Technical and legal determinants of implementing a Tor exit node

Michael Sonntag, René Mayrhofer, Rudolf Hörmanseder
Starting an exit node

- Introducing an exit node is simple: Install software & start it
  - But this can lead to lots of complaints, police visits etc.
  - Your IP might end up on blacklists; your bandwidth gobbled up

- Especially problematic
  - A Graz court sentenced an exit node operator to jail (on probation)

- So to avoid problems, several precautions are necessary
  - Technical ones: Blacklists, bandwidth, attacks…
  - Legal ones: Police, ISP, your employer/family members…

- Additionally it is a research project → We need to do some research!
  - But: Without compromising the anonymity of users,
    - And not helping anyone else in this
  - and keeping to the legal limits,
    - Do we collect “personal data”?  
    - Do we now “know” about the traffic content (→ liability)?
  - and not disrupting the Tor network!
Tor @ JKU

- a) Input traffic
- b) Relay and Exit traffic
  - Two different IP addresses
- c) Mirroring of exit traffic only
  - Statistics server cannot send on that interface, only listen
  - Used for research: traffic investigation

- Firewall is for security and bandwidth limitation; does not keep logs
  - All management interfaces limited to local network

- Infrastructure and publicity: DNS and webserver are separate
  - https://www.ins.tor.net.eu.org

- Not shown in picture: JKU operates an IDS for all Internet traffic
  - Special configuration for counting rule matches only

- Separate hardware in a separate room
Technical aspects (1)

- De-anonymization can be performed in two variants
  - Directly based e.g. on the content (we are an exit node!)
    - At the moment we do not collect/investigate any content data at all
  - “Tracing back” one step, i.e. aligning output data to input data
    - We store IP addresses & ports with a timestamp, but only on exit side

- How to prevent correlation?
  - Split incoming, outgoing, and relay traffic
    - Patch submitted to Tor which has now been integrated
    - We only investigate outgoing traffic, so reconfiguration/addition would be necessary to obtain anything useful; data seizure would be useless
  - Stored data is aggregated, and dropped if too few connections
    - One-hour slots; needs at least 10 connections to be shown, e.g. as destination country
      - Otherwise: 1 connection to Germany → Correlation trivial
    - Only top 50 destination countries/Autonomous Systems listed
    - Note: Total aggregate (bytes, connections…) includes everything!
Technical aspects (2)

- Bandwidth limitation to 200 Mbit/s:
  - Based on Tor node, firewall, and uplink provider (JKU/Aconet)
    - None keep logs, but all limit the amount
  - Note: We have a low priority. When the university was attacked via DDoS, our Tor node suffered significant bandwidth reduction
    - Recommended: Tor should be secondary to “core business” traffic

- Separate IP address space: To prevent blacklisting the university
  - A common threat in abuse reports
    - We do not know whether this has been put in action by anyone
  - Typical example: Spam on port 25 – but we don’t allow this port

- Similarly: Separate domain name (+reverse DNS and WhoIs entry)
  - Could be blacklisted, but also enables a custom abuse contact
  - Allows disclosure of this being a Tor server, too

- Operate equipment at capacity
  - Any additional task, e.g. storing data, should overwhelm them
Legal aspects (1)

- Physical and logical access restricted to the professor
  - “Research privilege” – also includes the right **not** to publish
    - This does not allow keeping data secret if a court asks, but courts would have to overcome a further argument if requiring disclosure
  - Important elements to be weighted:
    - Freedom of speech: End users, whose traffic would be investigated
      - Note that it is technically challenging to filter out a single user!
      - Investigating “everything” → Mandatory data retention, struck by ECJ
    - Freedom of research: Monitoring the exit node would lead to different behavior/exclusion of the node → No meaningful research possible
    - Legal investigation: Victims/prosecutors have a right to identify the responsible person/criminal
      - Might be useless, as 2 other systems are involved. So our data is only helping at all if the other nodes are simultaneously investigated too
  - Additionally: Without login the system must be forcefully shut down, which cannot go unnoticed
    - The statistics server would notice, as well as the firewall
Legal aspects (2)

- Incoming traffic is never touched for analysis, only exit traffic
  - Disclosing existing data is much easier than requiring someone to collect data in the first place
    - Requires expertise, equipment → Expenses!
    - Argument (JKU: useless): “How would I do this? I have no knowledge!”
      - Note: We are NOT a “telecommunication service provider” which is required to support & fulfill such requests!

- No human inspection: Averting liability for content
  - If we know about illegal traffic (who/IP & what), we are required to take action to stop it
  - We make sure we don’t know about it (legally: no obligation for searching/verification exists; explicitly stated by EU directive)
  - If unavoidable (e.g. the IP addresses; but when/which of them are illegal?) delay it → no meaningful action possible any more

- Be prepared to block destinations
  - We offer this in our abuse replies – no one requested this yet
Legal aspects (3)

- High bandwidth: Many connections per second imply that typical timestamps (second precision) are useless to identify a single connection for traceback
  - 100 outgoing connections per second → 100 incoming ones
  - The interesting ones are therefore 1% of disclosure
  - This can be argued as “useless” and targeting too many innocents

- Associated webserver: Public disclosure
  - Everyone can trivially find out that this is not “our” traffic
    - So we do not claim this as our own traffic (→ liability)
    - May practically help to reduce problems or inquiries

- Separating in- and output to different providers (theory only!)
  - So no single point exists where both in- and outgoing streams could be monitored simultaneously
  - Ideally also in different jurisdictions, e.g. through a tunnel
Legal aspects (4)

- Criminal law problems specific (?) to Austria
  - § 119a StGB: Data interception
    - Not a problem as it requires to use the data for financial gain or any kind of damage to someone else (includes third parties!)
    - But: Not only use, but also making available or publishing
      - We cannot publish raw data, as this might be useful for building profiles (ads), extortion (damage need not be money!) etc.
  - § 119 StGB: Secrecy of telecommunication
    - You or someone else without rights uses a “device” (incl SW) to obtain knowledge of message content which is sent by telecommunication
      - Message content: Human content (“text”), not metadata (IP addresses, ports etc). Special protection, secrecy etc. not needed
    - Using the device is sufficient, reading the message etc is not required
    - Potential solution: We are not interested in the content of “a” message, but only in anonymous statistics of “all” messages
    - Practical solution: Prosecution only with permission of the victim
      - Who would have to prove that he actually is a “victim”
Current results – Flows/bytes
Current results – Encrypt. ratio (Bytes)
Current results – Destination countries

Traffic

<table>
<thead>
<tr>
<th>Country</th>
<th>Last hour (if in top 50)</th>
<th>Total average per hour</th>
</tr>
</thead>
<tbody>
<tr>
<td>US</td>
<td>16GB</td>
<td></td>
</tr>
<tr>
<td>DE</td>
<td>15GB</td>
<td></td>
</tr>
<tr>
<td>NL</td>
<td>14GB</td>
<td></td>
</tr>
<tr>
<td>AT</td>
<td>13GB</td>
<td></td>
</tr>
<tr>
<td>FR</td>
<td>12GB</td>
<td></td>
</tr>
<tr>
<td>RU</td>
<td>11GB</td>
<td></td>
</tr>
<tr>
<td>ES</td>
<td>10GB</td>
<td></td>
</tr>
<tr>
<td>IT</td>
<td>9GB</td>
<td></td>
</tr>
<tr>
<td>GB</td>
<td>8GB</td>
<td></td>
</tr>
<tr>
<td>RO</td>
<td>7GB</td>
<td></td>
</tr>
<tr>
<td>CN</td>
<td>6GB</td>
<td></td>
</tr>
<tr>
<td>LU</td>
<td>5GB</td>
<td></td>
</tr>
<tr>
<td>CZ</td>
<td>4GB</td>
<td></td>
</tr>
<tr>
<td>JP</td>
<td>3GB</td>
<td></td>
</tr>
<tr>
<td>KR</td>
<td>2GB</td>
<td></td>
</tr>
<tr>
<td>CA</td>
<td>1GB</td>
<td></td>
</tr>
<tr>
<td>EU</td>
<td>0GB</td>
<td></td>
</tr>
<tr>
<td>FI</td>
<td>0GB</td>
<td></td>
</tr>
<tr>
<td>DK</td>
<td>0GB</td>
<td></td>
</tr>
<tr>
<td>VN</td>
<td>0GB</td>
<td></td>
</tr>
</tbody>
</table>
IDS report


- All blocked attacks:
  1. 63502  Joomla Object Injection Vulnerability
  2. 17506  Sqlmap HTTP Request
  3. 10213  Hydra Brute Force Attempt
  4. 7371   IIS Extended Unicode Directory Traversal
  5. 5183   IP Reputation
  6. 3079   Nikto HTTP Request
  7. 2669   WordPress XMLRPC Pingback Usage
  8. 1575   Backdoor: Zeus Botnet C&C Phone Home Request
  9. 1221   %252f Double Encoded / in URI
 10. 1069   SlowHTTPTest DoS Tool

....

50. 14   Cisco Prime Data Center Network Manager Arbitrary File Upload Vulnerability

Sum of Top 50: 119,677 blocked requests (≈ 161/hour)
Further research

- Extensions planned: IPv6
  - How many persons are using it, are differences in countries/AS visible, encryption ratio changes?
  - Potentially a problem with statistics, if few users or sites

- Extensions planned: Wider exit policy
  - Allowing more ports for communication and restricting only definitely problematic ones, like 25 (SMTP)

- Switching off the IPS of the JKU for the exit node
  - Has to be tried very carefully: Currently the abuse report load is easily manageable (19.11.2015-4.4.2017: 195 reports)
  - Will take place in brief slices of increasing length

- Most traffic is web surfing: Ports 80 and 443
  - Which sites are visited? What encryption is used?
  - Note that this is legally much more difficult (§ 119 StGB)!
Conclusions

- Separate the exit node from your normal network:
  - IP addresses, DNS, logical, physical
- Do not create “tiny” exit nodes; only “bulk” traffic helps anonymity
- Try to fill up the hardware and bandwidth to their limit
- Prepare to be able to block destinations
- Ensure that you do not know about the traffic content
  - Don’t look inside
  - Any investigation must be fully automatic without manual intervention, ideally even in case of problems
- Ensure timely reply to requests (police inquiries, abuse reports)
- Public disclosure: This is an exit node and not our traffic

Tor@JKU: tor2e.ins.tor.net.eu.org / www.ins. ... / 193.171.202.146
THANK YOU FOR YOUR ATTENTION!

Michael Sonntag
michael.sonntag@ins.jku.at
+43 (732) 2468 - 4137
S3 235 (Science park 3, 2nd floor)