Digging into Anonymous Traffic: A Deep Analysis of the Tor Anonymizing Network

Abdelberi Chaabane, **Pere Manils**, Mohamed Ali Kaafar

INRIA Rhônes-Alpes, FRANCE

pere.manils@inrialpes.fr

NSS, September 3rd, 2010

Outline

Outline

Tor

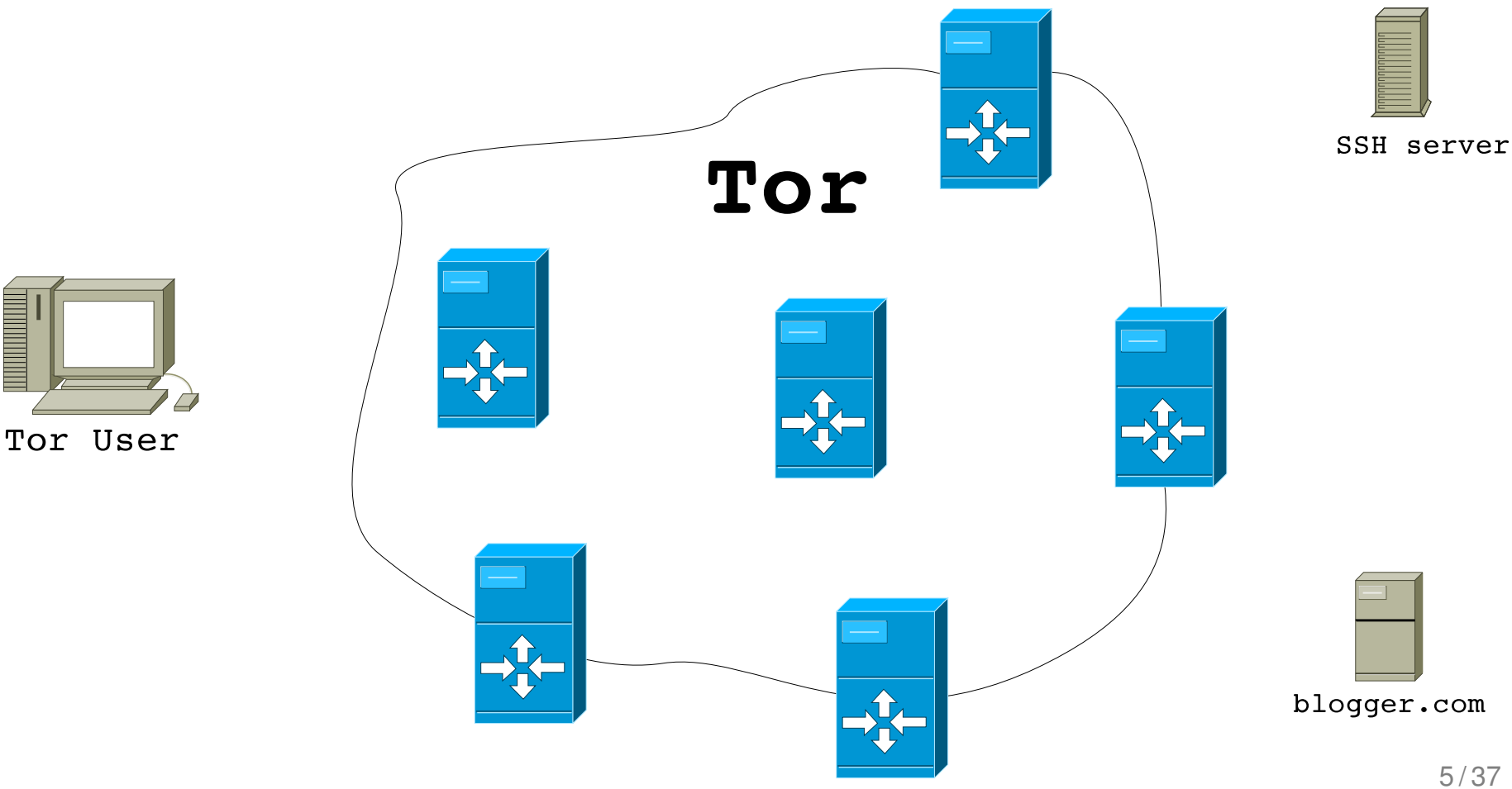
What is Tor?

- A low-latency anonymizing network.
- Only TCP traffic.
- Volunteer-based infrastructure.
- 1.800 nodes.

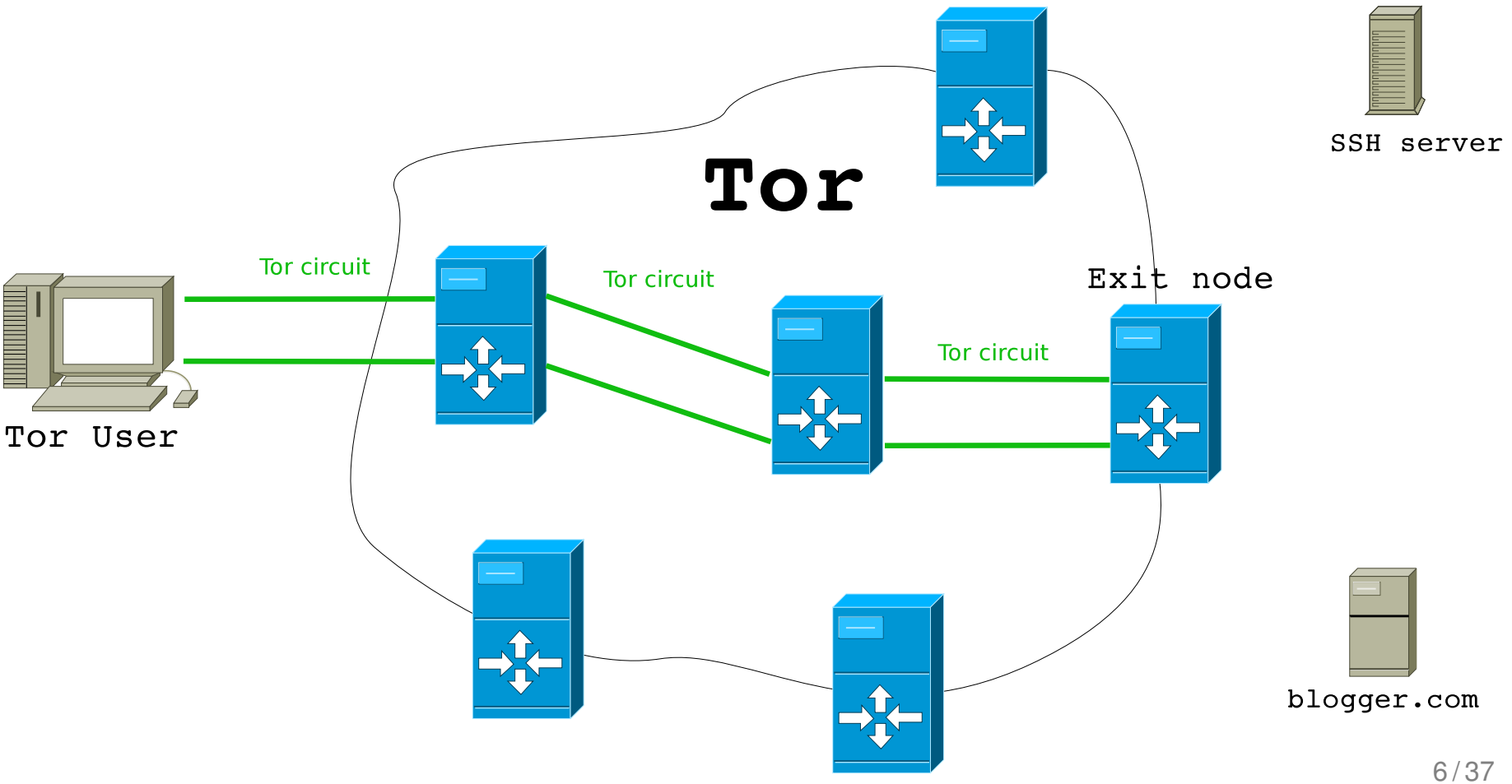
Main goal

- Prevent linking communication partners.

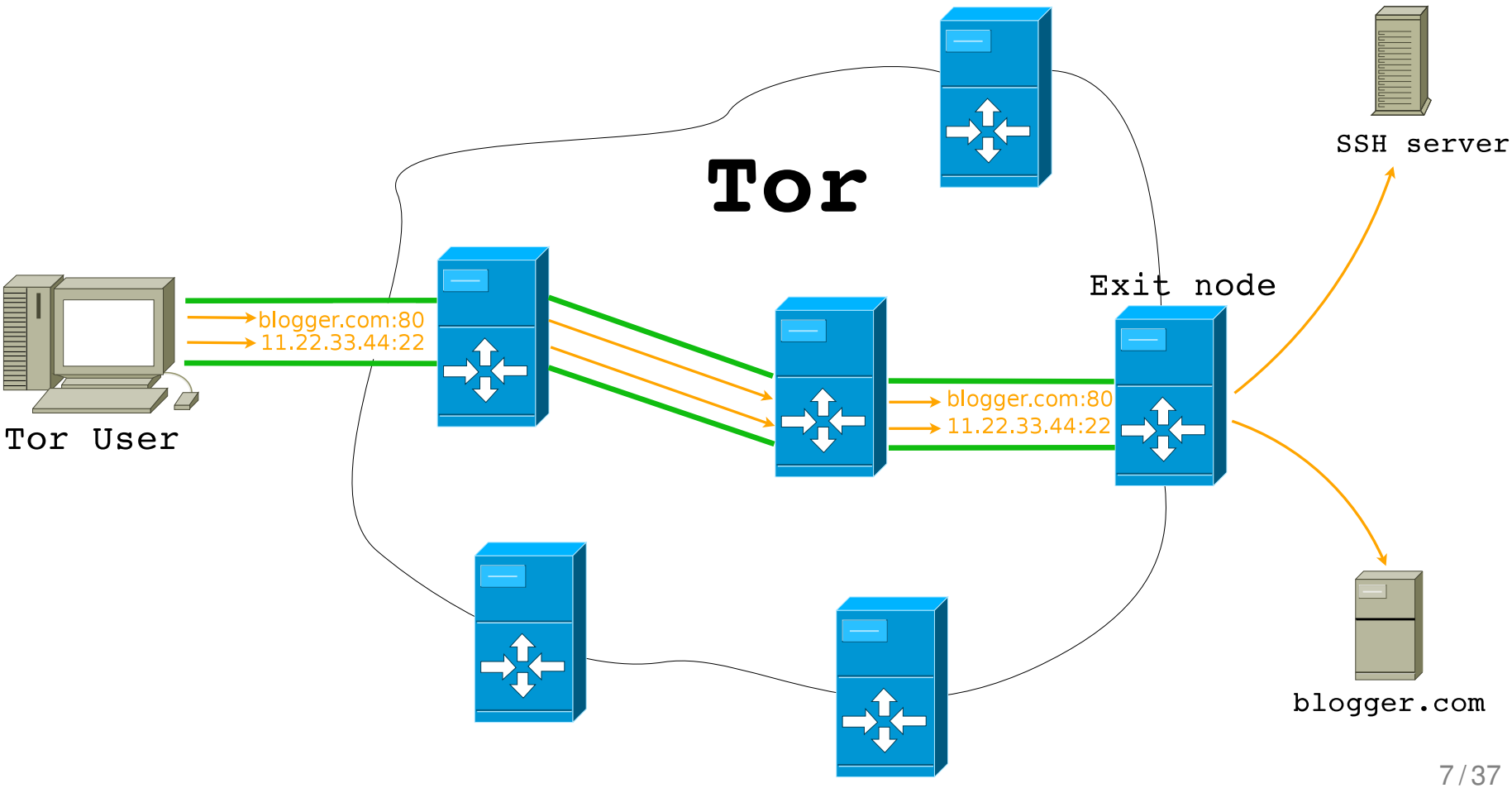
Tor: Illustration (1)



Tor: Illustration (2)

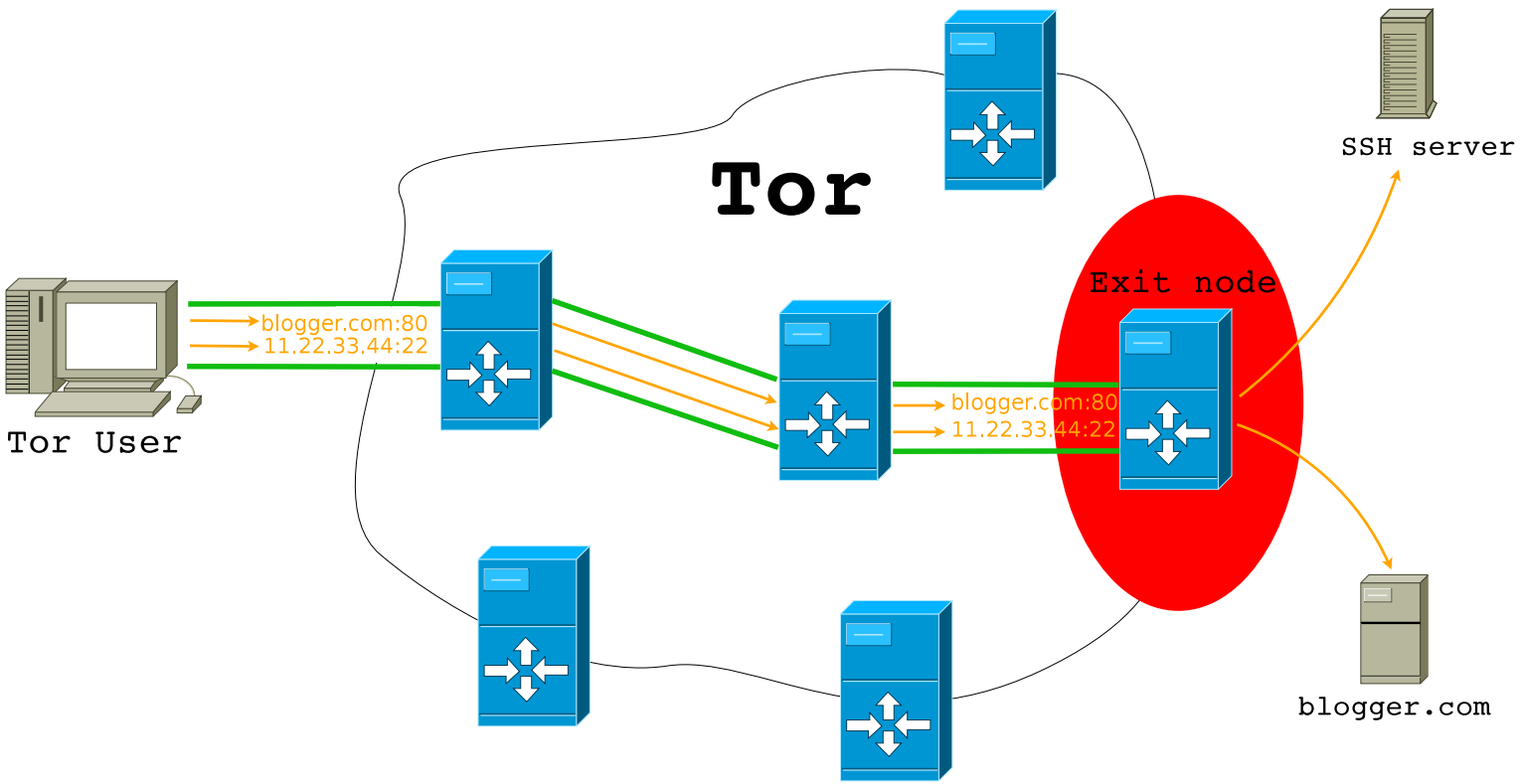


Tor: Illustration (3)



Our Experiments

- Took place at the exit node side.
- Deployed 6 exit nodes.
- Monitored them.



Outline

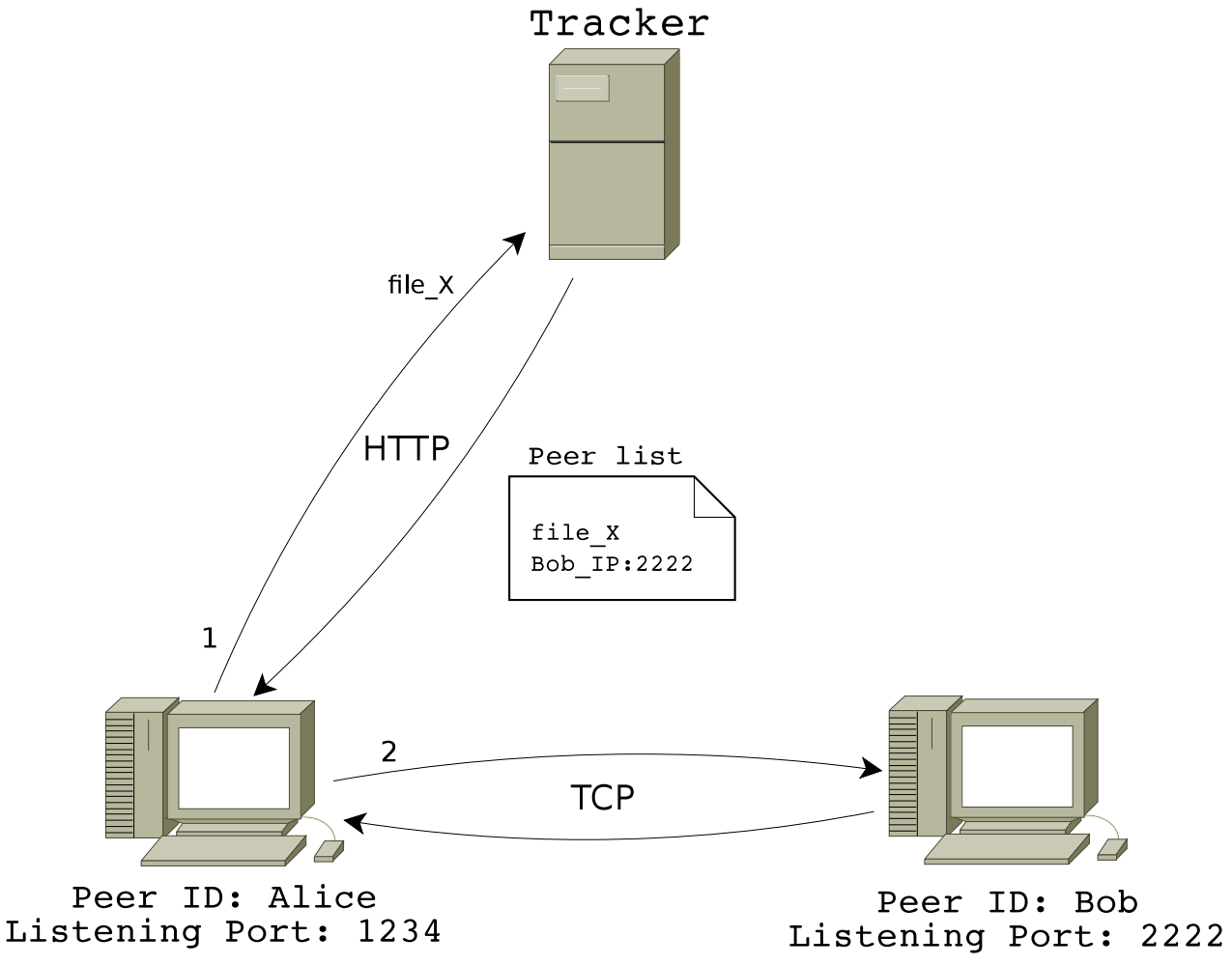
BitTorrent

- The most used P2P network.
- Peers share files.

Main entities

- **Peers:** share content between them (TCP).
- **Trackers:** help peers to know which other peers share a particular content (HTTP-TCP).

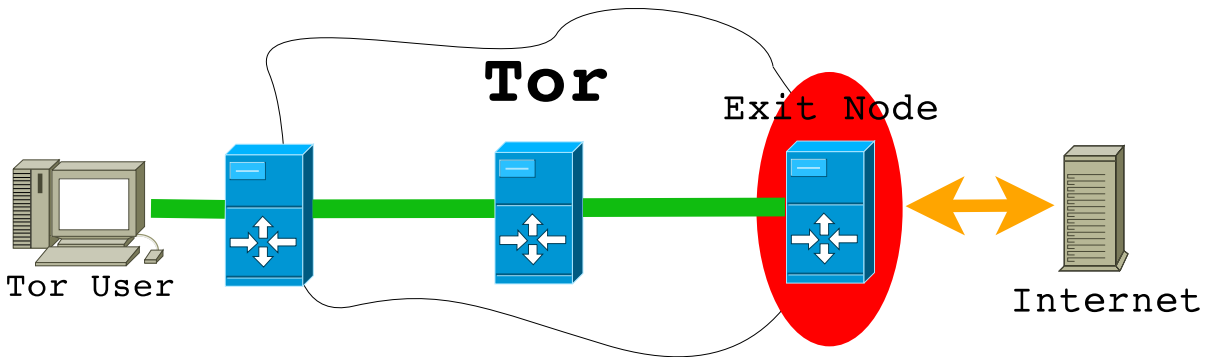
BitTorrent: Illustration



Outline

Which Protocols are Run on Top of Tor?

- Exit nodes establish connections on behalf of Tor users...



$\Rightarrow$ we can obtain aggregated statistics from them.

McCoy et al.[1] already did it in 2008.

[1] *Shining Light in Dark Places: Understanding the Tor Network*, PETS '08

Deep Packet Inspection

Why DPI?

- More accurate results than a port-based approach.

What is it?

- Technique that digs into packets (header+payload) to collect useful information (application recognition, viruses, protocol non-compilance, etc.).

How?

- Self-modified version of OpenDPI (http://www.opendpi.org).

DPI Results

- 40% of discarded flows (less than 4 packets)

Protocol	Size (%)	Flows (%)
HTTP	**36.44**	**68.57**
BitTorrent (clear)	**24.92**	**4.64**
SSL	5.37	1.83
Others P2P/ file sharing	1.17	0.22
Insecure (ftp, email, etc.)	0.32	0.09
Instant Messaging	0.26	1.72
Other well-known protocols	6.04	16.99
Unknown	**25.47**	**5.94**
Total	373.6 GB	6905 K

Outline

Digging into the Unknown Traffic

- Unknown: 25% of traffic, 6% of flows.

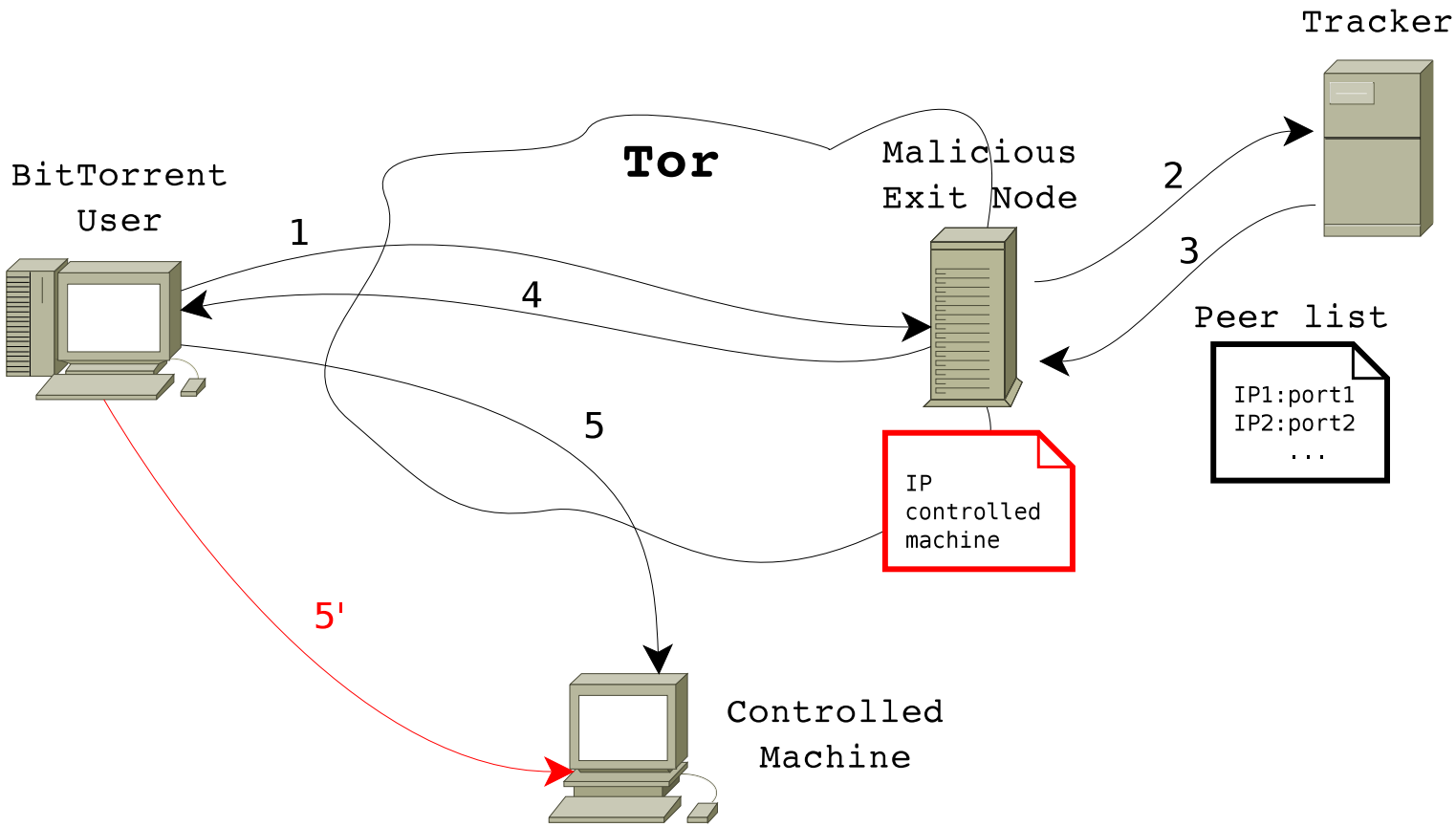
BitTorrent evidences

- P2P traffic: few connections and high traffic volume.
  $\Rightarrow$ Identified BitTorrent: 25% traffic, 4.5% flows
- Random destination ports (contacting peers).

Why our DPI did not reconize it???

- Encrypted data (high entropy).

Hijacking Tracker Responses



53% of **encrypted** BitTorrent handshakes!

Outline

HTTP Usage over Tor

- What kind of content they visit through Tor?
- What kind of websites?

Category of Visited Web Pages

Method

- Extract the `Host` header from HTTP requests.
- Use Trend Micro URL Query service to get the category.
- Group web sites into categories.

Rank	Category	Percentage
1	Search Engines/Portals	14.45%
2	Pornography	11.50%
3	Computers/Internet	11.45%
4	Social Networking	9.52%
11	Blogs/Web Communications	2.26%
13	StreamingMedia/MP3	1.82%
14	Software Downloads	1.66%
36	Hacking	0.3%
40	Political	0.18%

Outline

BitTorrent Usage over Tor

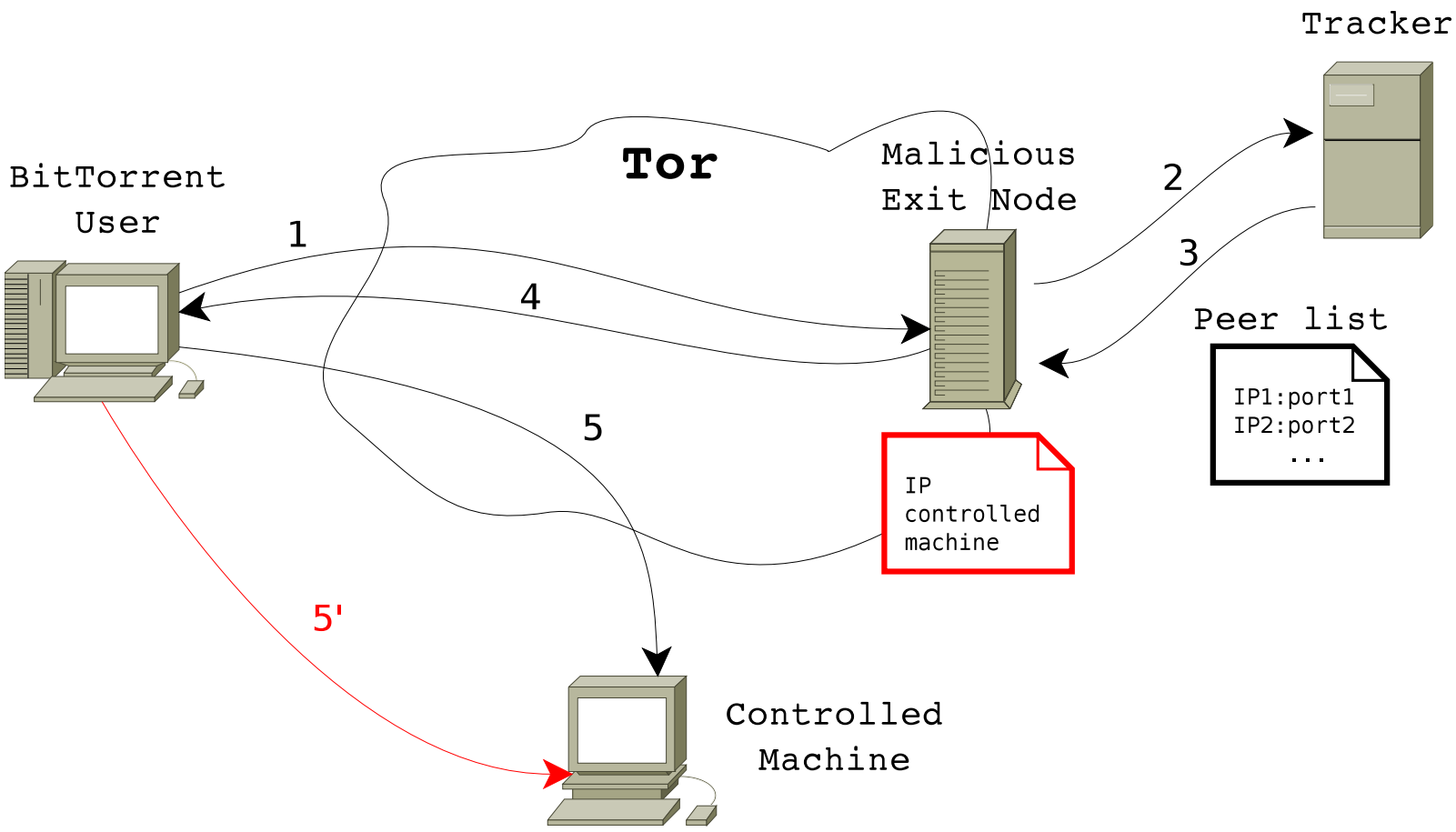
- How do BitTorrent users use Tor?
- What are they downloading through Tor?

BitTorrent Usage over Tor

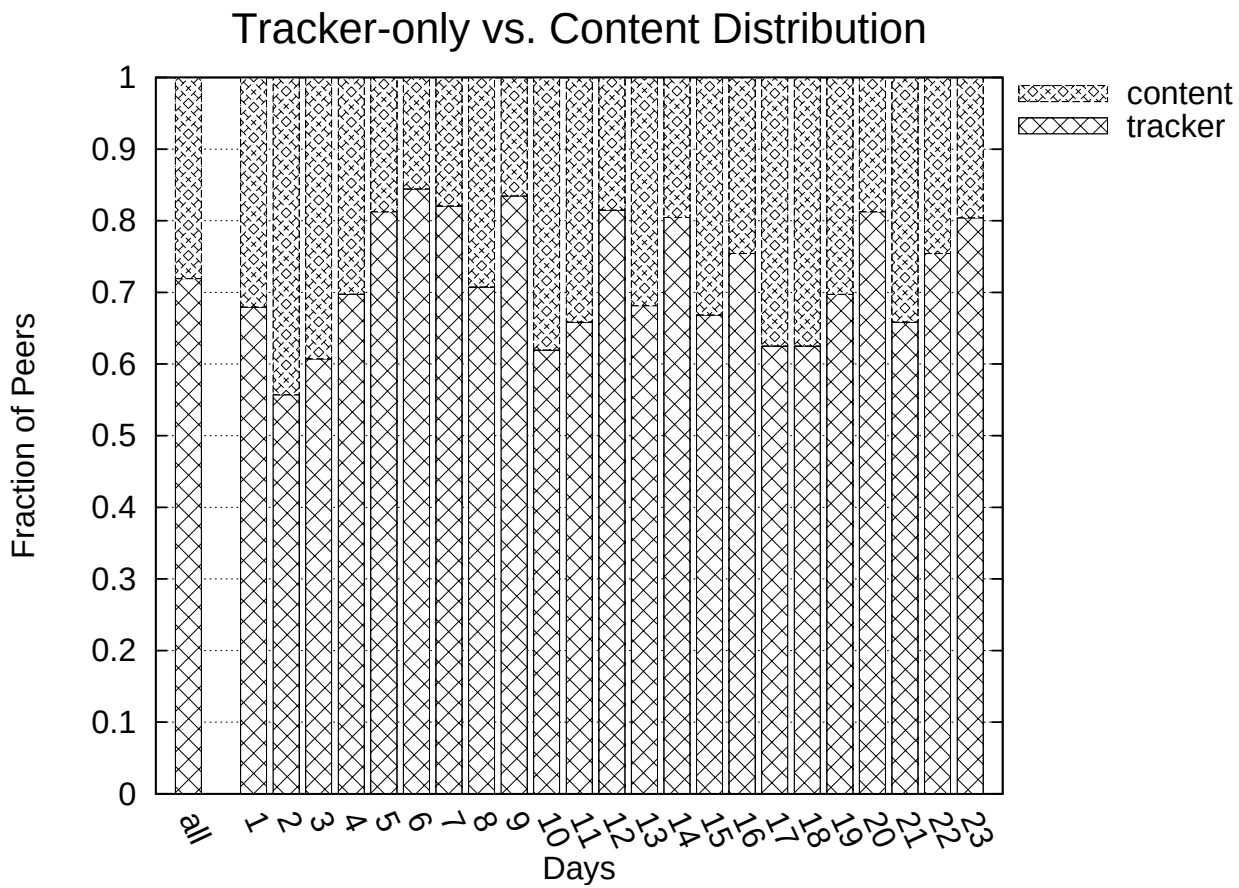
BitTorrent configured with Tor to anonymize...

- Tracker connections.
- Peer connections (harmful!).
- Both (harmful!).

Remember: Hijacking the Tracker Responses



BitTorrent + Tor: Usage



Only **30%** of BitTorrent-over-Tor users download content

Downloaded Files

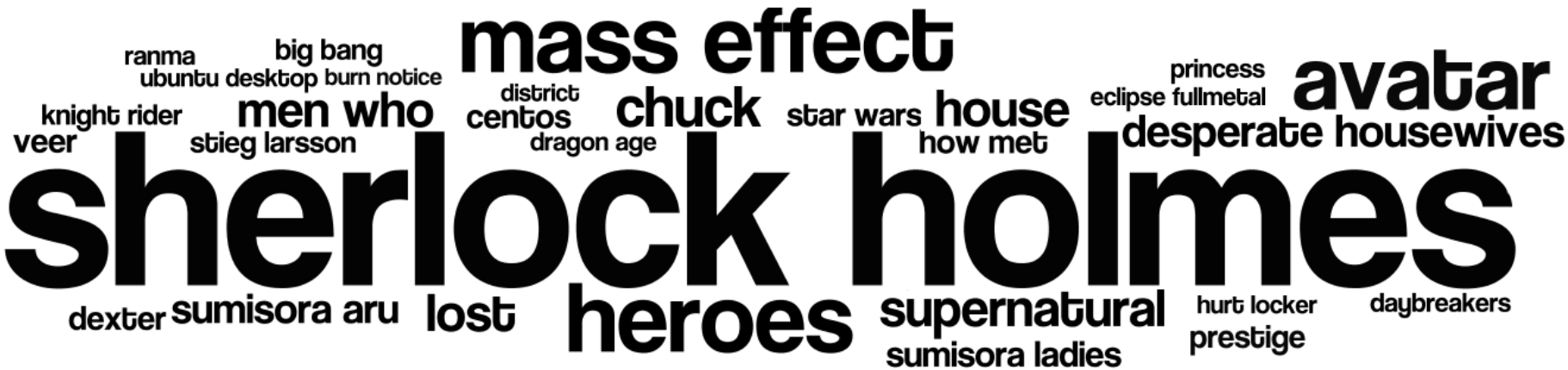
Method

- Extract infohashes from BitTorrent messages.
- Resolve the infohash into the file name.
- Consider the frequency to draw a word cloud.

- 36% infohashes not resolved $\Rightarrow$ exsitence of *darknets*[2]

[2] Zhang et al. *BitTorrent Darknets*, Infocom 2010

First 30 Most Requested Files



Copyrighted files!!!

Outline

Exit Nodes as 1-hop Proxies

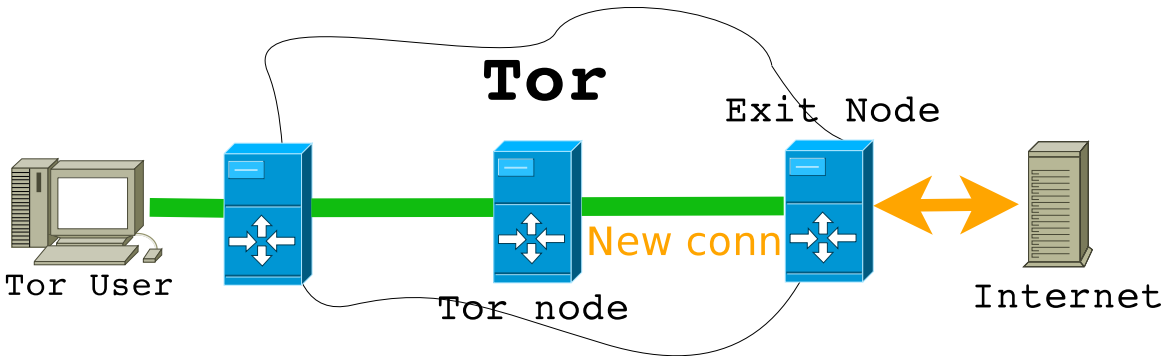
Tor is... (for bad guys)

- A reliable SOCKS proxy.
- Encrypted traffic.
- But **slow**!! (3 hops).

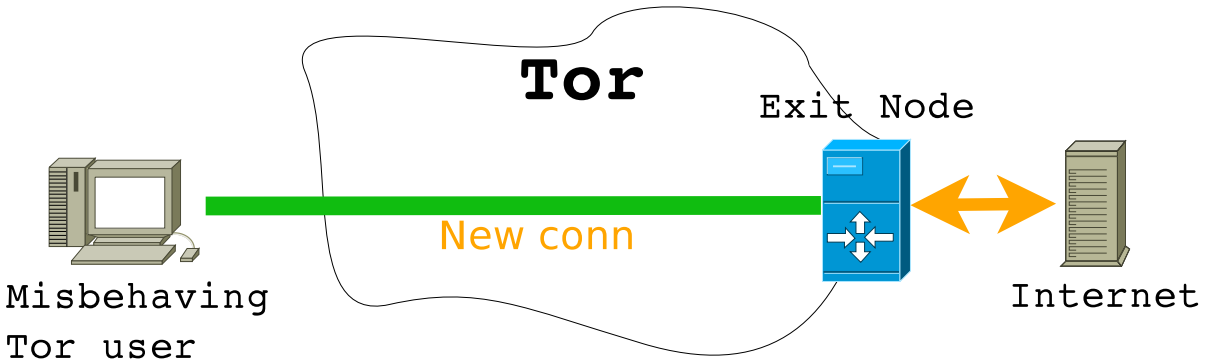
$\Rightarrow$ Use the exit nodes directly! (Tor tunnel)

Exit Nodes as 1-hop Proxies

- Normal use (3 hops):



- Tor tunnel (1 hop):



Detecting Tor Tunnels

Method

- Monitor Tor control messages asking the exit node to initiate new Internet connections.
- Are the messages coming from an other Tor node or from an unknown IP address?

OR* connections	Unique IP addresses	Once OR	Once non OR	Always OR	**Always non OR**
299977	6393	6234	504	5889	**159**

*OR = Onion Router = Tor node

Outline

Conclusions

- Detailed analysis of the anonymized traffic traveling through Tor.
  - HTTP
  - BitTorrent
- Demonstrated the importance of BitTorrent traffic over Tor (50%).
- How some users abuse Tor exit nodes as 1-hop proxies.

Questions

Thank you for your attention.
Any questions?