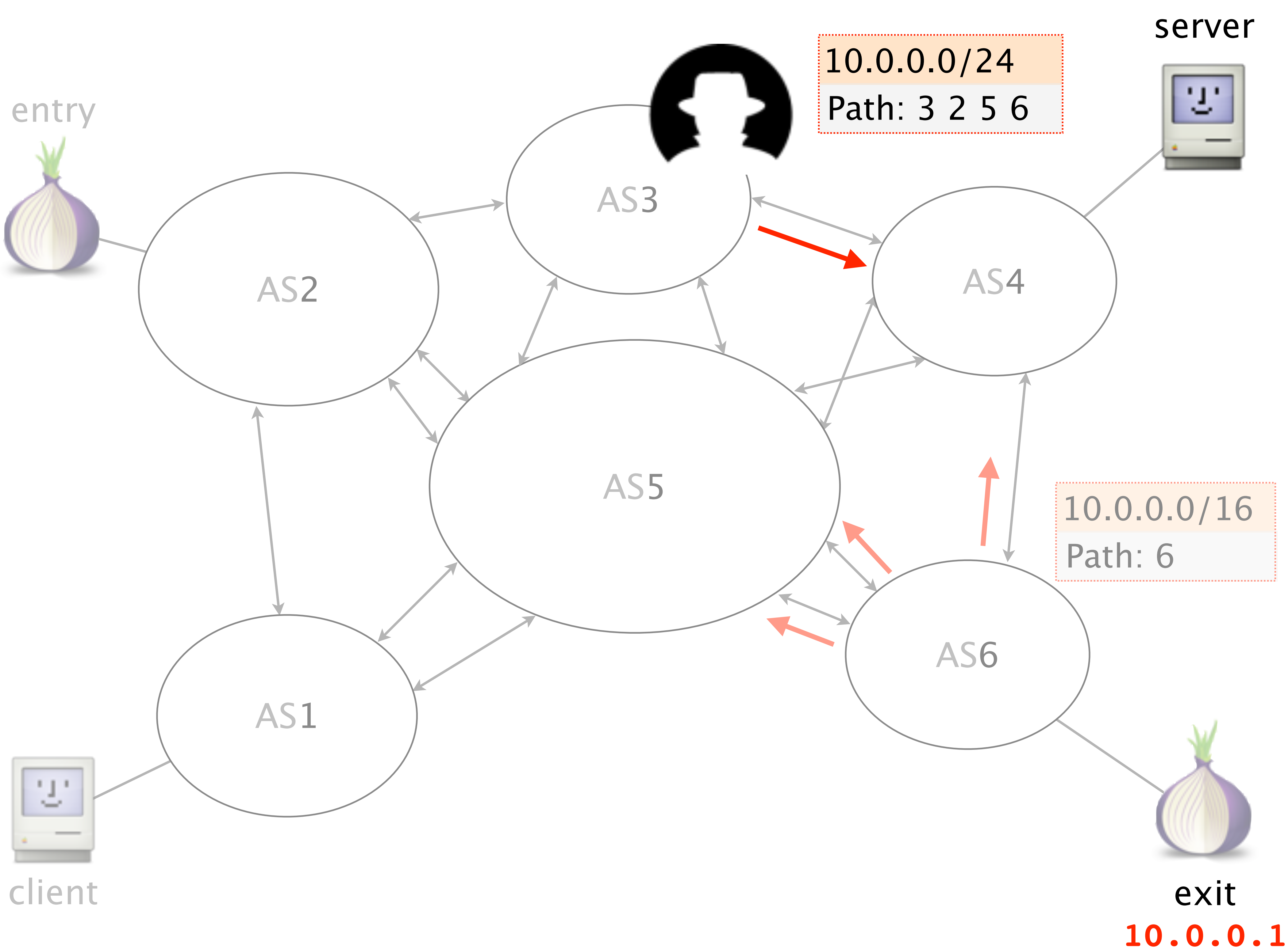
Assume that AS3 is a malicious AS, and wants to observe Tor traffic

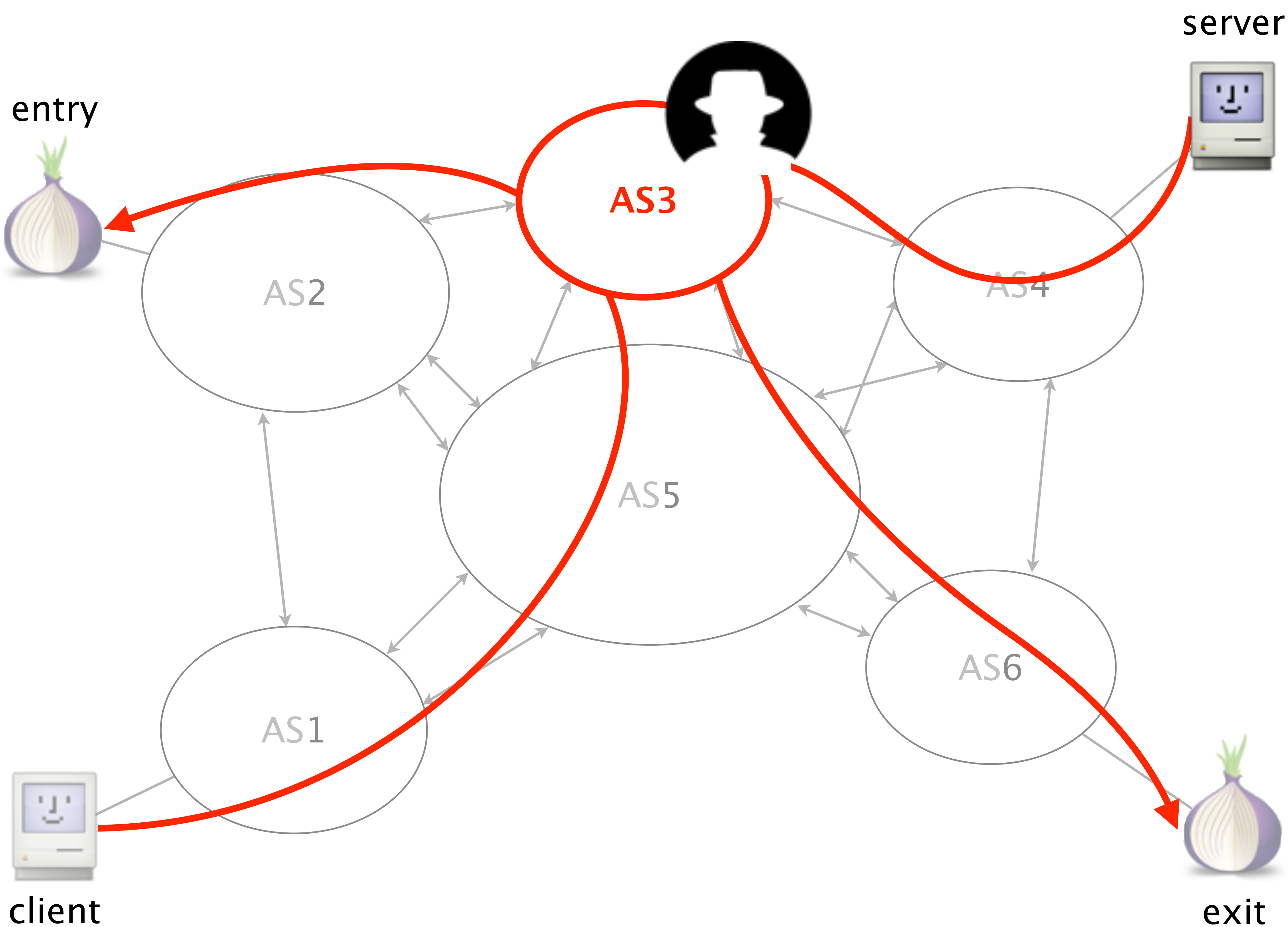


# AS3 can put itself on server-to-exit paths by hijacking Tor prefixes



# AS3 can put itself on server-to-exit paths by hijacking Tor prefixes




server
entry
AS3
AS2
AS4
AS5
AS1
AS6
client
exit

# In November 2010,
# China Telecom hijacked 50k prefixes during ~20 min

NOVEMBER 18, 2010 | COMMENTS (10) | VIEWS: 778 | ENGINEERING, INTERNET, POLITICS, SECURITY | JIM COWIE

## China's 18-Minute Mystery

When the US-China Economic and Security Review Commission released its report to Congress this week, something slightly unusual happened: *people read it*. And there, buried on pages 236-247, a mystery was revealed, and the media have greedily amplified it.

Did China's government really divert 15% of the Internet's traffic for eighteen minutes in April, effortlessly intercepting sensitive traffic in flight, and generally creating a massively embarrassing man-in-the-middle attack on vulnerable global communications?

China Telecom

always sees traffic between
its customer and entry relays

During the attack, it also

saw traffic to/from exit relays
for a non-trivial fraction of traffic

Intentional? No one knows.

# Anonymity on Quicksand

## Using BGP to compromise Tor

Attacks

All your traffic belongs to me

2 Results

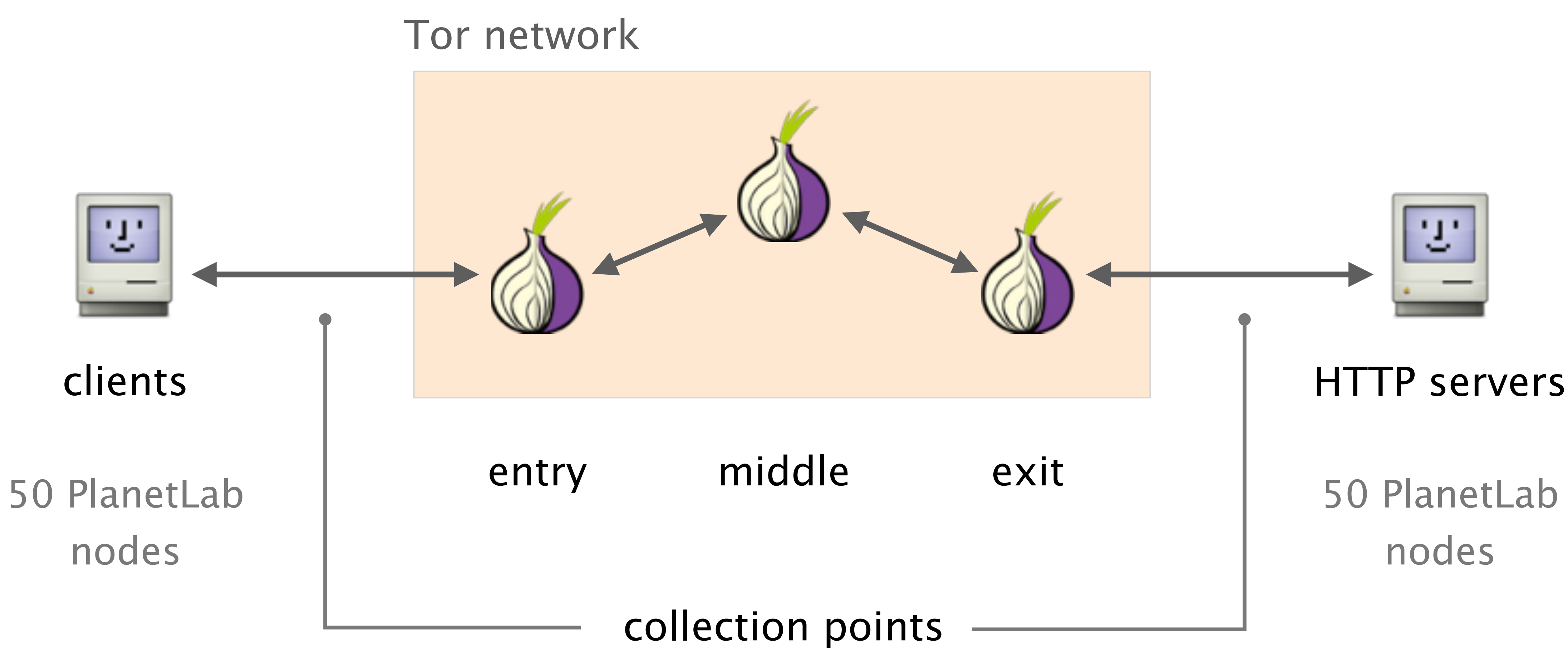
Eyes wide open

Countermeasures

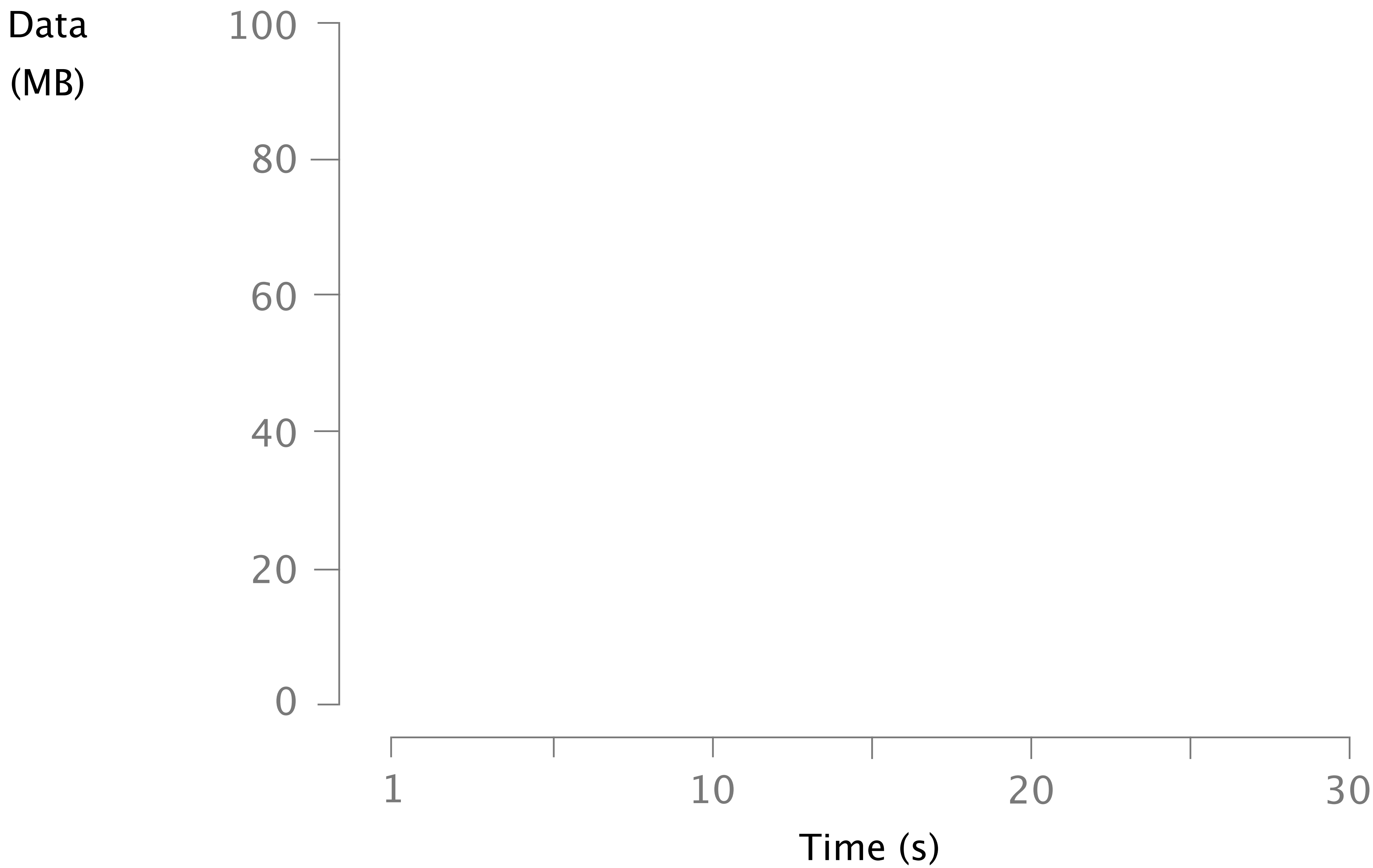
Close the curtains

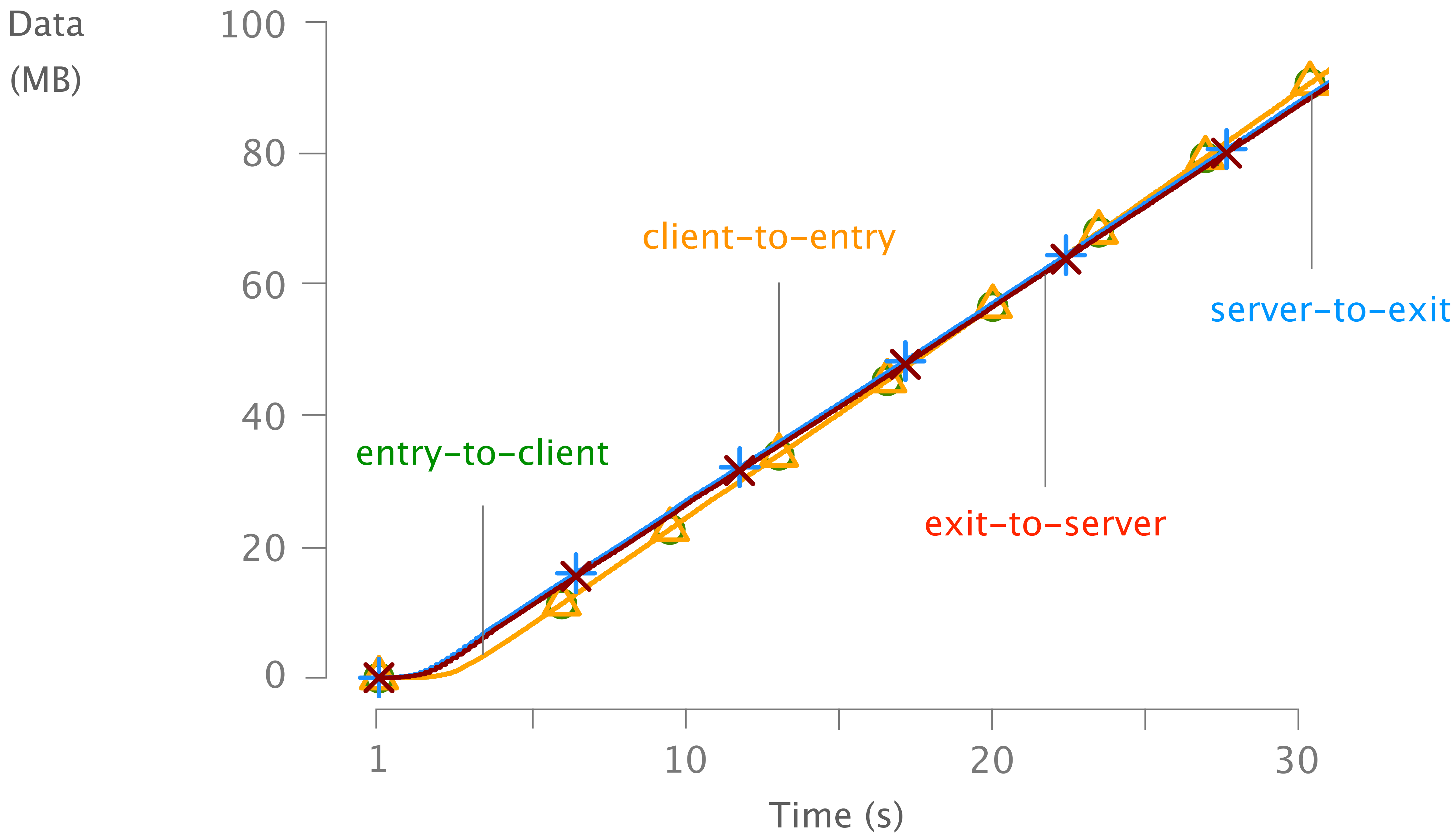
## #1. Asymmetric traffic analysis is highly efficient

# We collected traces by downloading 100 Mb files through Tor

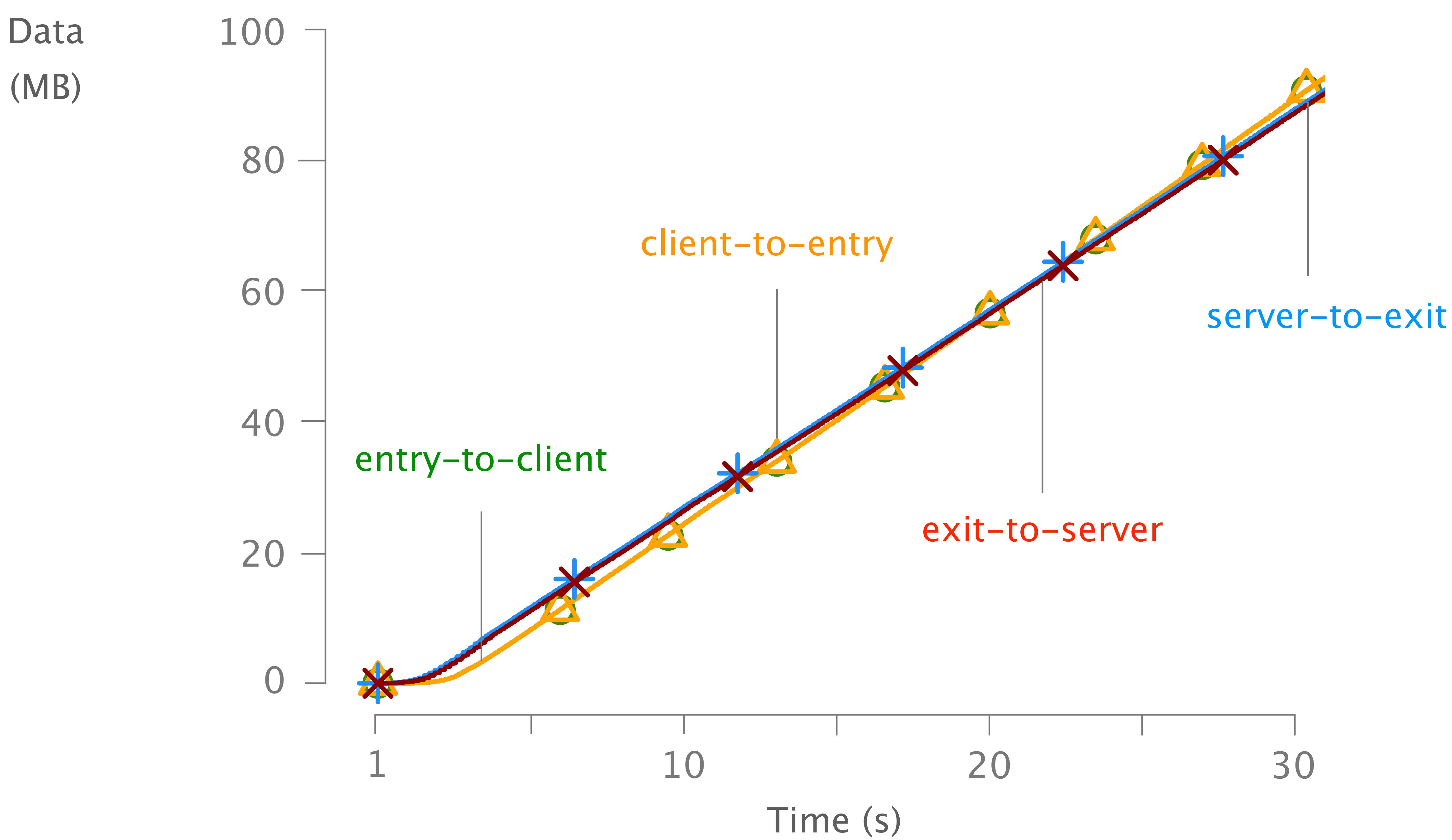


# We analyzed the evolution of the data sent & acknowledged, in each direction




Data
(MB)
100
80
60
40
20
0
client–to–entry
entry–to–client
exit–to–server
server–to–exit
1
10
20
30
Time (s)

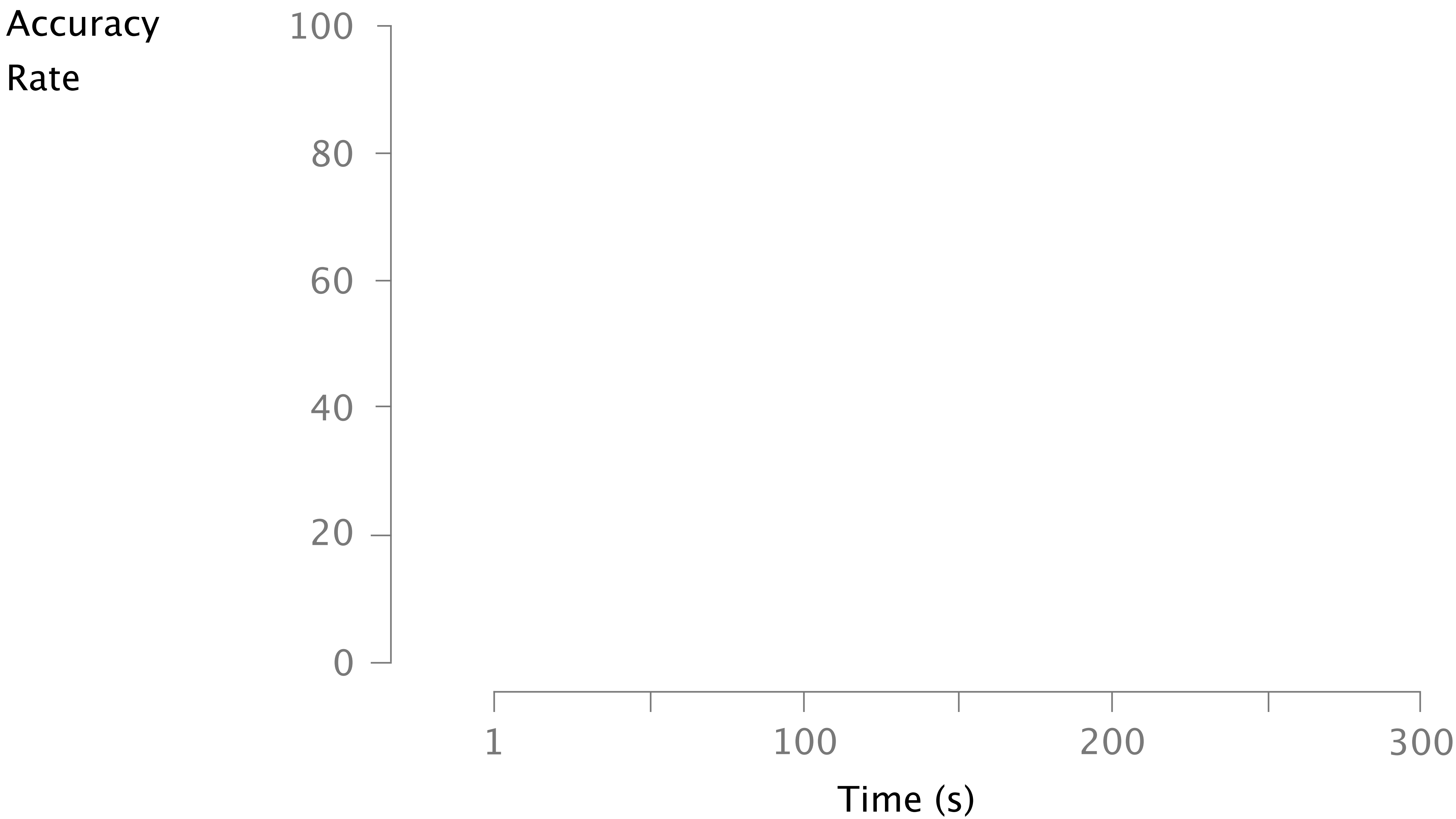
# Data sent in one direction is nearly identical to data acknowledged in the other



# After 5 min, we were able to deanonymize ~95% of the pairs—with *no* false positives

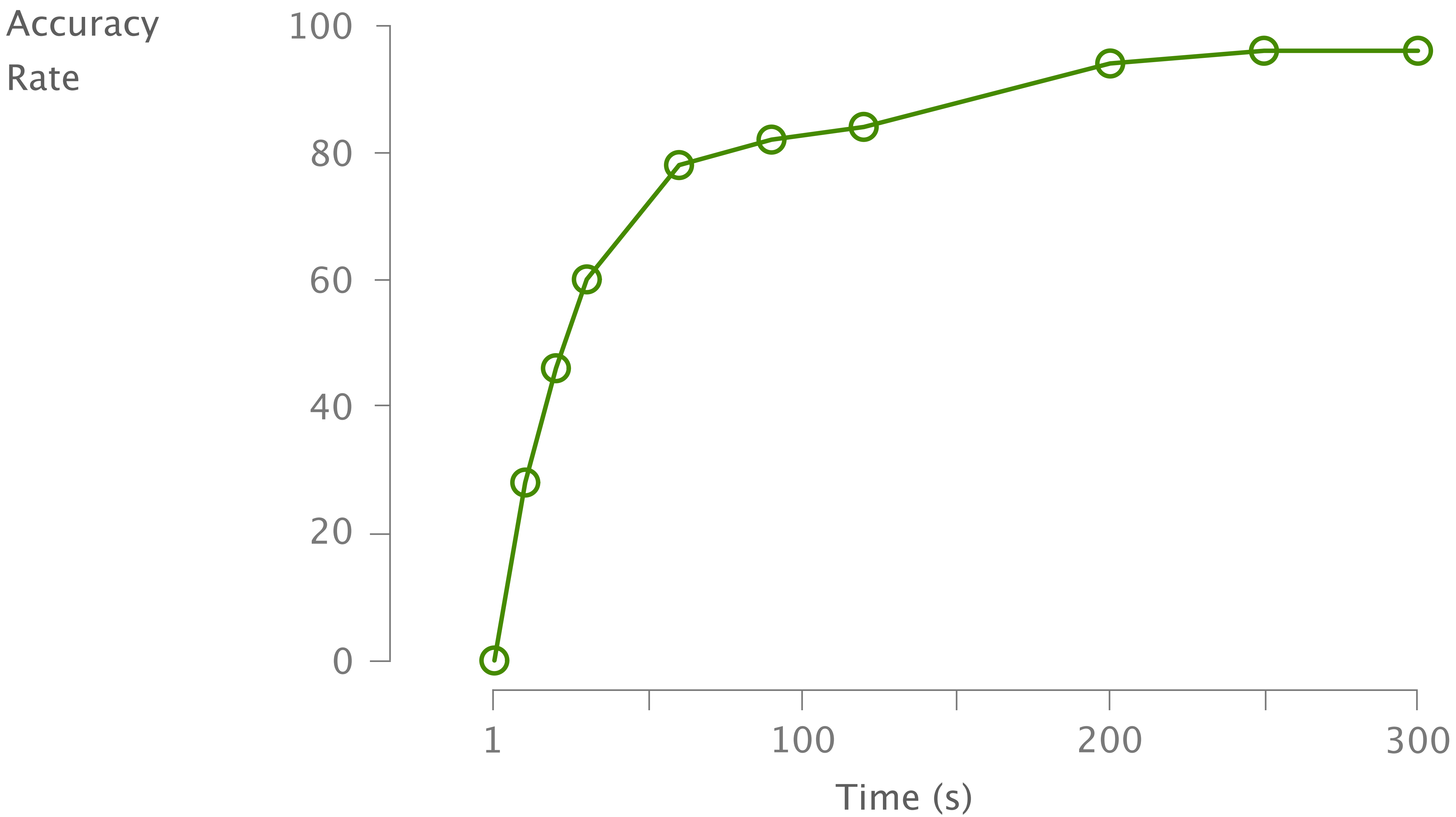
| | client ACK & server ACK | client ACK & server data | client data & server ACK | client data & server data |
|---|---|---|---|---|
| detection rate | 96 % | 94 % | 96 % | 94 % |
| false negative | 4 % | 6 % | 4 % | 6 % |
| false positive | 0 % | 0 % | 0 % | 0 % |

# Detection accuracy quickly increases with time



Detection accuracy quickly increases with time,

reaching 80% within only a minute

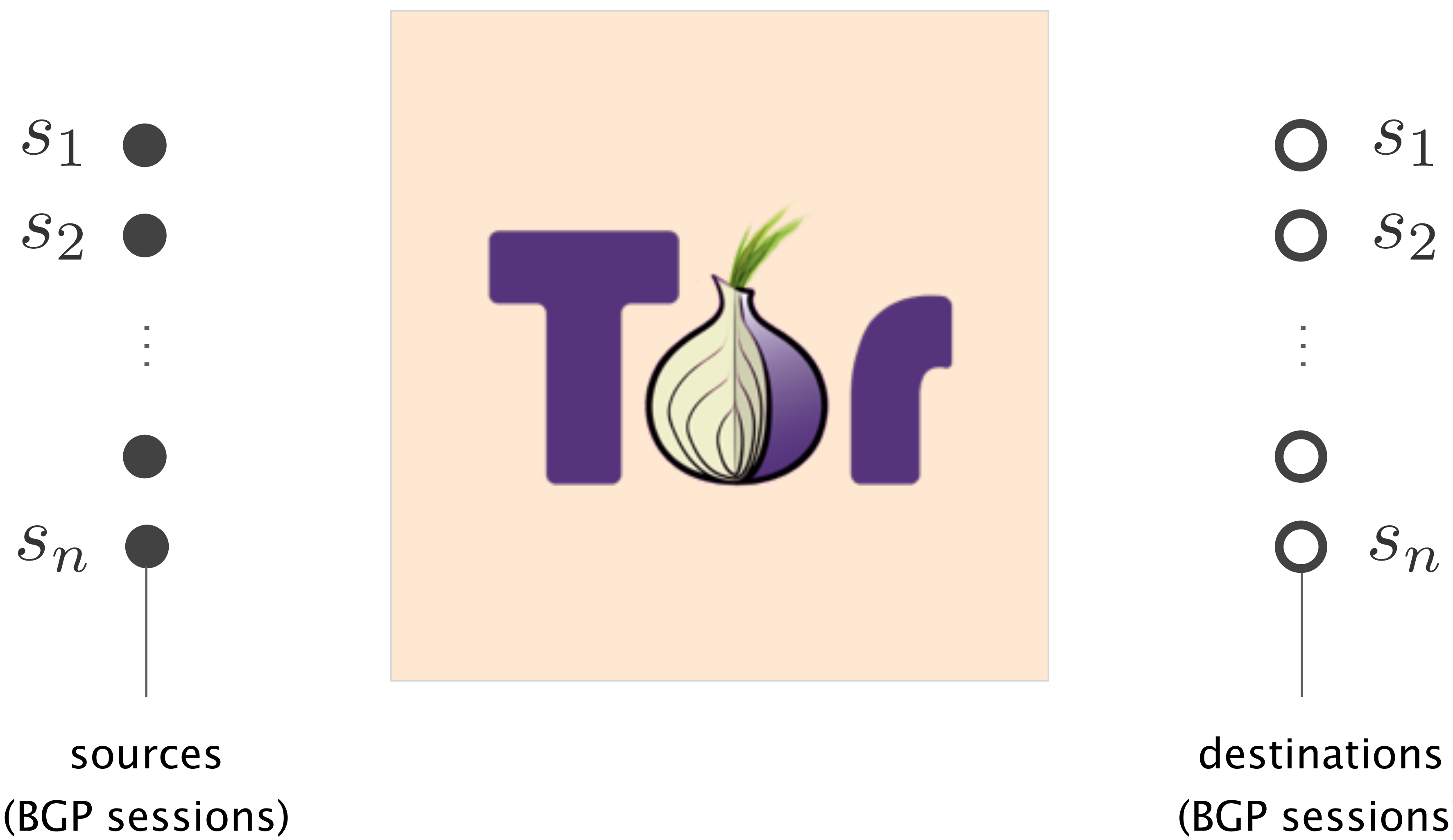


#2. Churn significantly increases the number of compromising ASes

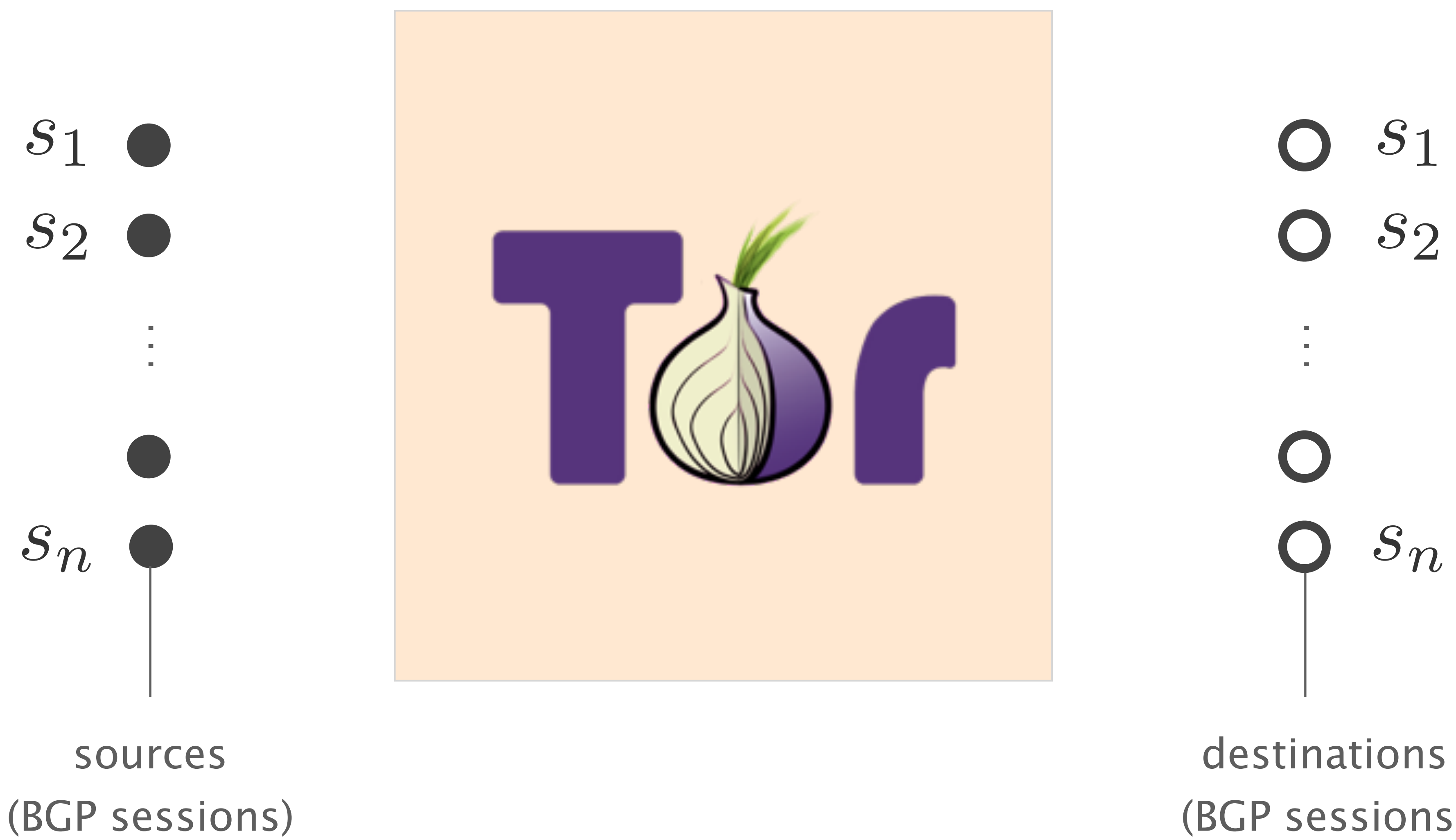
# We measured the effect of churn by collecting BGP updates for 1 month (Jan 15)

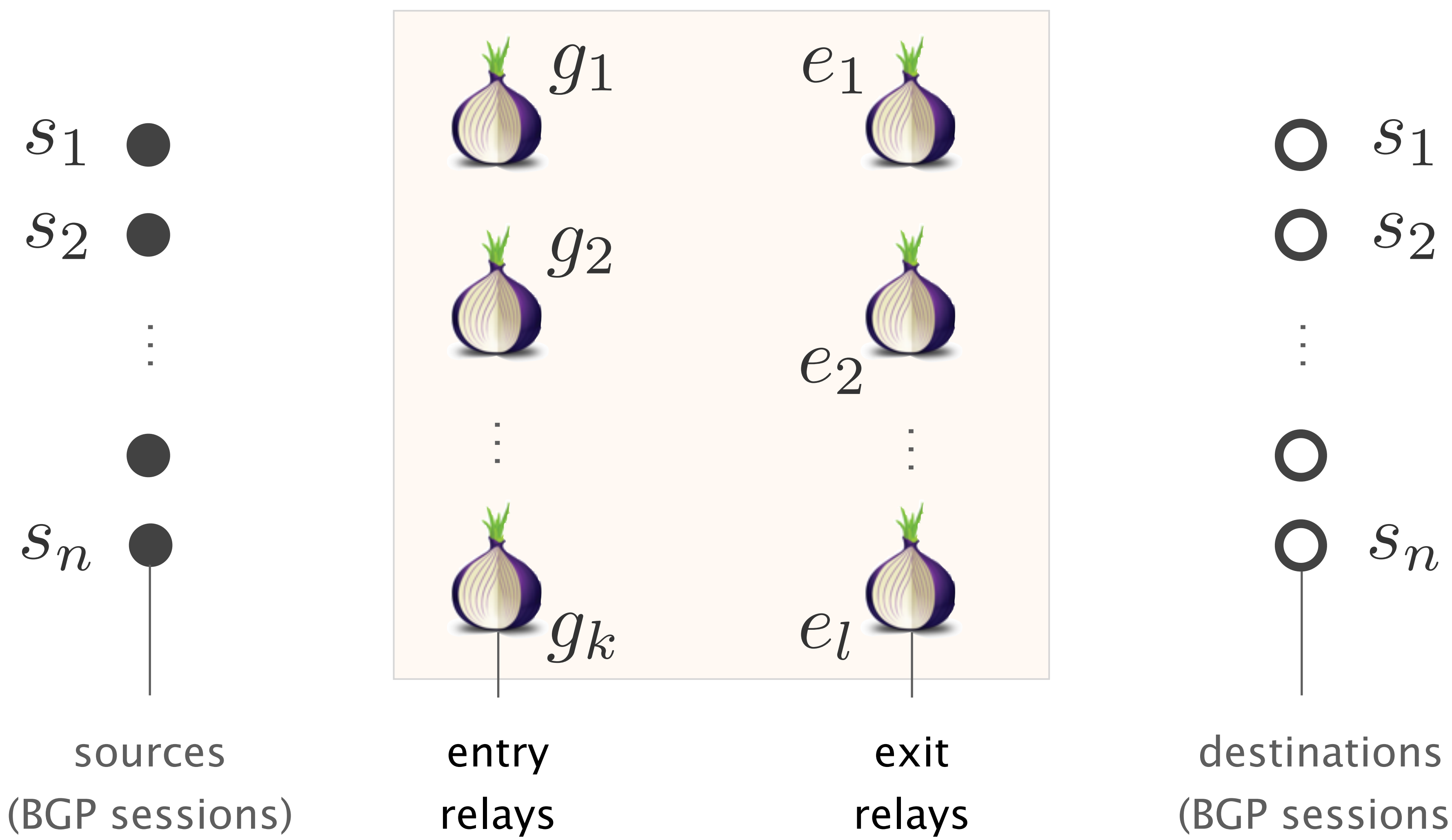
| | |
|---|---|
| # BGP sessions<br>(6 RIPE RIS collectors) | 250+ |
| # BGP prefixes | 550k |
| # BGP updates<br>announcements/withdraws | 612+ millions |

# We considered each BGP session as a Tor user or destination

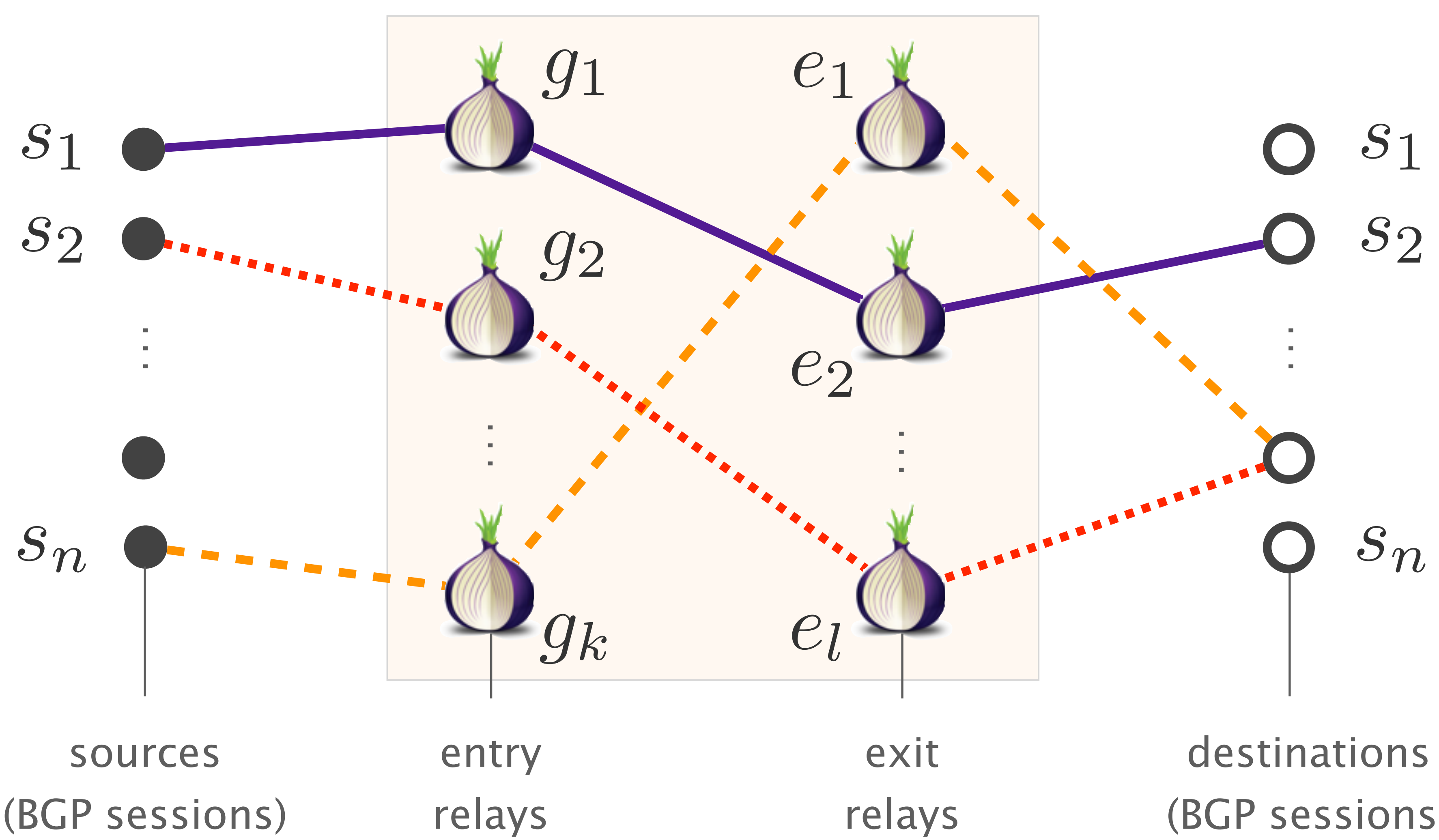


On each session, we computed the ASes used to reach each entry and exit relays

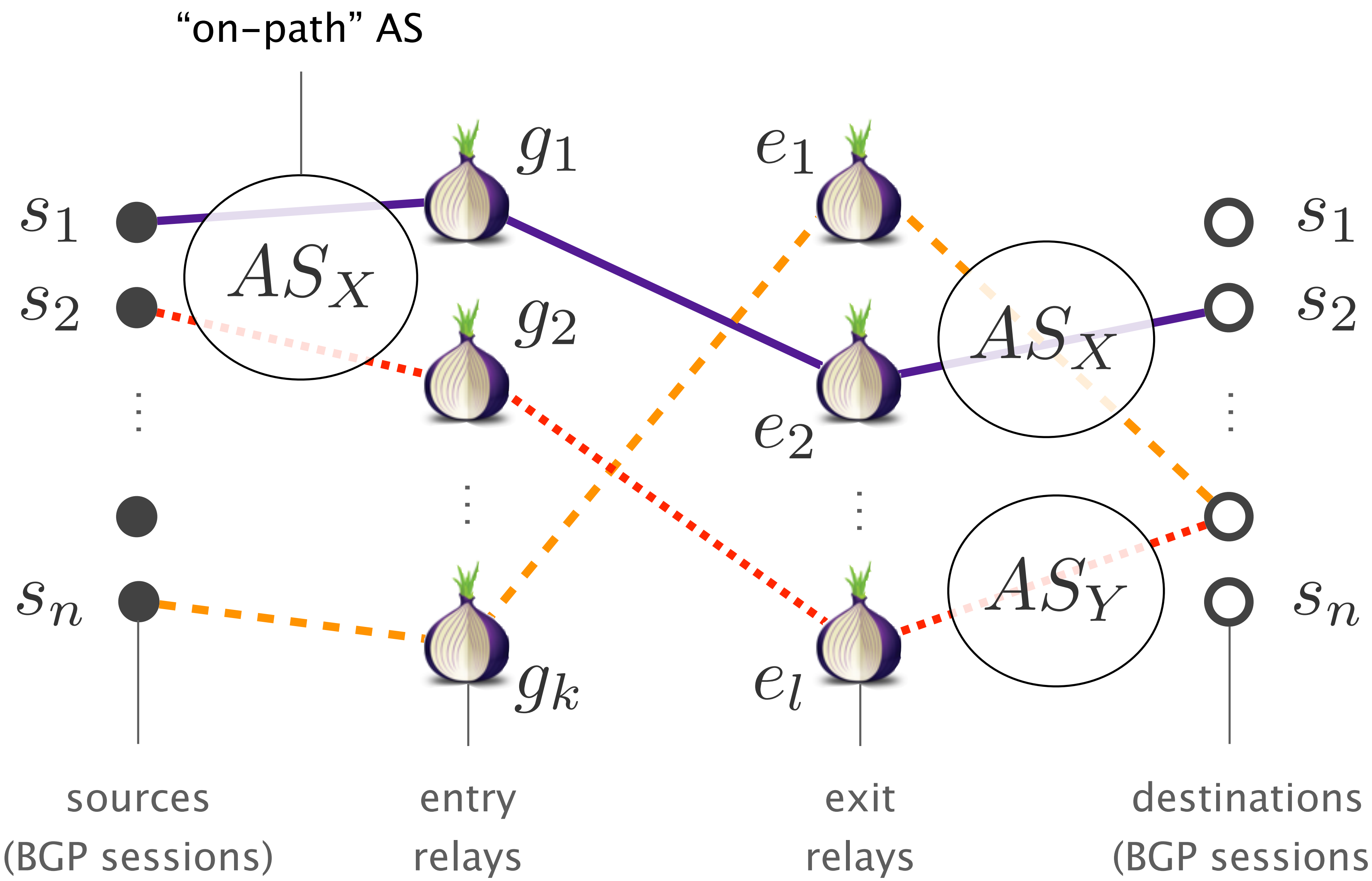




$s_1$
$s_2$
$s_n$
$g_1$
$g_2$
$g_k$
$e_1$
$e_2$
$e_l$
$s_1$
$s_2$
$s_n$
sources
(BGP sessions)
entry
relays
exit
relays
destinations
(BGP sessions)

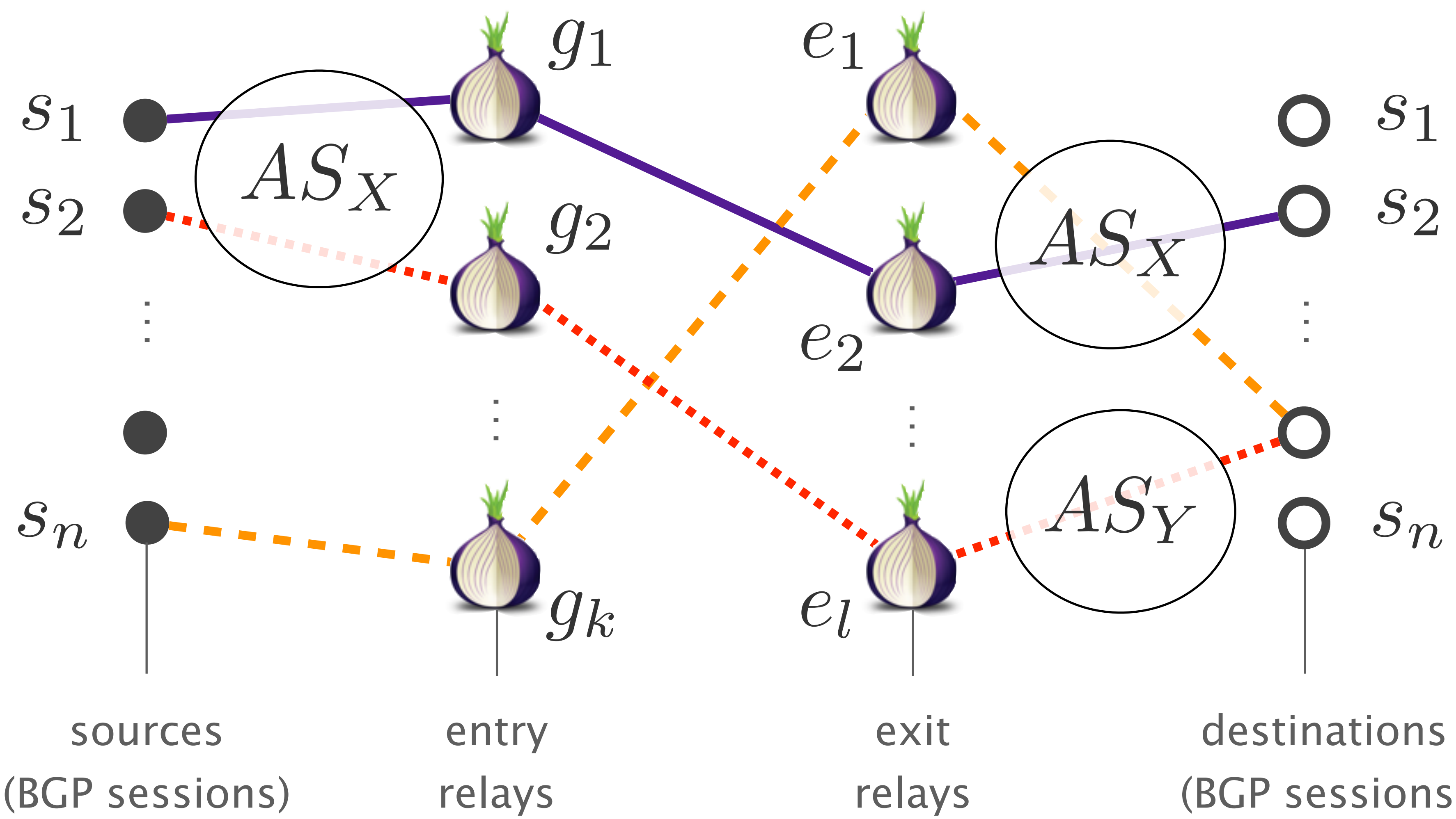


$s_1$
$s_2$
$s_n$
$g_1$
$g_2$
$g_k$
$e_1$
$e_2$
$e_l$
$s_1$
$s_2$
$s_n$
sources (BGP sessions)
entry relays
exit relays
destinations (BGP sessions)

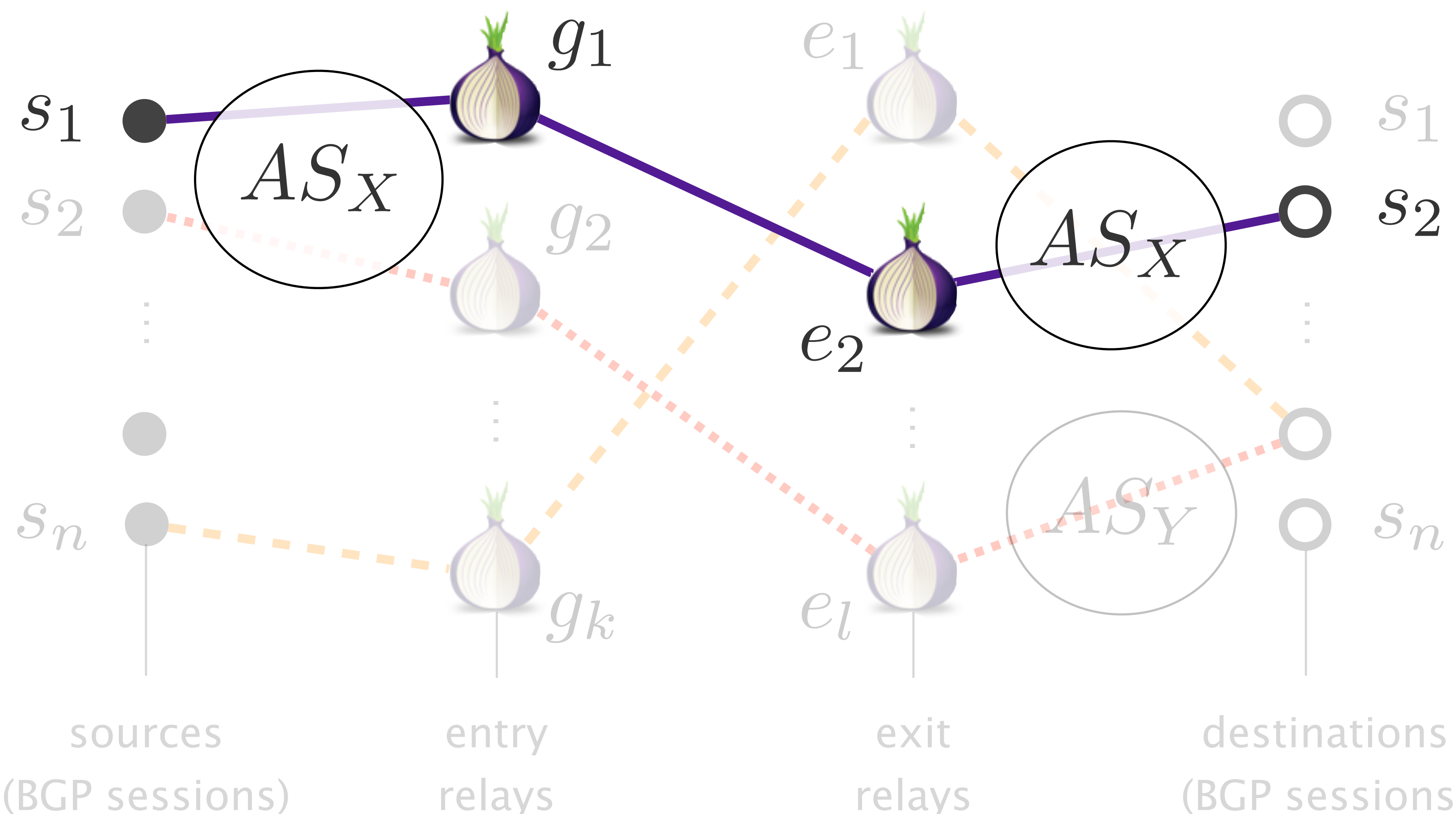


“on-path” AS
$s_1$
$s_2$
$s_n$
$AS_X$
$g_1$
$g_2$
$g_k$
$e_1$
$e_2$
$e_l$
$AS_X$
$AS_Y$
$s_1$
$s_2$
$s_n$
sources
(BGP sessions)
entry
relays
exit
relays
destinations
(BGP sessions)

An AS is compromising when it ends up simultaneously on a (src, entry) and (exit, dest) path



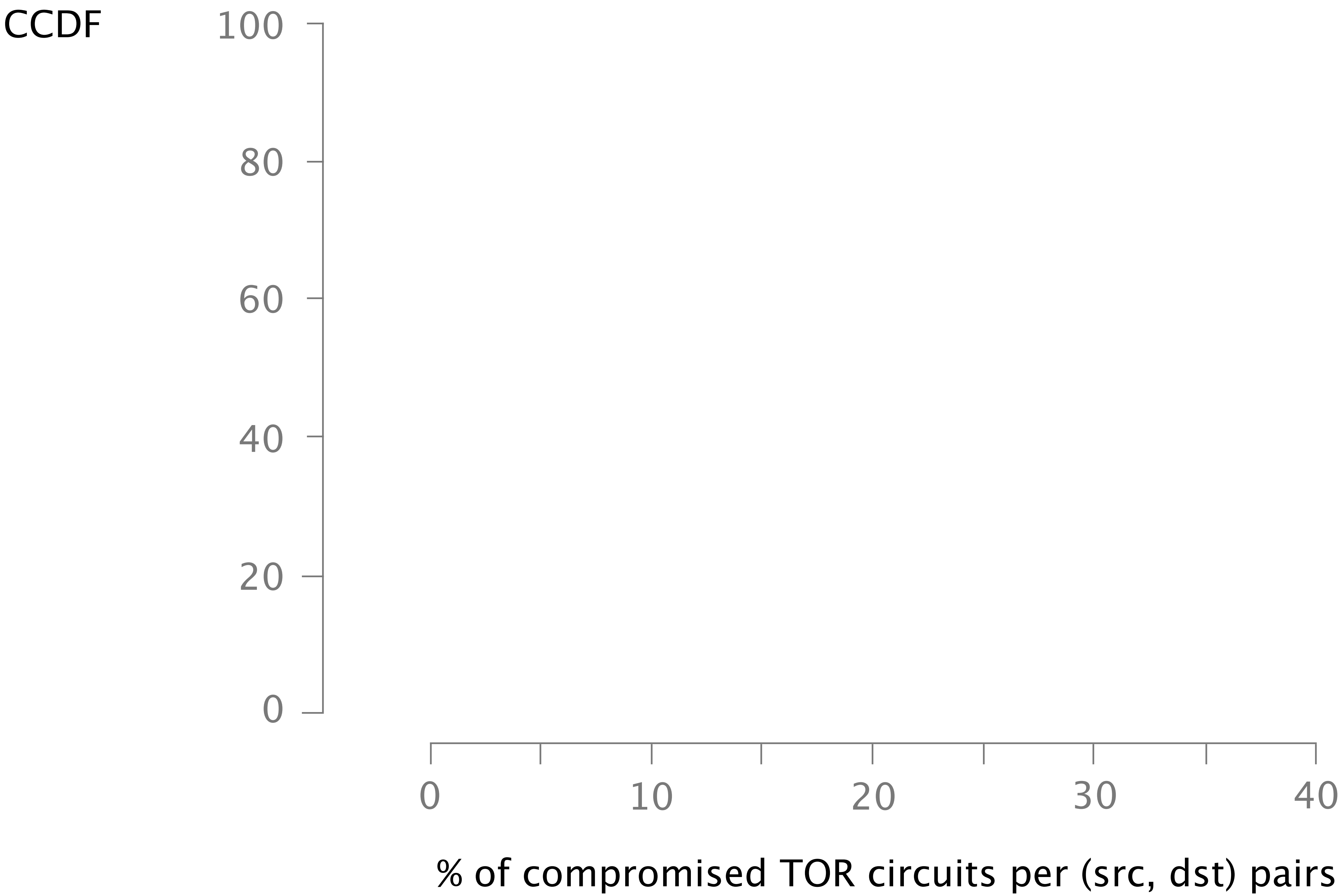
ASX is compromising for the TOR circuit (g1, e2) and (s1,s2)

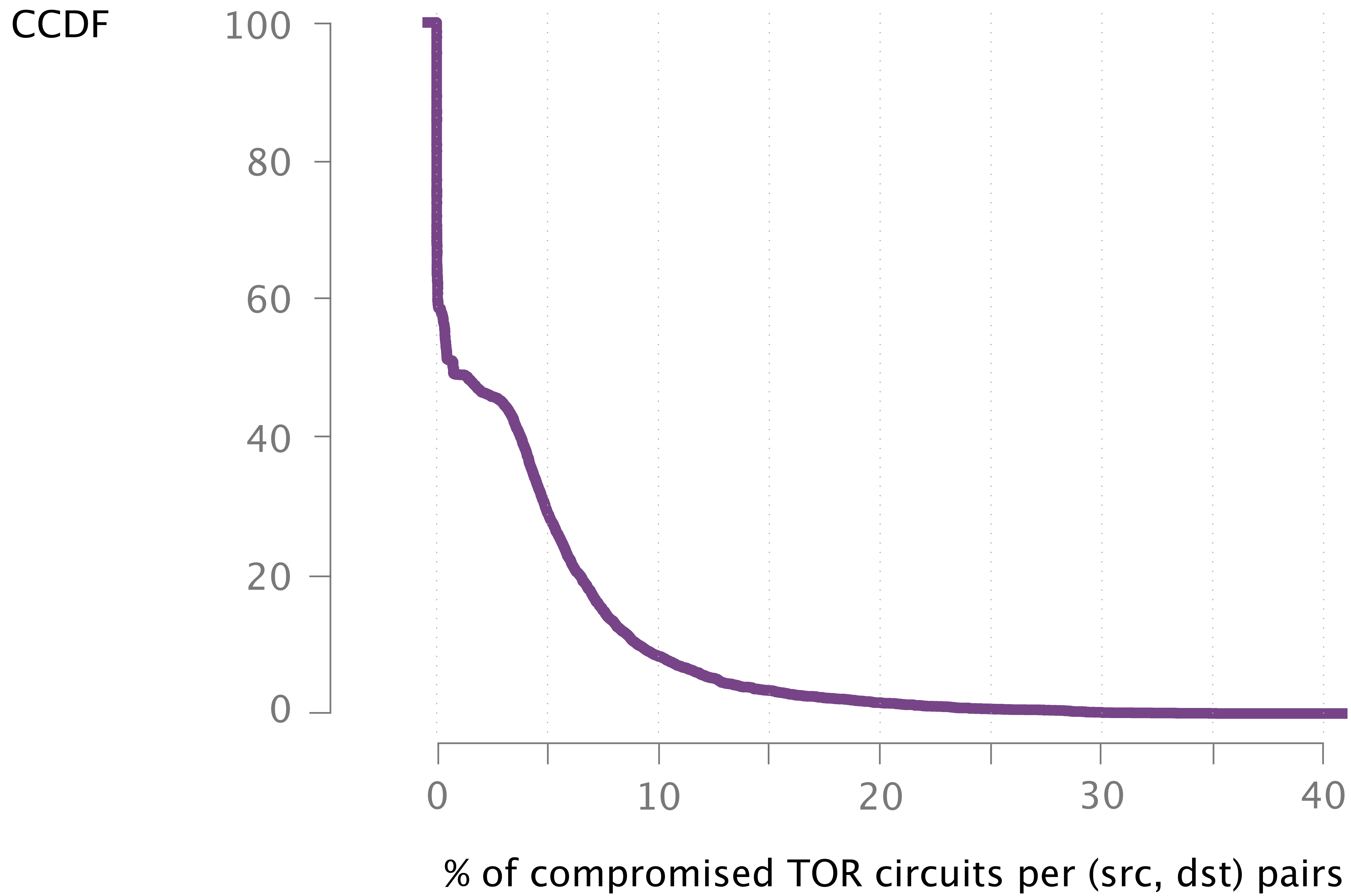


Without considering churn…

How many ASes are compromising,

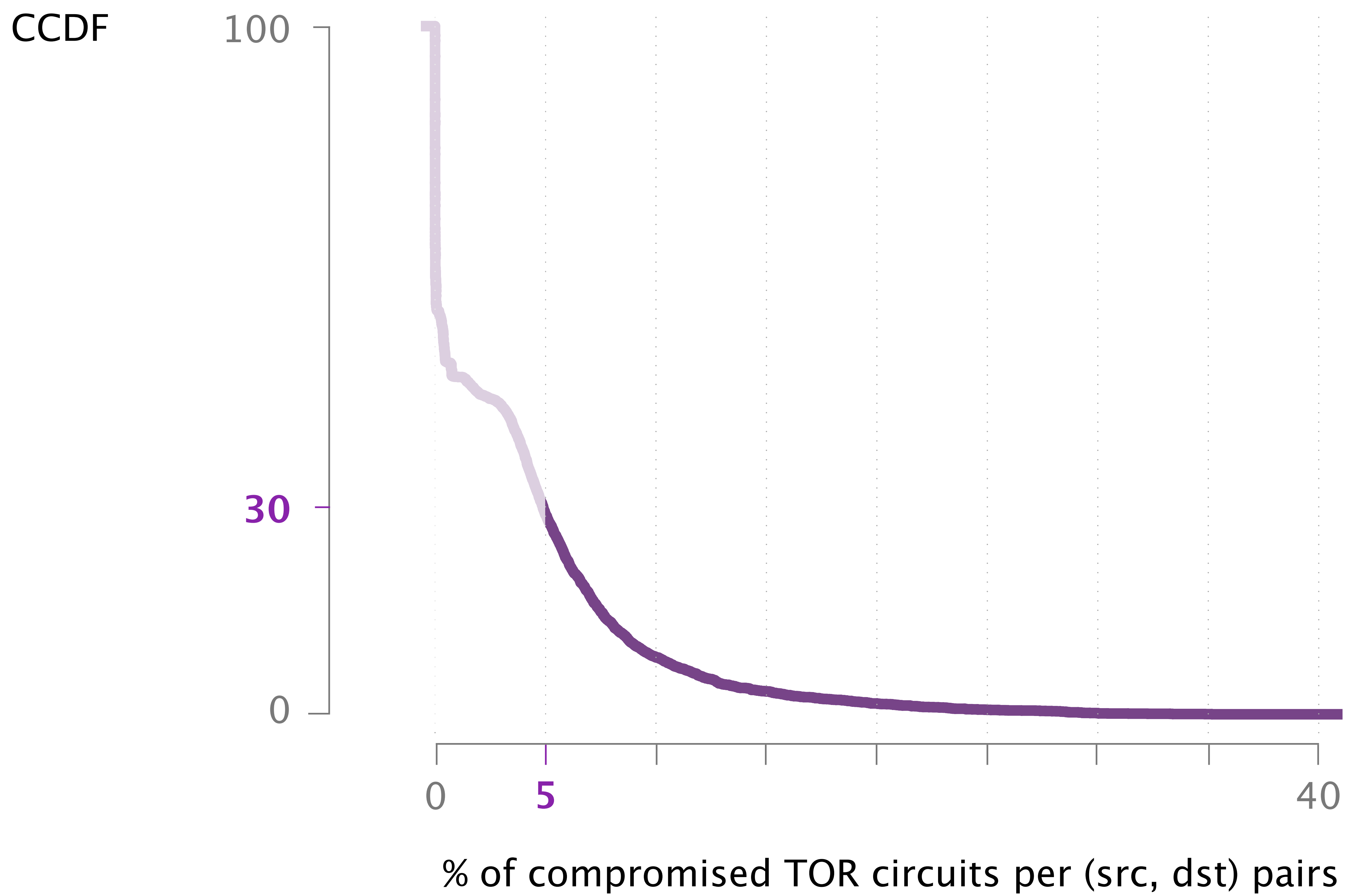
and for how many Tor circuits?


CCDF
100
80
60
40
20
0
0
10
20
30
40
% of compromised TOR circuits per (src, dst) pairs



CCDF
100
80
60
40
20
0
0
10
20
30
40
% of compromised TOR circuits per (src, dst) pairs

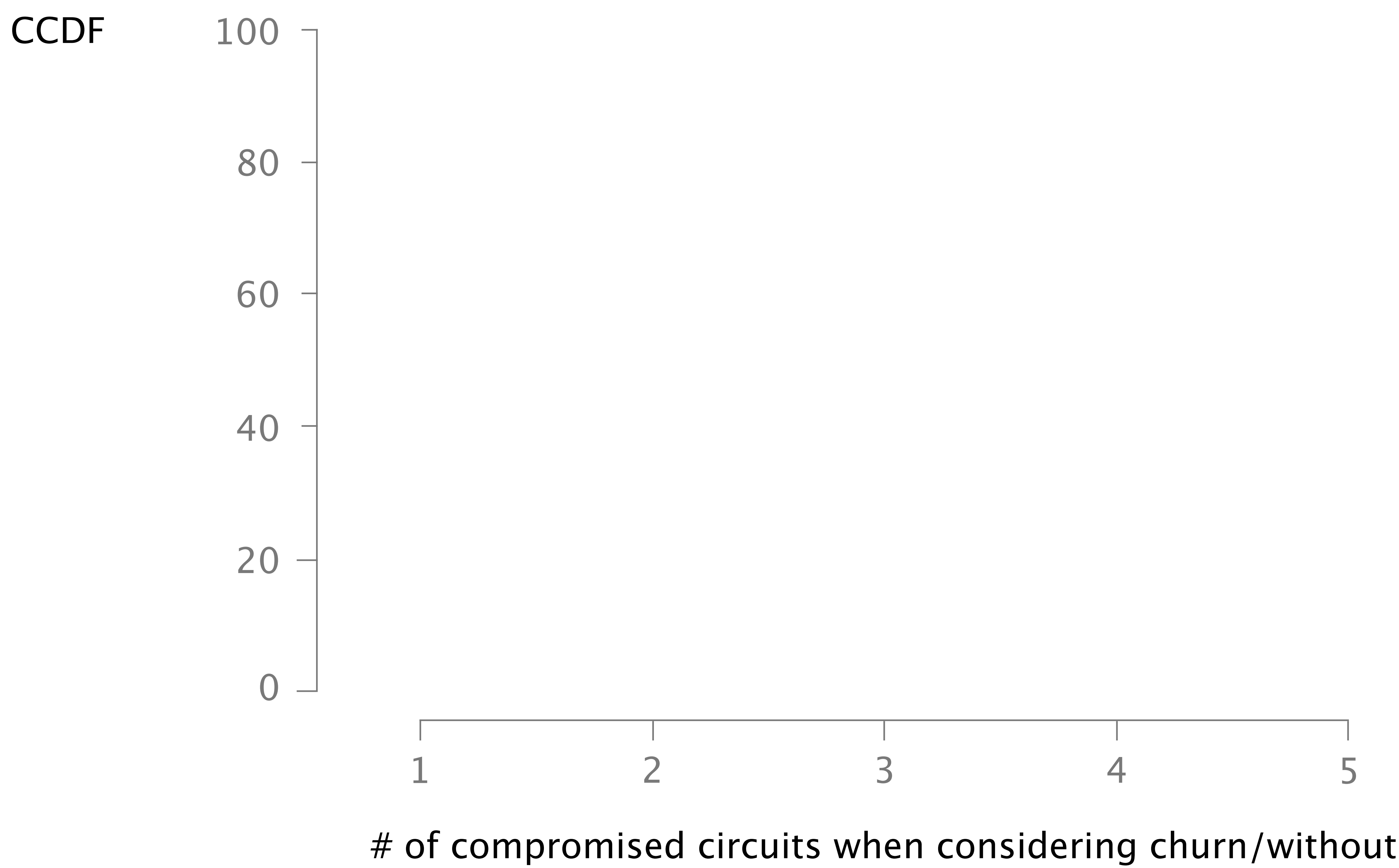
# 30% of the time, >5% of the circuits are compromised by at least 1 AS



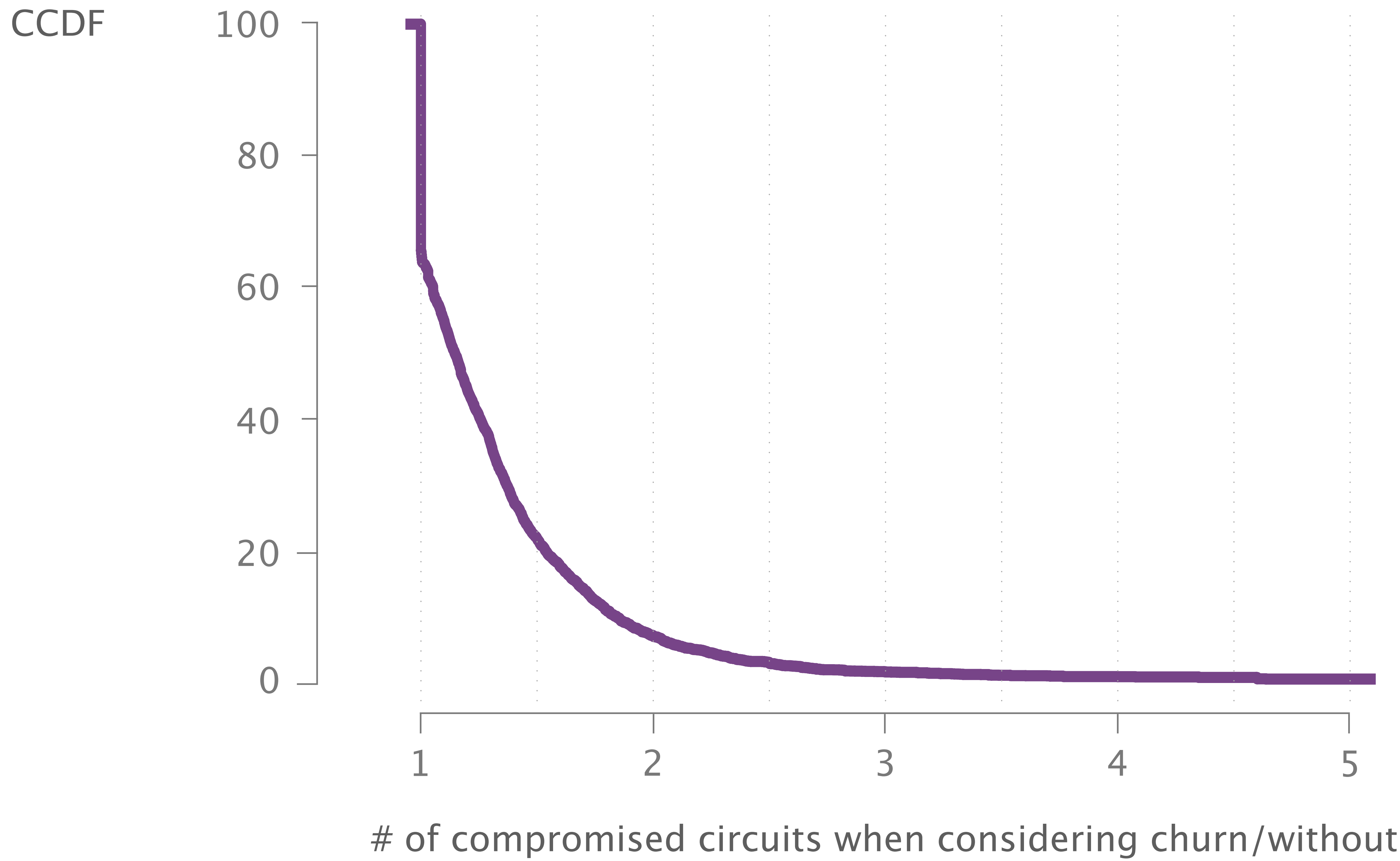
When considering churn…

How many more ASes are compromising,

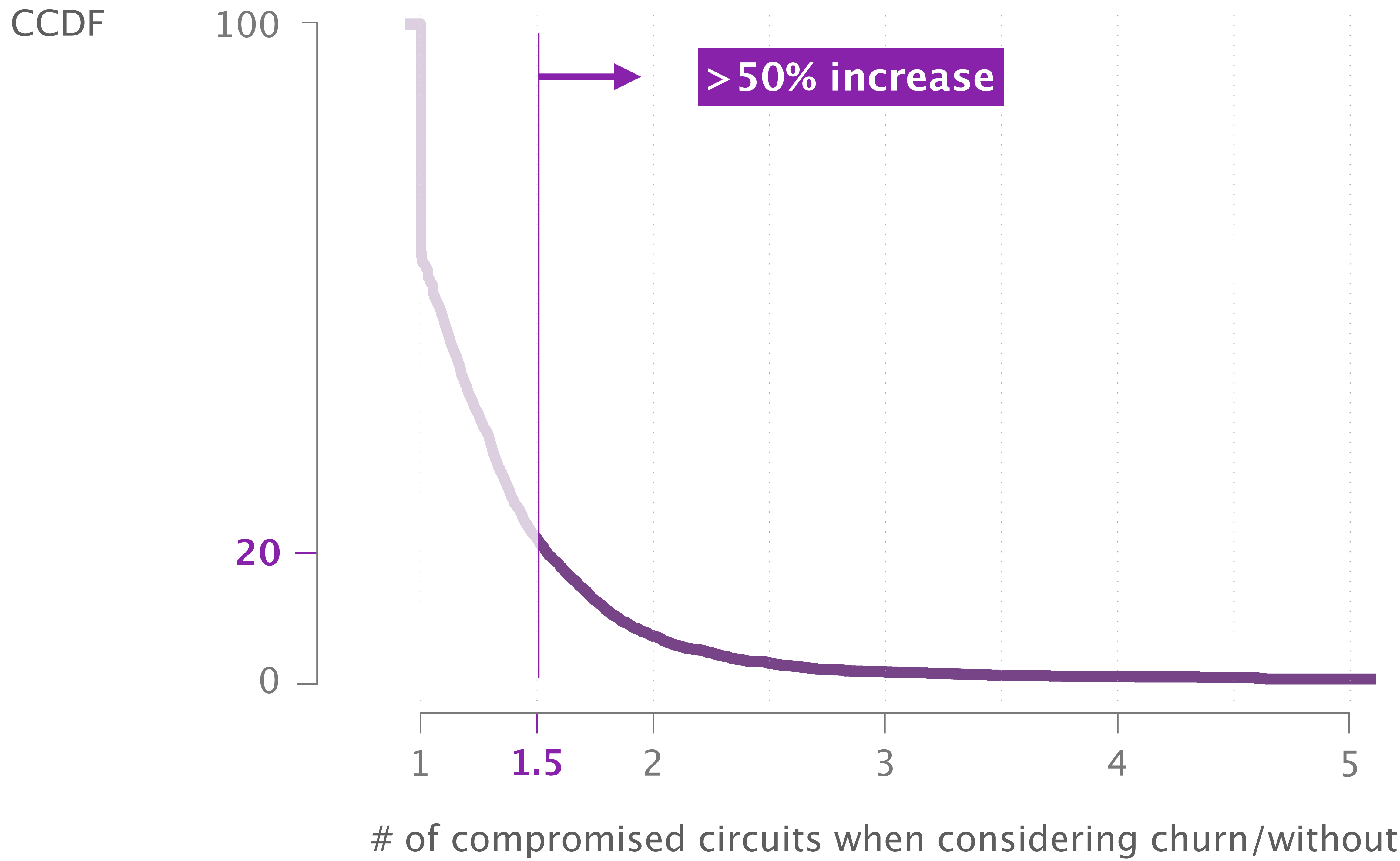
and for how many TOR circuits?


CCDF
100
80
60
40
20
0
1
2
3
4
5
# of compromised circuits when considering churn/without

# 60% of the pairs (src, dst) sees an increase of compromised circuits



20% of the pairs sees an increase of more than 50%!



# #3. BGP hijack works in the wild

We successfully performed a BGP attack on an existing Tor entry relay

# Our experiments did not compromise the privacy or safety of real Tor users

We attacked our own traffic

not actual user-generated Tor traffic

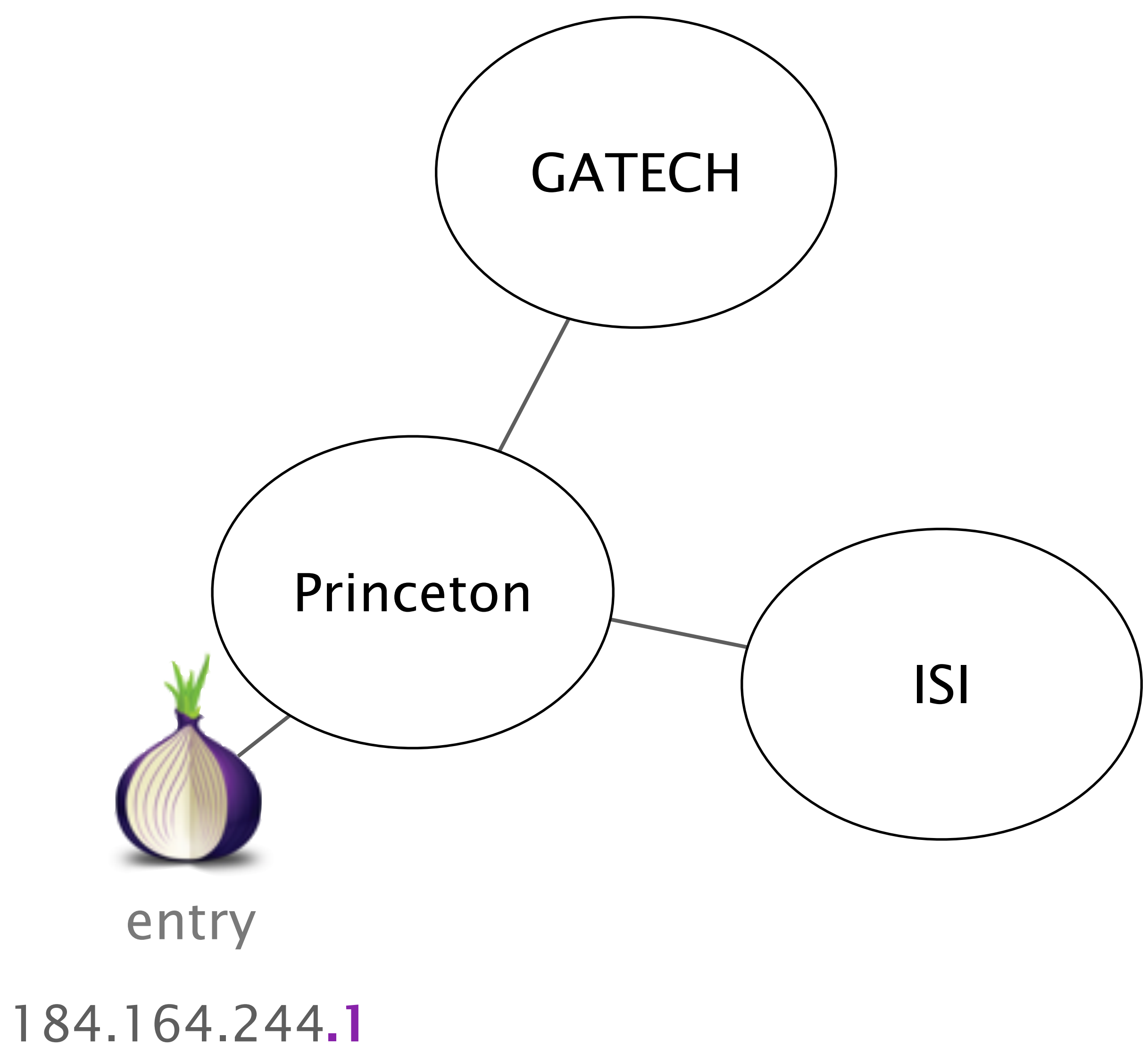
We attacked our own relay

hijacking our own IP prefix

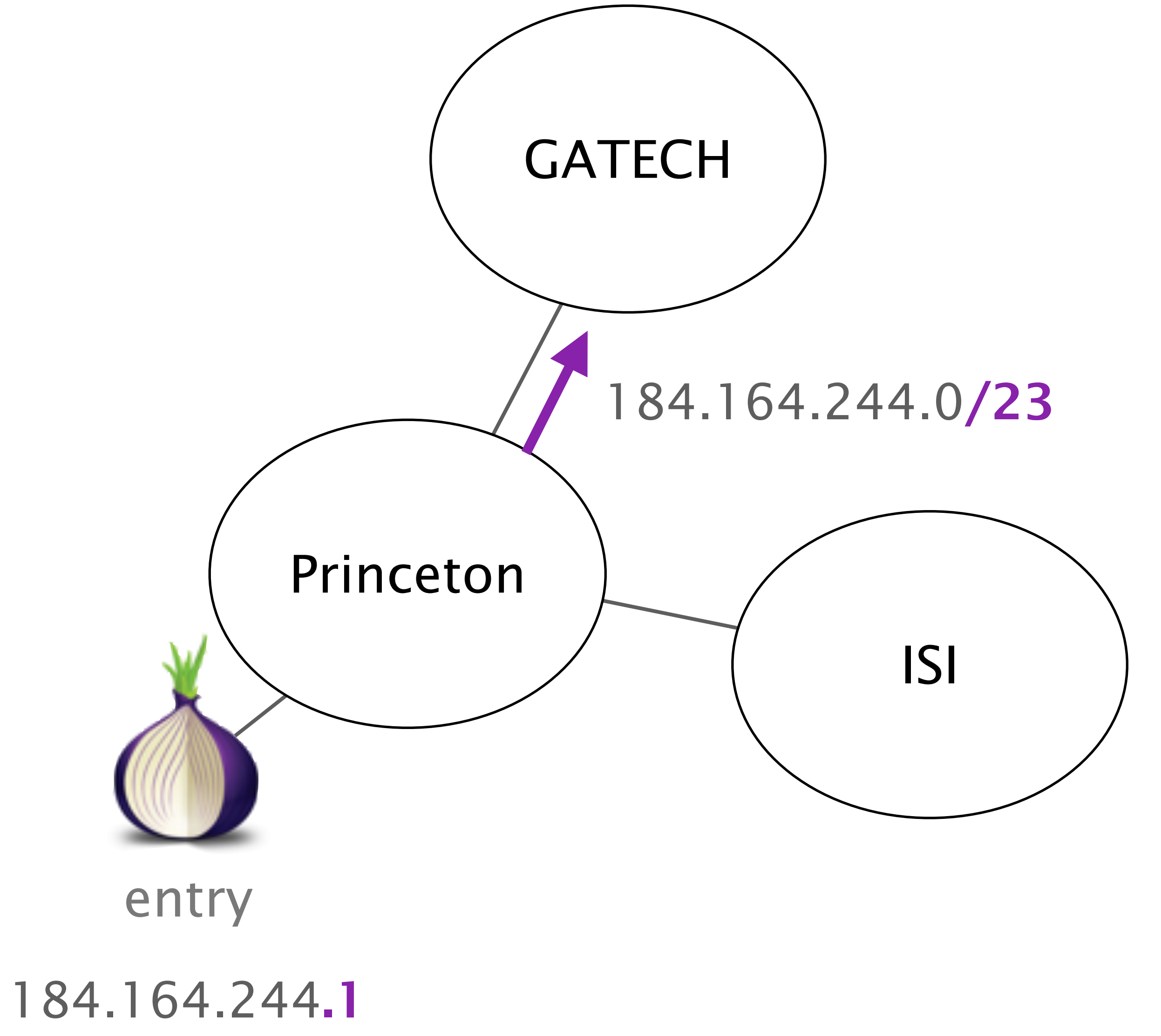
We firewalled our relay

dropping any traffic not generated by us

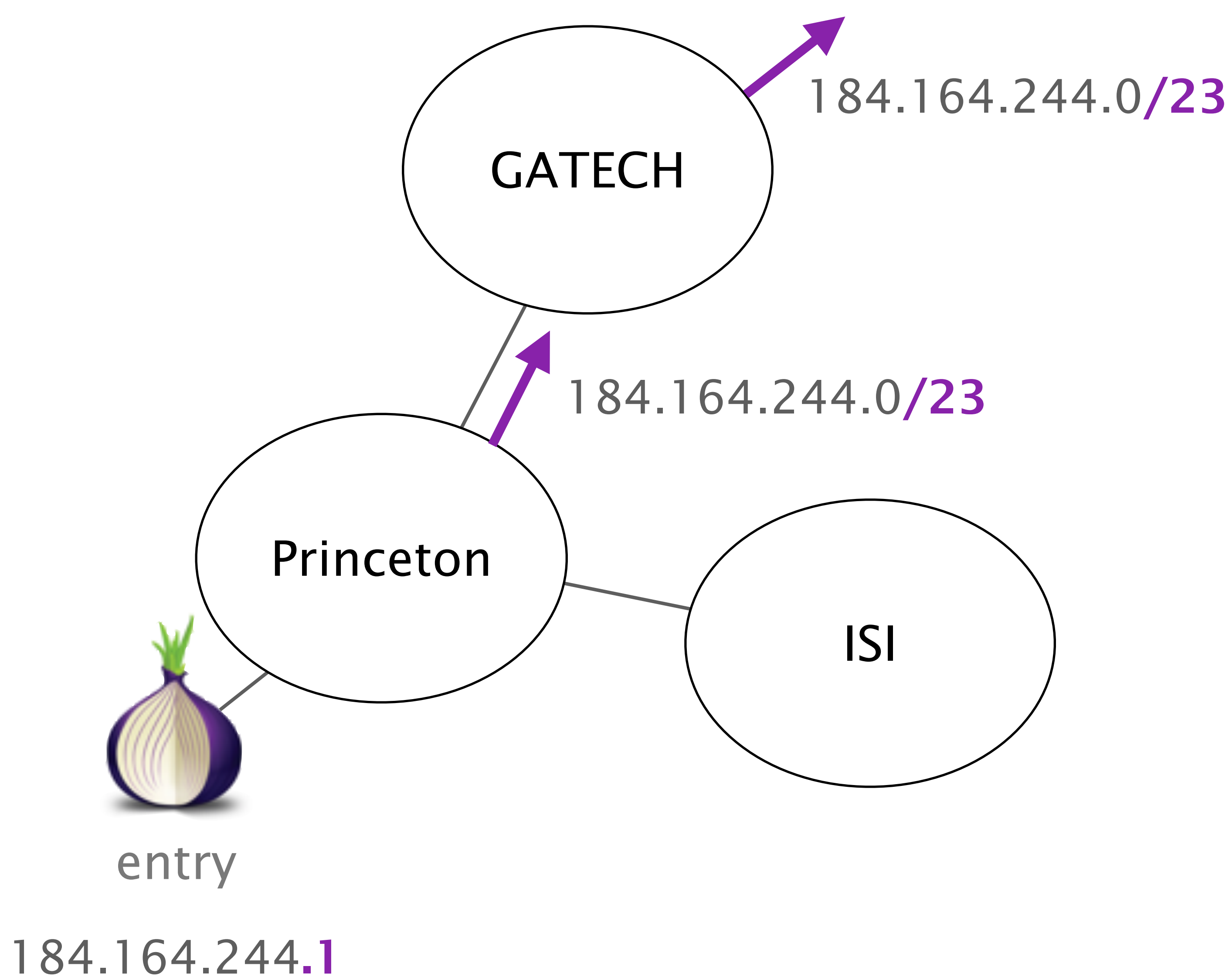
# We hosted an entry relay in Princeton



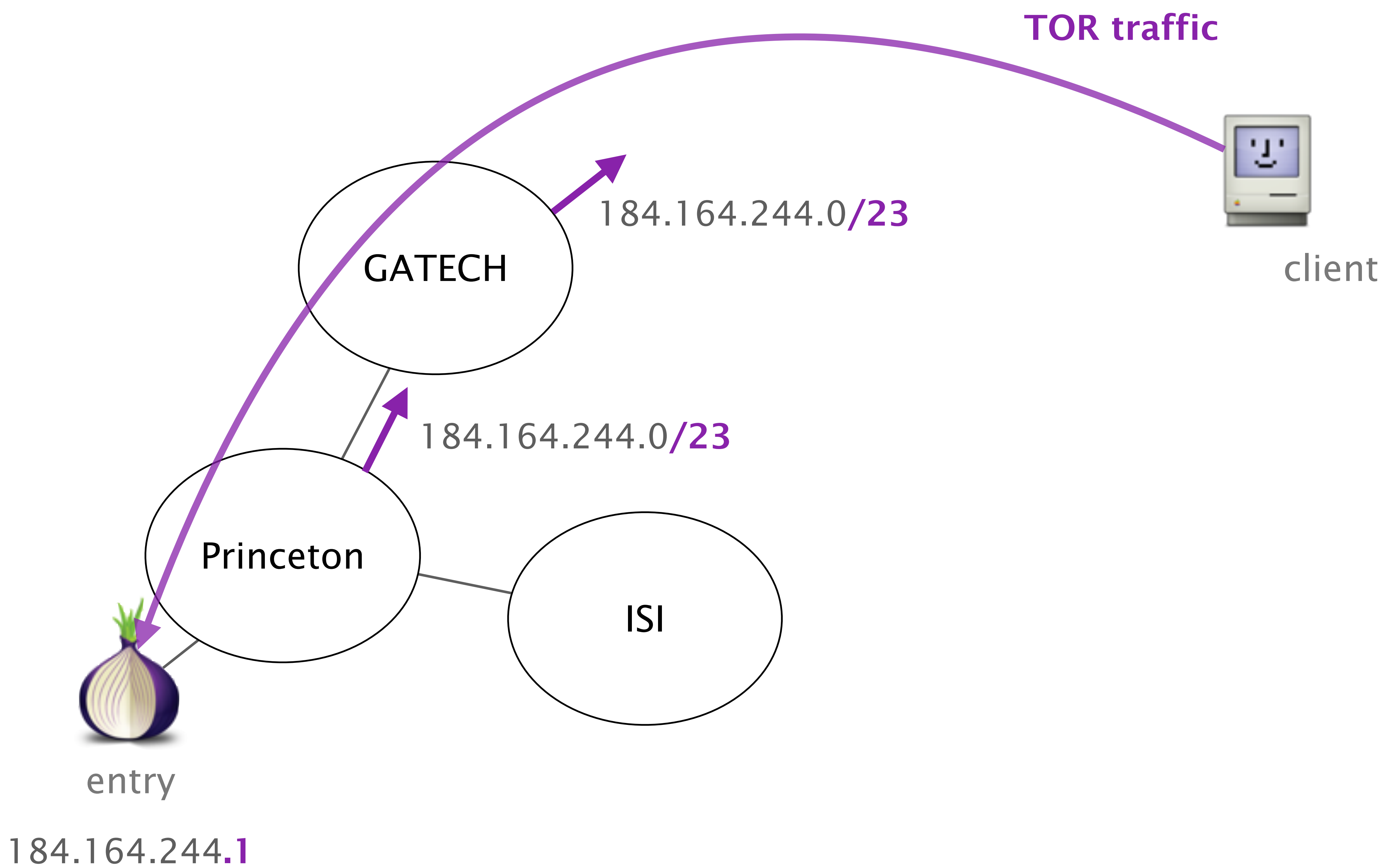
# We advertised the covering IP prefix via GATECH



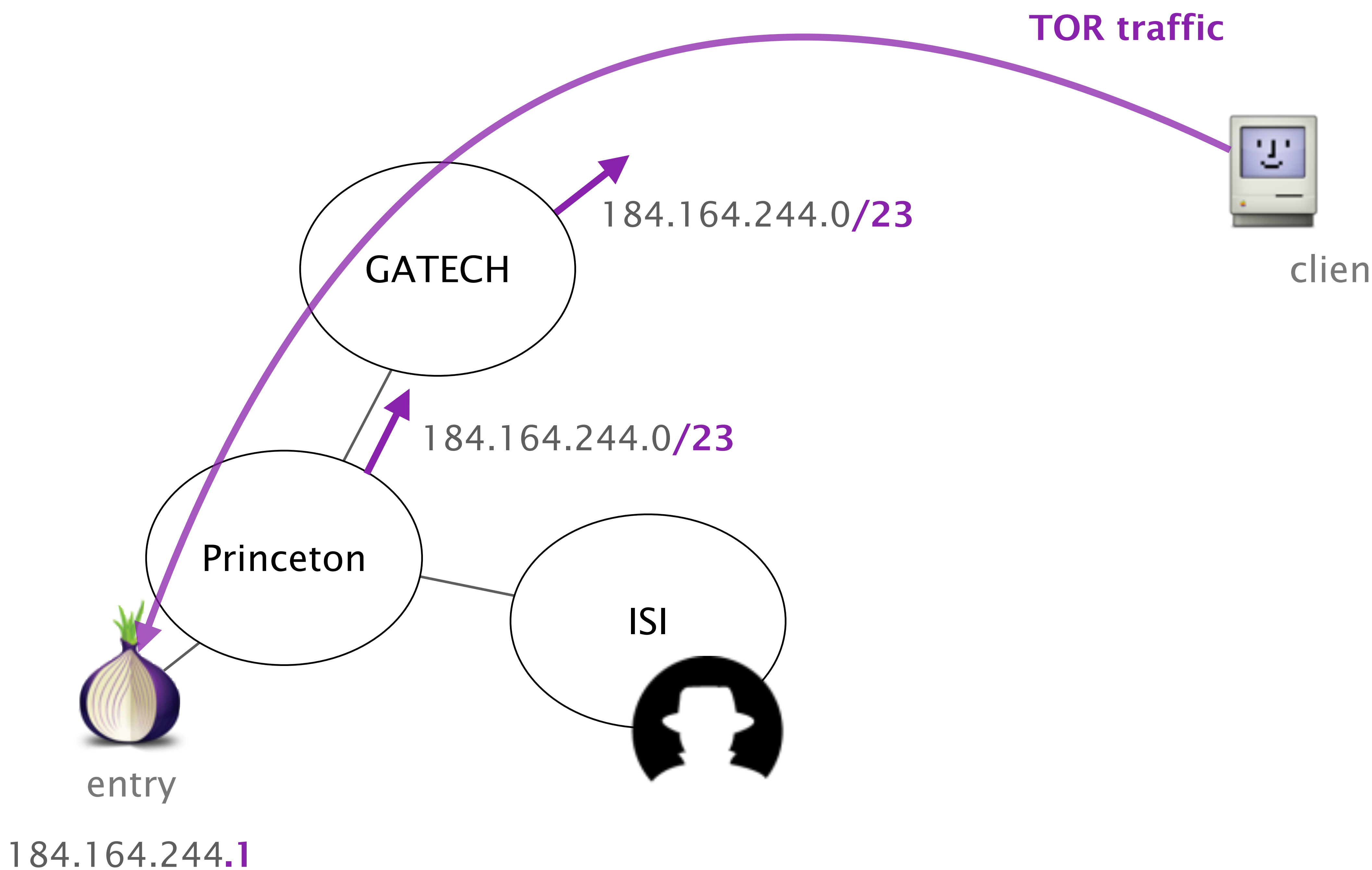
# GATECH relayed on prefix to the entire Internet



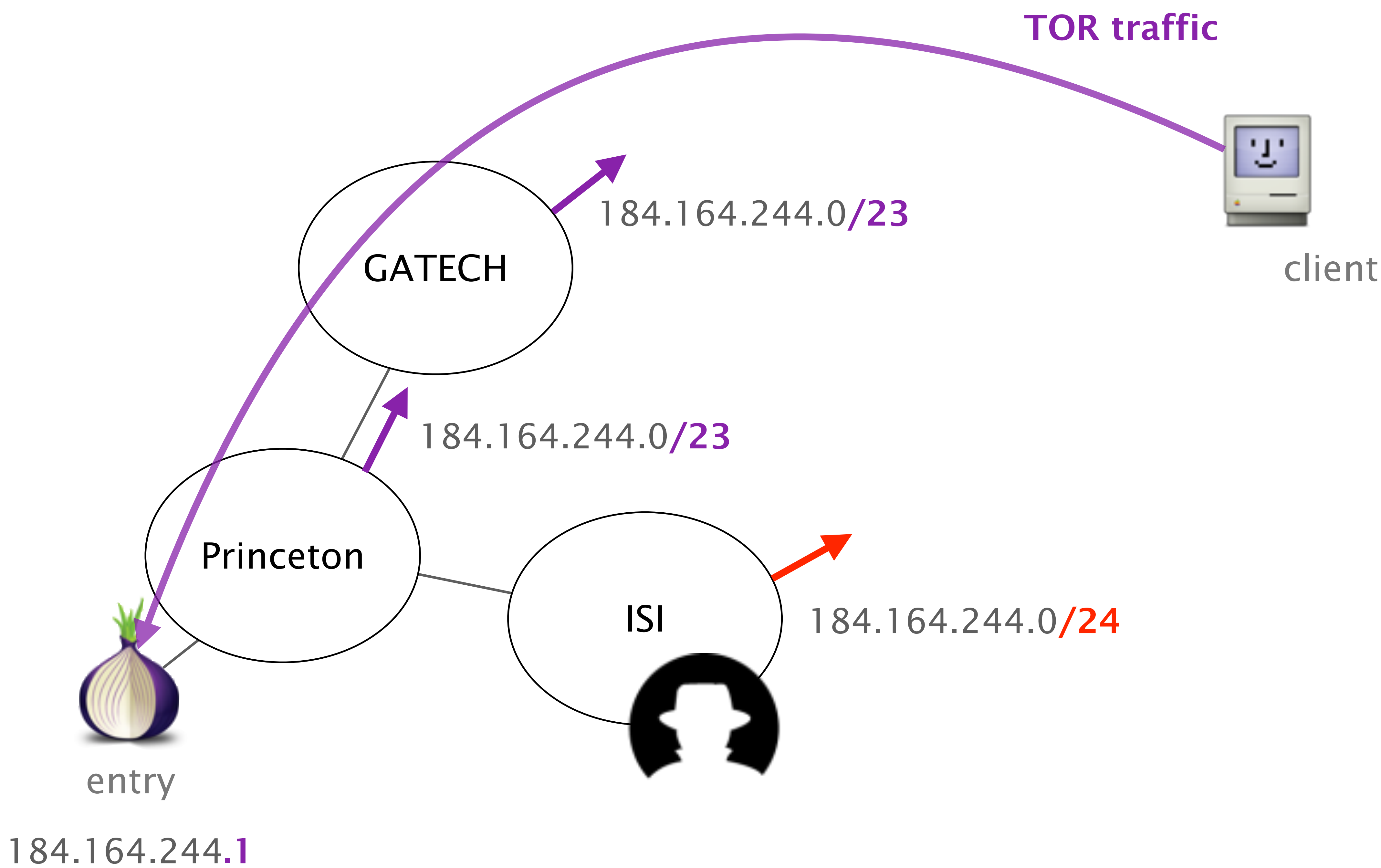
# Tor traffic started to flow



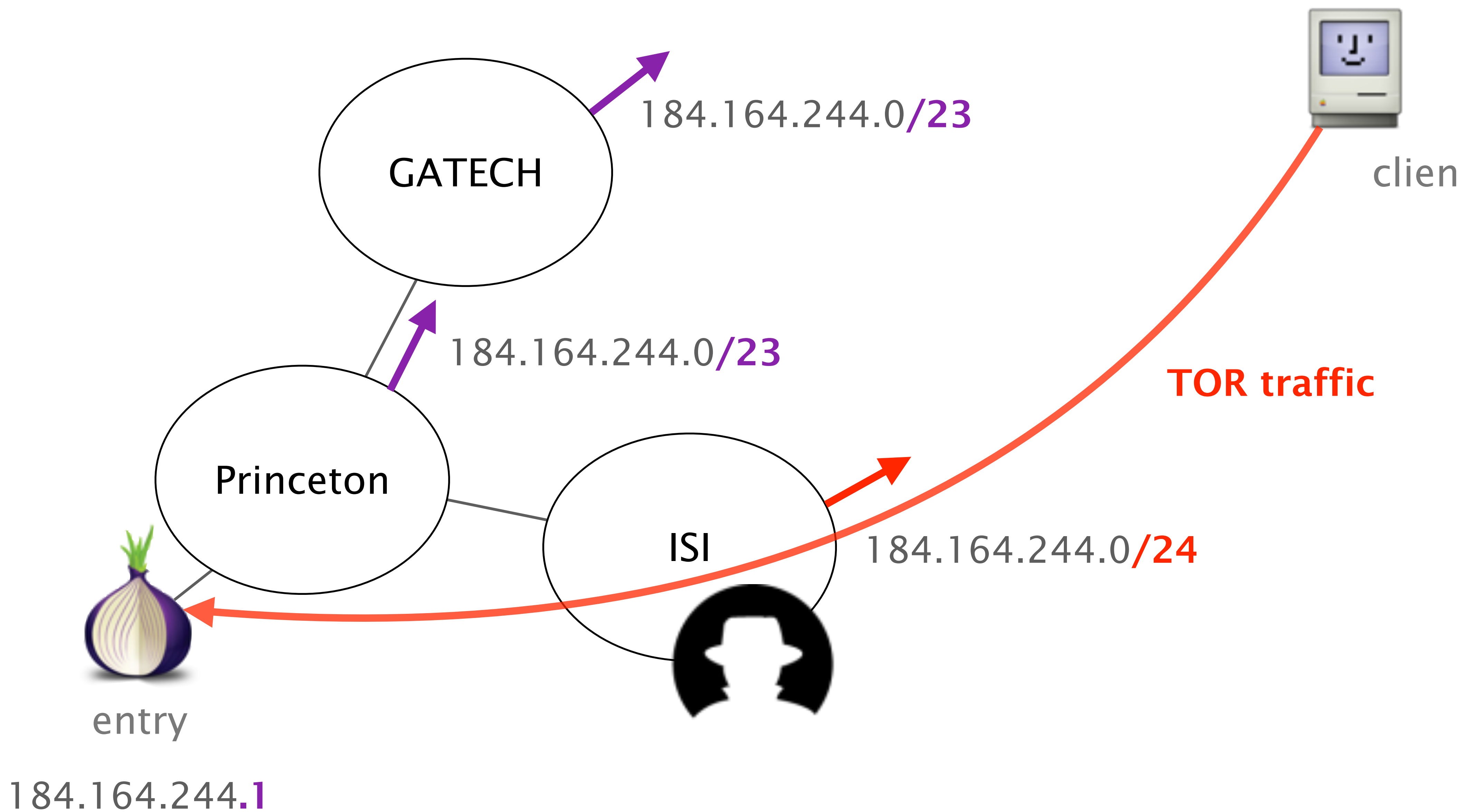
# After 5 minutes, we announced a more-specific prefix via ISI

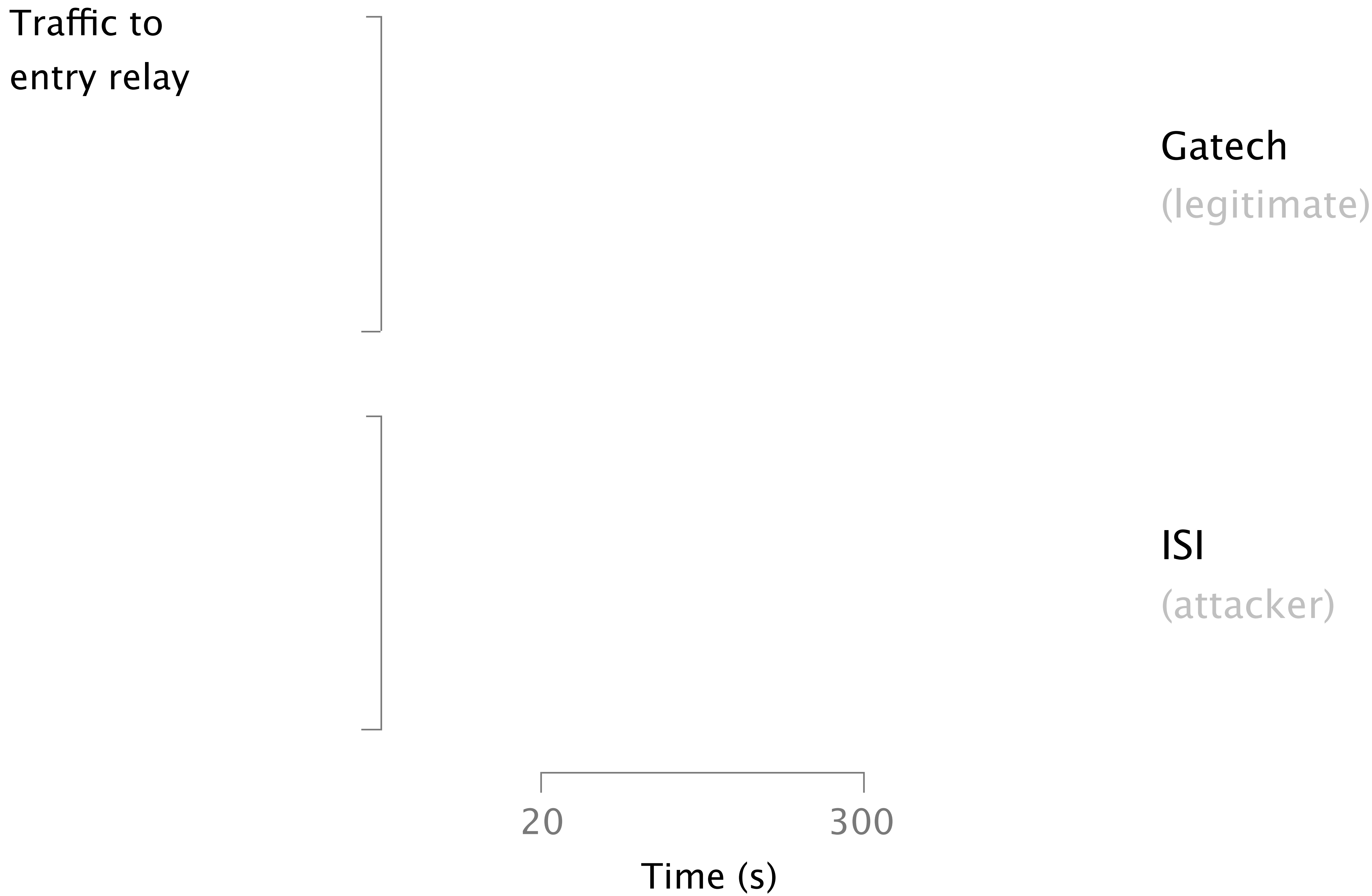


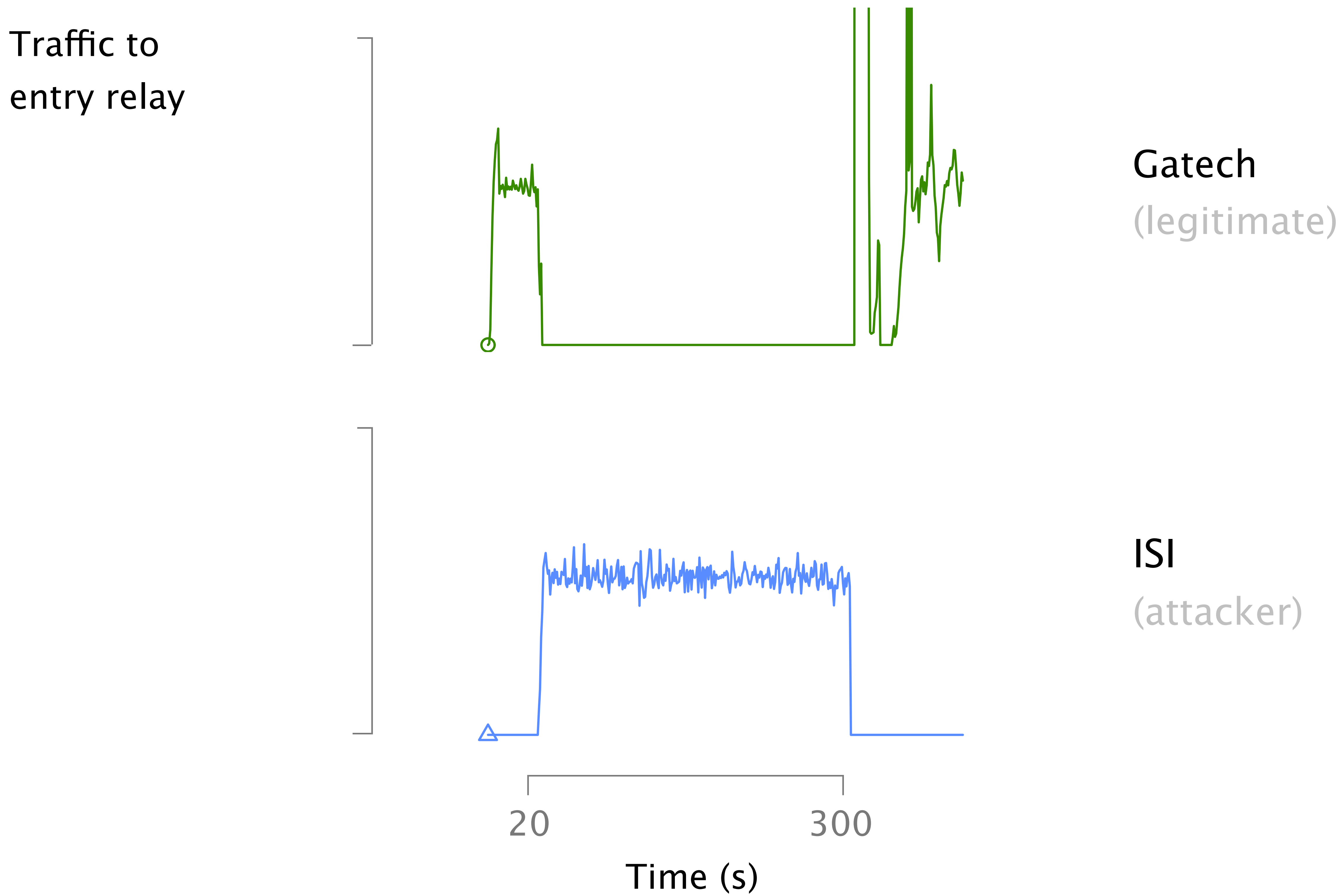
# After 20 sec, we announced a more-specific prefix via ISI

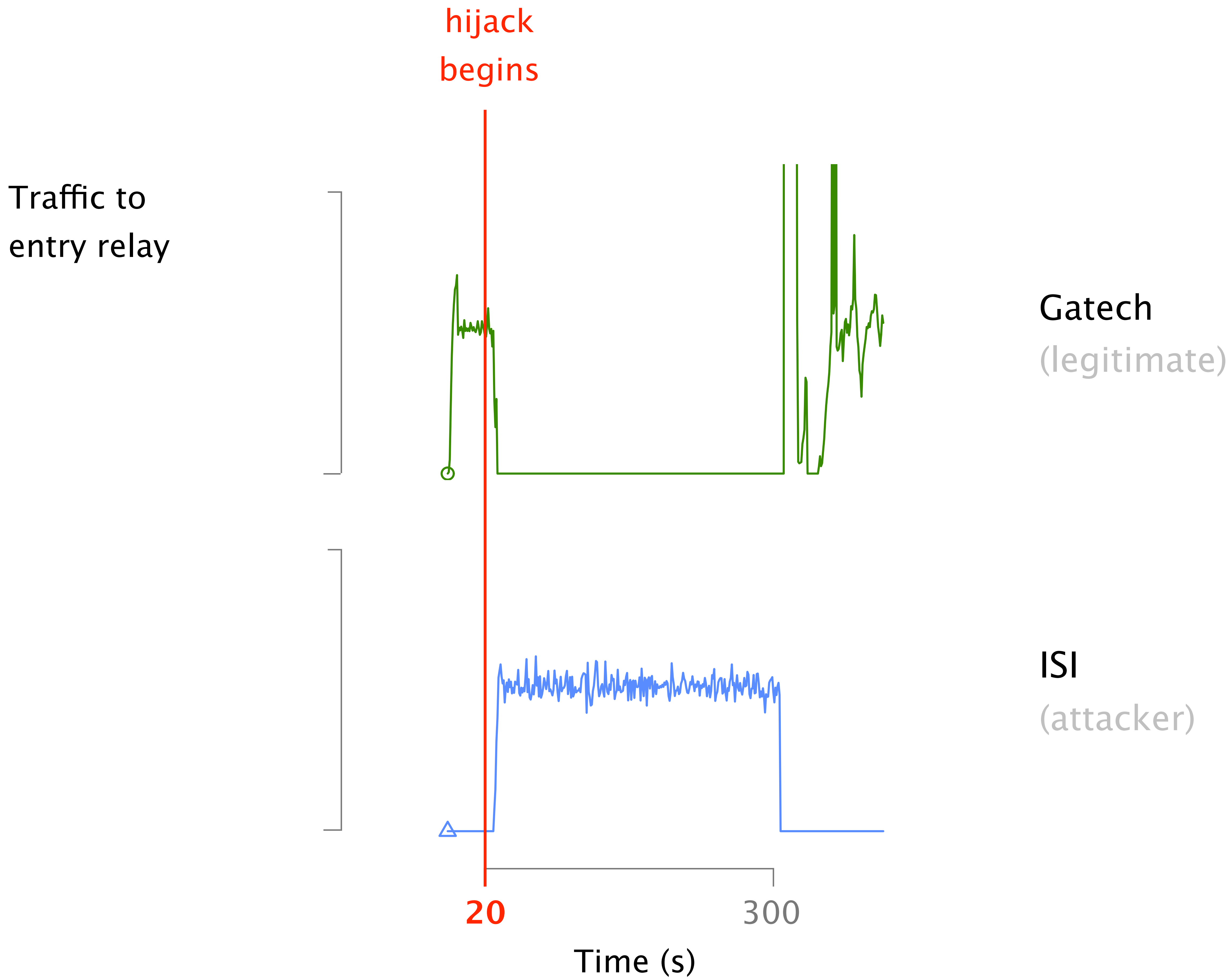


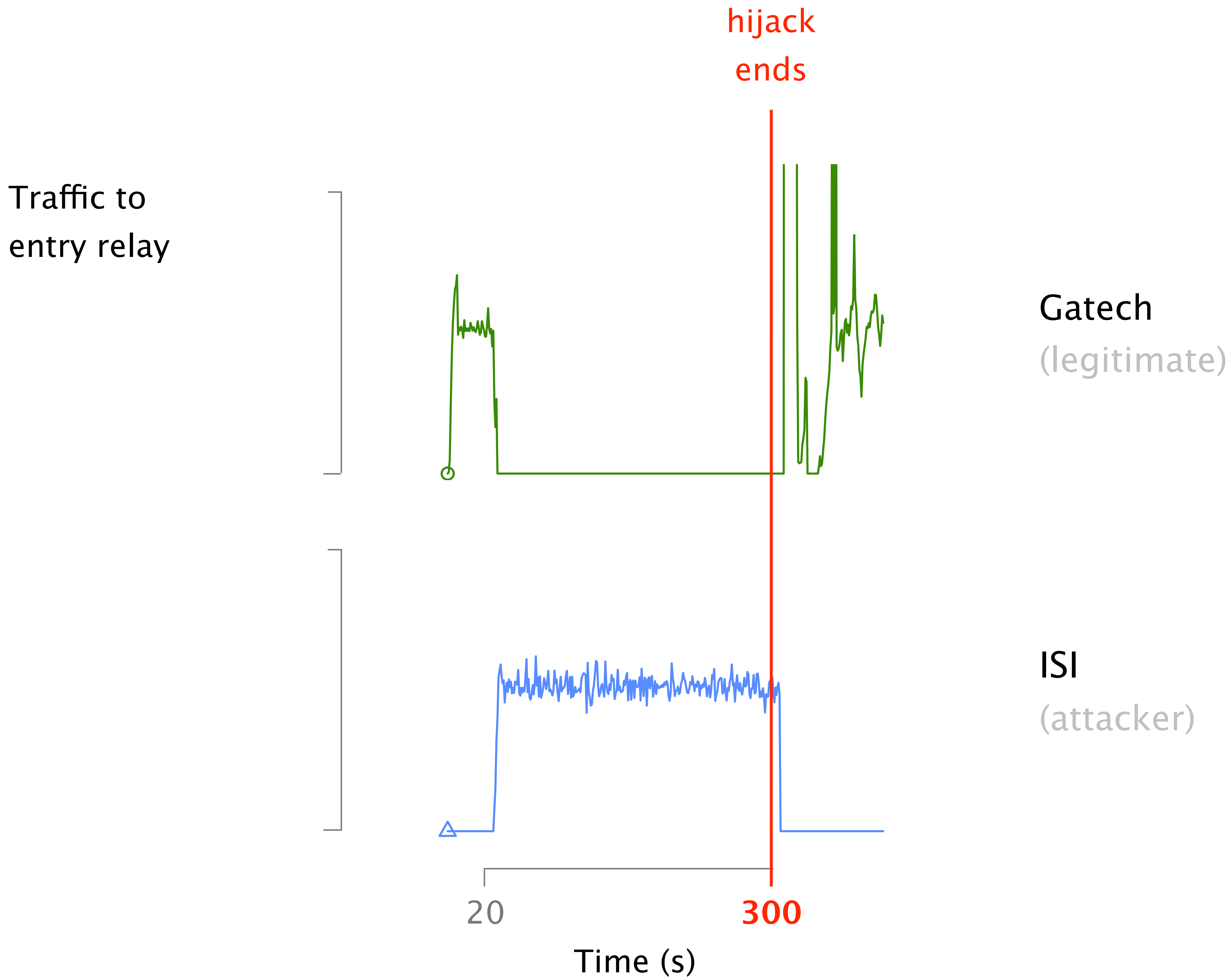
As forwarding is based on the longuest-match, all traffic soon started to enter via ISI




Traffic to
entry relay
Gatech
(legitimate)
ISI
(attacker)
20
300
Time (s)


Traffic to
entry relay
Gatech
(legitimate)
ISI
(attacker)
20
300
Time (s)


hijack
begins
Traffic to
entry relay
Gatech
(legitimate)
ISI
(attacker)
20
300
Time (s)


hijack
ends
Traffic to
entry relay
Gatech
(legitimate)
ISI
(attacker)
20
300
Time (s)

# BGP interception attacks are concerning

70 prefixes host ~30% of all entries & exits

announced by only 6 ASes

90% of the prefixes hosting relays are shorter than /24

making them vulnerable to more-specific attacks

Known attacks did *already* intercept Tor traffic

e.g., Indosat in 2011 (~5 relays) and 2014 (~44 relays)

# Anonymity on Quicksand

## Using BGP to compromise Tor

Attacks

All your traffic belongs to me

Results

Eyes wide open

3 Countermeasures

Close the curtains

# To protect itself, Tor should become more aware of the network underlying it

| | Countermeasures | Tools |
|---|---|---|
| Natural dynamism | | |
| Route manipulation | | |
| Asymmetric analysis | | |

| | Countermeasures | Tools |
|---|---|---|
| Natural dynamism | prefer stable relays | BGP monitoring |
| Route manipulation | | |
| Asymmetric analysis | | |

| | Countermeasures | Tools |
|---|---|---|
| Natural dynamism | | |
| Route manipulation | discard “suspicious” relays<br>prefer close relays | BGP monitoring +<br>BGPsec |
| Asymmetric analysis | | |

| | Countermeasures | Tools |
|---|---|---|
| Natural dynamism | | |
| Route manipulation | | |
| Asymmetric analysis | encrypt transport header | IPsec |

# These countermeasures help, but come with tradeoffs

| | Countermeasures |
|---|---|
| Natural dynamism | prefer stable relays |
| Route manipulation | discard “suspicious” relays<br>prefer close relays |
| Asymmetric analysis | encrypt transport header |

# These countermeasures help, but come with tradeoffs

| | Countermeasures | Tradeoffs |
|---|---|---|
| Natural dynamism | prefer stable relays | more power to fewer relays |
| Route manipulation | discard “suspicious” relays<br>prefer close relays | |
| Asymmetric analysis | encrypt transport header | not widely used<br>(easier to detect) |

# Anonymity on Quicksand

## Using BGP to compromise Tor

### Attacks

All your traffic belongs to me

### Results

Eyes wide open

### Countermeasures

Close the curtains

## BGP is not only a problem for Tor…

THREAT LEVEL

# Hacker Redirects Traffic From 19 Internet Providers to Steal Bitcoins

BY ANDY GREENBERG 08.07.14 | 1:00 PM | PERMALINK



… A bitcoin thief redirected a portion of online traffic from no less than 19 Internet service providers, including data from the networks of Amazon and other hosting services like DigitalOcean and OVH, with the goal of stealing cryptocurrency from a group of bitcoin users…

… A bitcoin thief redirected a portion of online traffic from no less than 19 Internet service providers, including data from the networks of Amazon and other hosting services like DigitalOcean and OVH, with the goal of stealing cryptocurrency from a group of bitcoin users…

**OVH** is the second AS in terms of # Tor relays hosted

… A bitcoin thief redirected a portion of online traffic from no less than 19 Internet service providers, including data from the networks of Amazon and other hosting services like DigitalOcean and **OVH**, with the goal of stealing cryptocurrency from a group of bitcoin users…

# Internet routing matters when it comes to user anonymity

BGP dynamics decreases user anonymity over time

natural & induced, exacerbated by asymmetric routing

The threat is real. Attacks are efficient

validated in the field, on the live Tor network

Countermeasures help—to an extent

we need a better understanding of their impacts

# Anonymity on Quicksand

## Using BGP to compromise Tor

Laurent Vanbever

www.vanbever.eu

Columbia University

May, 8 2015