Anonymity and Privacy

50.520 Systems Security
Paweł Szałachowski

Privacy

- “*Ability of individuals, groups, or institutions to determine for themselves when, how, and to what extent information about them is communicated to others.*”

Anonymity

- “*Anonymity ensures that a user may use a resource or service without disclosing the user’s identity. The requirements for anonymity provide protection of the user identity. [...] Anonymity requires that other users or subjects are unable to determine the identity of a user bound to a subject or operation.*”

Anonymity vs Privacy

- Anonymity is about hiding identity
- Privacy is about hiding information/actions
- Anonymity in the context of (Internet) communication
    - Very difficult to achieve
    - Adversary
        - MITM (eavesdropping or active)
        - Contacted endpoint (e.g., a website operator)
    - Unlinkability, indistinguishability, and anonymity set

Why we need these properties?

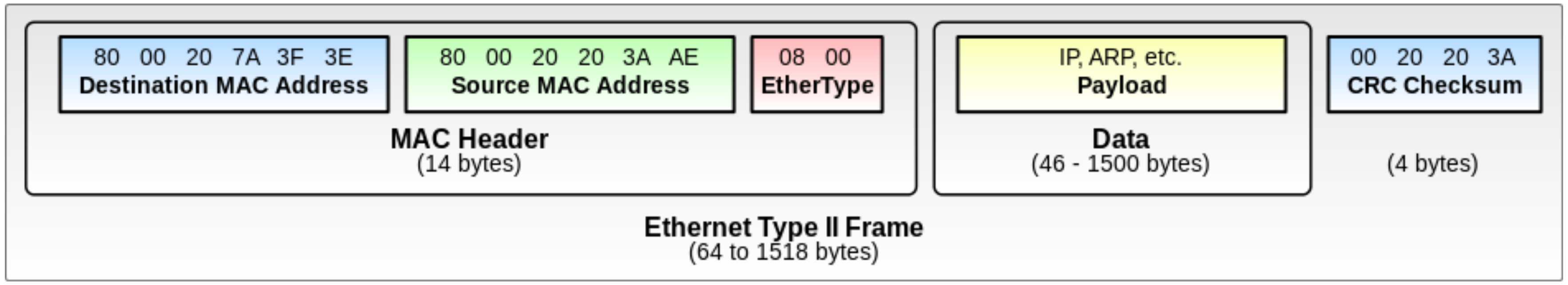
- Social and Political Motivations
  - People tend to be more honest
- Work
  - Legal or HR departments, Police, Journalists, …
- Economical Motivations
  - Why so many services are for free?
    - *“If there is no product you are the product.”*
- Snowden (2013)
  - PRISM, XKeyscore, Tempora, …

Interfering Privacy and Anonymity

Physical Layer

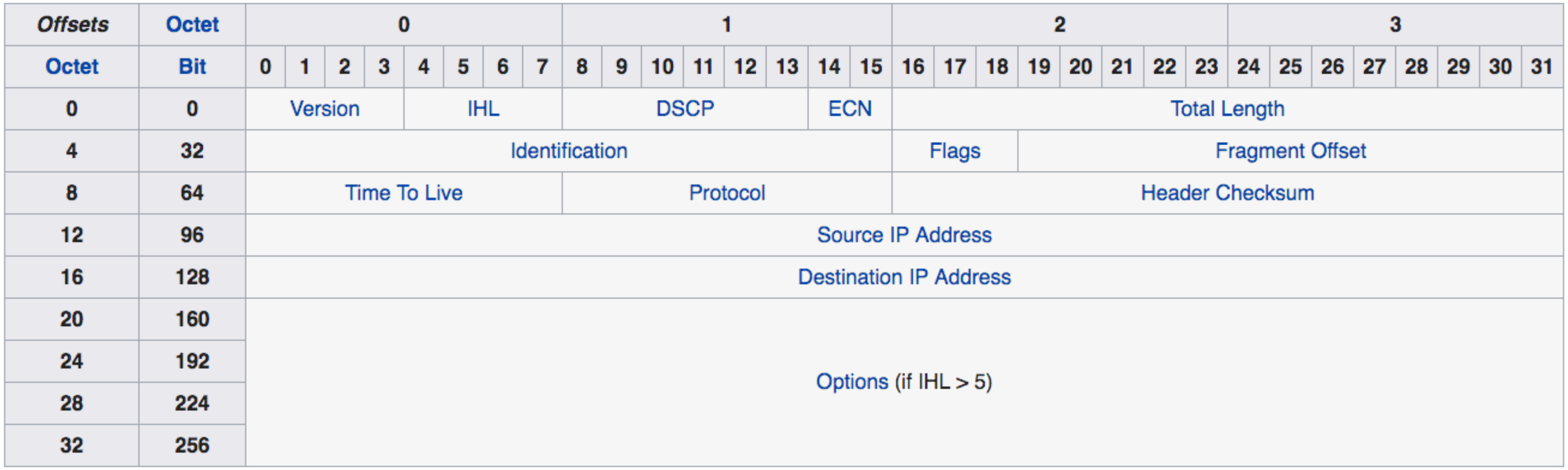
- Requires access to hardware/medium involved in the network
  - Network taps (to monitor traffic)
- Powerful adversary able to find a physical location

Data Link Layer



- Media Access Control (MAC) sublayer
  - Reminder: MAC addresses have to be unique
    - Manufacturers take care of that
  - MAC addresses reveal manufacturers (sometimes models, factories, series, …)
- Limited scope of observation (LAN)
  - However, (according to Snowden) NSA heavily uses it for tracking people
  - How to prevent?

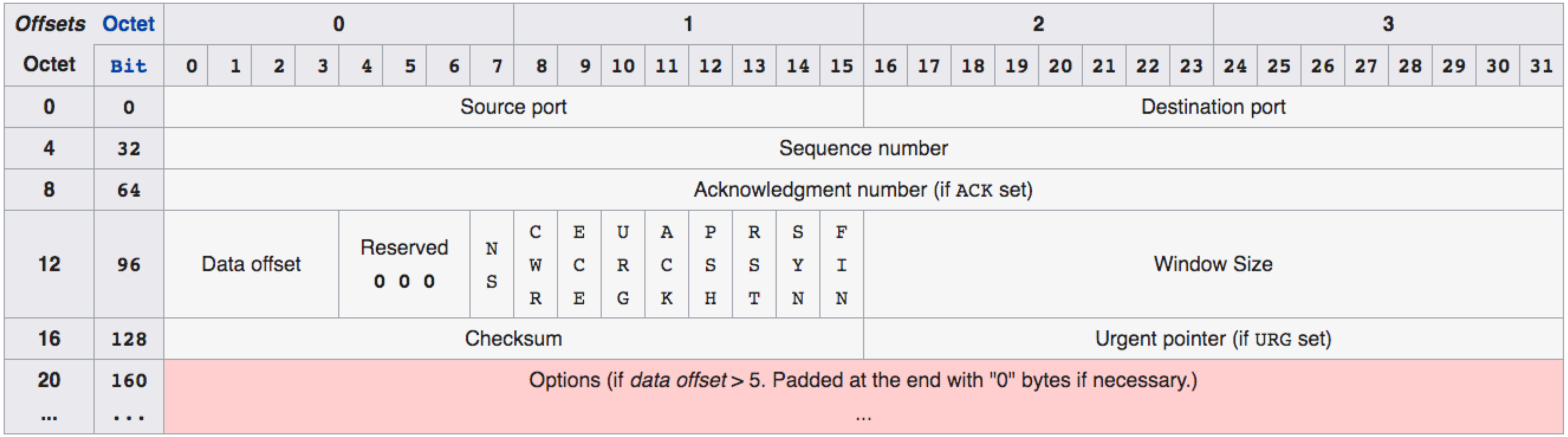
Network Layer



Offsets	Octet	0								1								2								3							
Octet	Bit	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
0	0	Version				IHL				DSCP						ECN		Total Length															
4	32	Identification																Flags			Fragment Offset												
8	64	Time To Live								Protocol								Header Checksum															
12	96	Source IP Address																															
16	128	Destination IP Address																															
20	160	Options (if IHL > 5)																															
24	192																																
28	224																																
32	256																																

- Addresses are required for routing and communication
    - Main target for revealing identities
    - Address ranges are allocated to AS (ownership can be easily checked)
    - Often addresses are static and bound to a person/host/department/…
        - Mapping between IPs and domain names
- NAT helps but not too much (anonymity set is still small)
- Statistical traffic analysis
- Active fingerprinting and other fields can reveal software used (e.g., OSes set different initial TTL)

Transport Layer

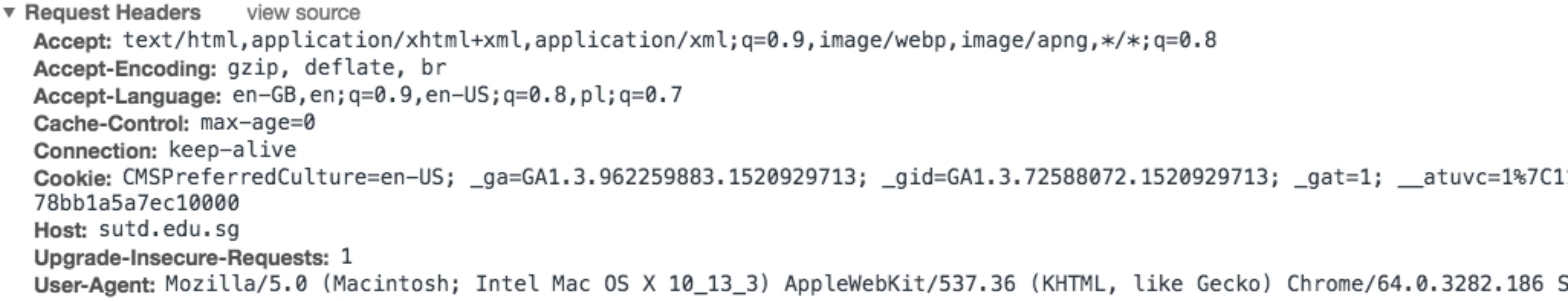


Offsets	Octet	0								1								2								3							
Octet	Bit	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
0	0	Source port																Destination port															
4	32	Sequence number																															
8	64	Acknowledgment number (if ACK set)																															
12	96	Data offset				Reserved 0 0 0			N S	C W R	E C E	U R G	A C K	P S H	R S T	S Y N	F I N	Window Size															
16	128	Checksum																Urgent pointer (if URG set)															
20 ...	160 ...	Options (if *data offset* > 5. Padded at the end with "0" bytes if necessary.) ...																															

- Ports can identify applications
- Ports, sequence numbers, congestion window, options, can passively identify software implementing the TCP stack
- Active fingerprinting is possible too
  - e.g., sending TCP segments with incorrect or unexpected flags

Application Layer

- Application-specific metadata
    - Session (tokens, usernames, …)
    - Location and language
    - Software version used
    - Encoding
- Data
- What sutd.edu.sg can learn about me (even with the incognito mode)?



```
Accept: text/html,application/xhtml+xml,application/xml;q=0.9,image/webp,image/apng,*/*;q=0.8
Accept-Encoding: gzip, deflate, br
Accept-Language: en-GB,en;q=0.9,en-US;q=0.8,pl;q=0.7
Cache-Control: max-age=0
Connection: keep-alive
Cookie: CMSPreferredCulture=en-US; _ga=GA1.3.962259883.1520929713; _gid=GA1.3.72588072.1520929713; _gat=1; __atuvc=1%7C1
78bb1a5a7ec10000
Host: sutd.edu.sg
Upgrade-Insecure-Requests: 1
User-Agent: Mozilla/5.0 (Macintosh; Intel Mac OS X 10_13_3) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/64.0.3282.186 S
```

Mechanisms to Improve Privacy and Anonymity

Encryption

- Hides upper layers
  - e.g., IPSec protects transport layer, TLS protects application layer, …
- Even if communication is encrypted a passive adversary still can learn some information
  - Timing and length
  - Sometimes it is enough
    - How would you attack privacy of an user browsing a subset of https://wikipedia.org?
- What if you would like to hide from contacted server?

Network-layer Anonymity

- Which layer(s) to anonymize?
  - Does it make sense to protect upper layers (transport or application) w/o protecting the network layer ?
    - Probably not, as IP gives a very good accuracy
  - Upwards from the network layer
    - It is good to protect lower layers too

Network-layer Anonymity

- Alice wants to send a message to Bob anonymously
  - Requirements
    - Low-latency (critical)
    - Bandwidth
    - Security
- Adversary model
  - Your ISP, state-level adversary, global adversary, or Bob

Proxy Servers

- Idea: Alice sends (securely) a message to a proxy server that will forward the message to Bob
- Different Implementations
    - SSL/TLS tunnels (stunnel)
    - SOCKS proxies
    - VPNs
- Pros and Cons
    - ○ latency (not too bad actually) ○ usually services are paid
    - ○ the proxy server is a trusted party

Onion Routing

- How to make sure that the proxy server does not know destination?

- Idea: introduce more “proxy servers” and route messages through them

- Design Goal: No proxy can learn both Alice and Bob

- Onion: a layer of encryption

Tor

- A low-latency open anonymity network
    - An overlay network with mixes
- Hidden services
    - Services that are accessible only within the Tor network
    - .onion TLD
- Software bundles
    - Browser, proxy servers, …

The anonymous Internet

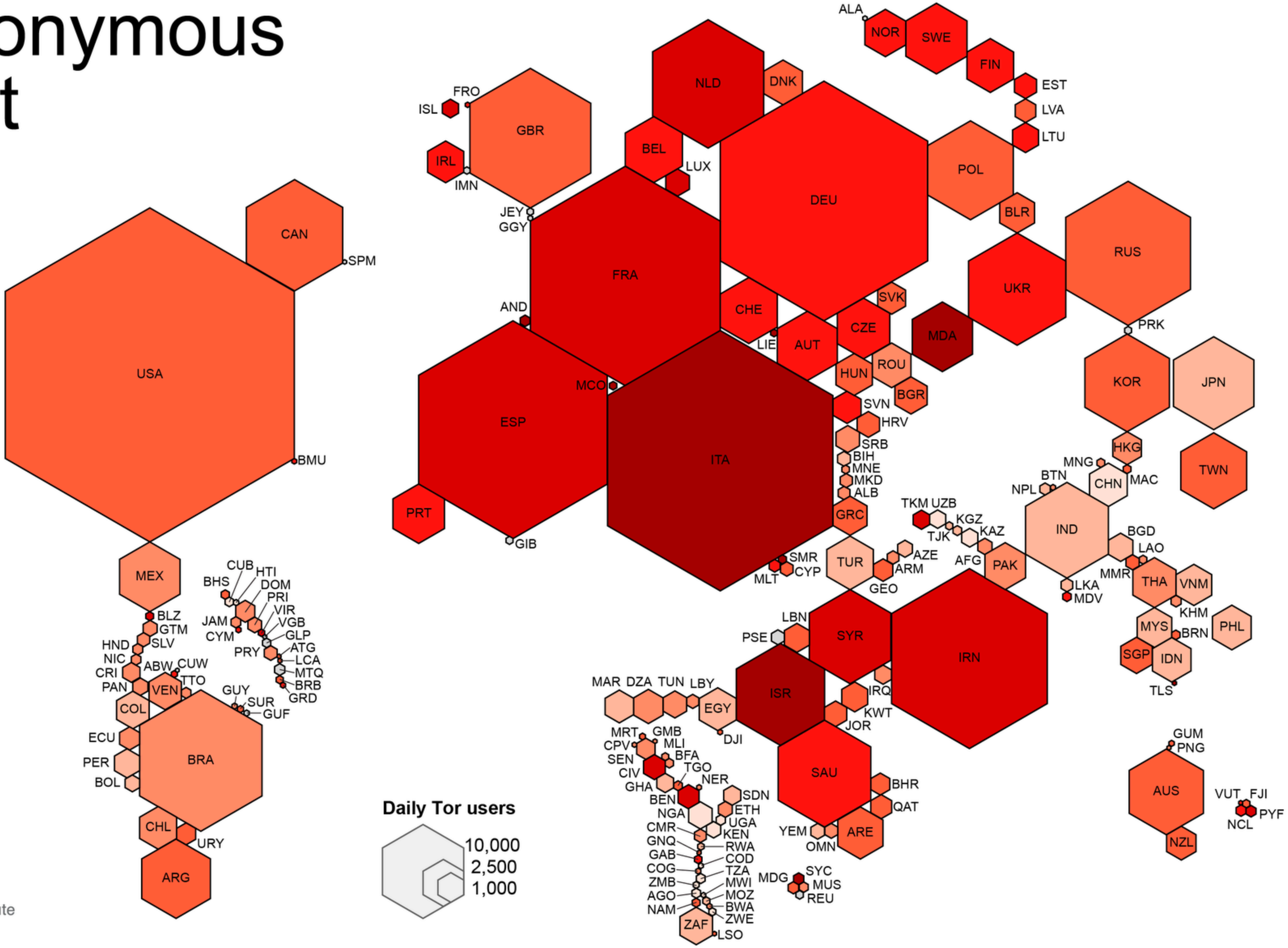
Daily Tor users per 100,000 Internet users

- ■ > 200
- ■ 100 - 200
- ■ 50 - 100
- ■ 25 - 50
- ■ 10 - 25
- ■ 5 - 10
- ■ < 5
- ■ no information

Average number of Tor users per day calculated between August 2012 and July 2013

data sources:
Tor Metrics Portal
metrics.torproject.org
World Bank
data.worldbank.org

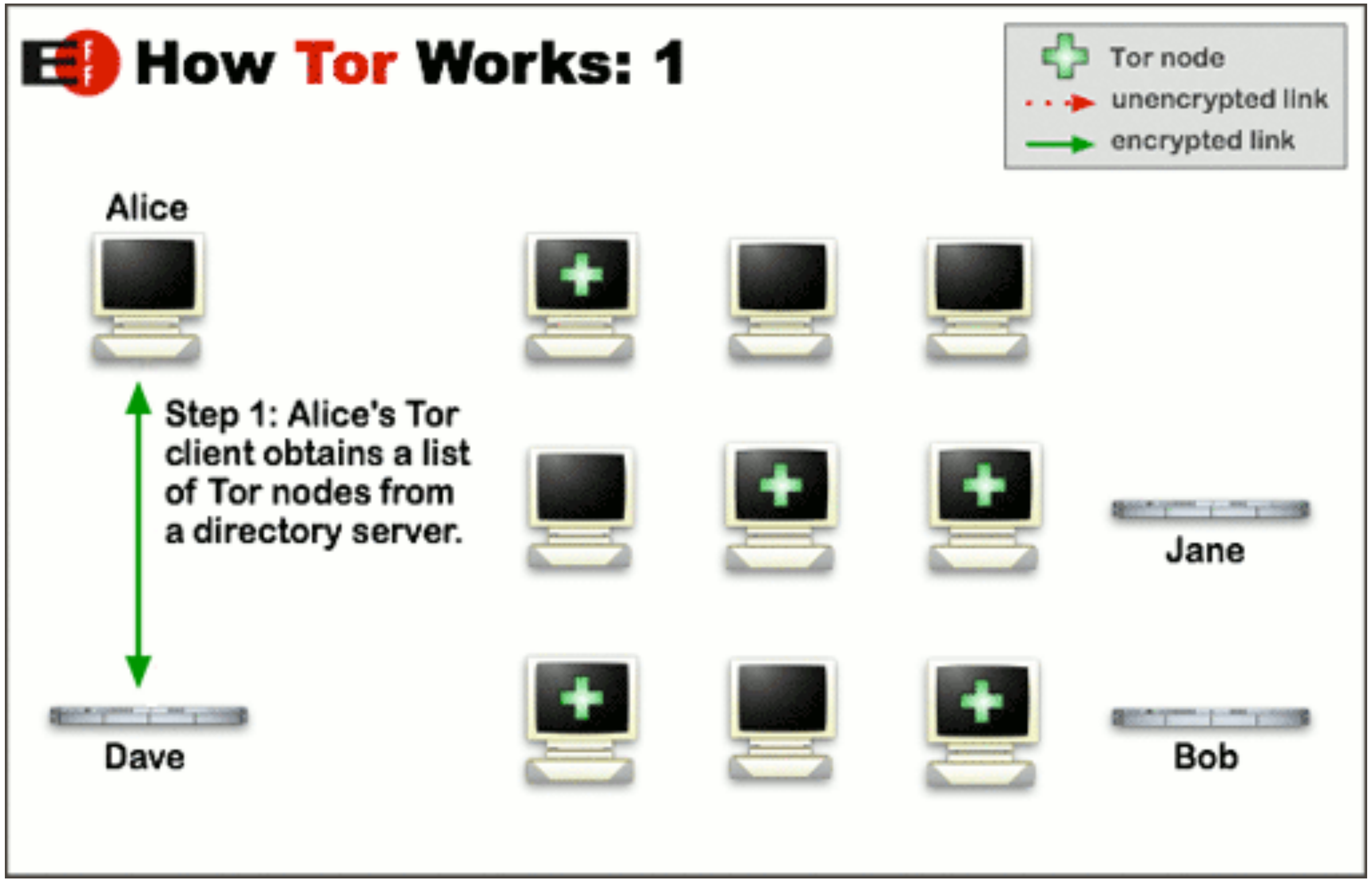
by Mark Graham (@geoplace) and Stefano De Sabbata (@maps4thought)
Internet Geographies at the Oxford Internet Institute
2014 • geography.oii.ox.ac.uk



Oxford Internet Institute
University of Oxford

Tor Network

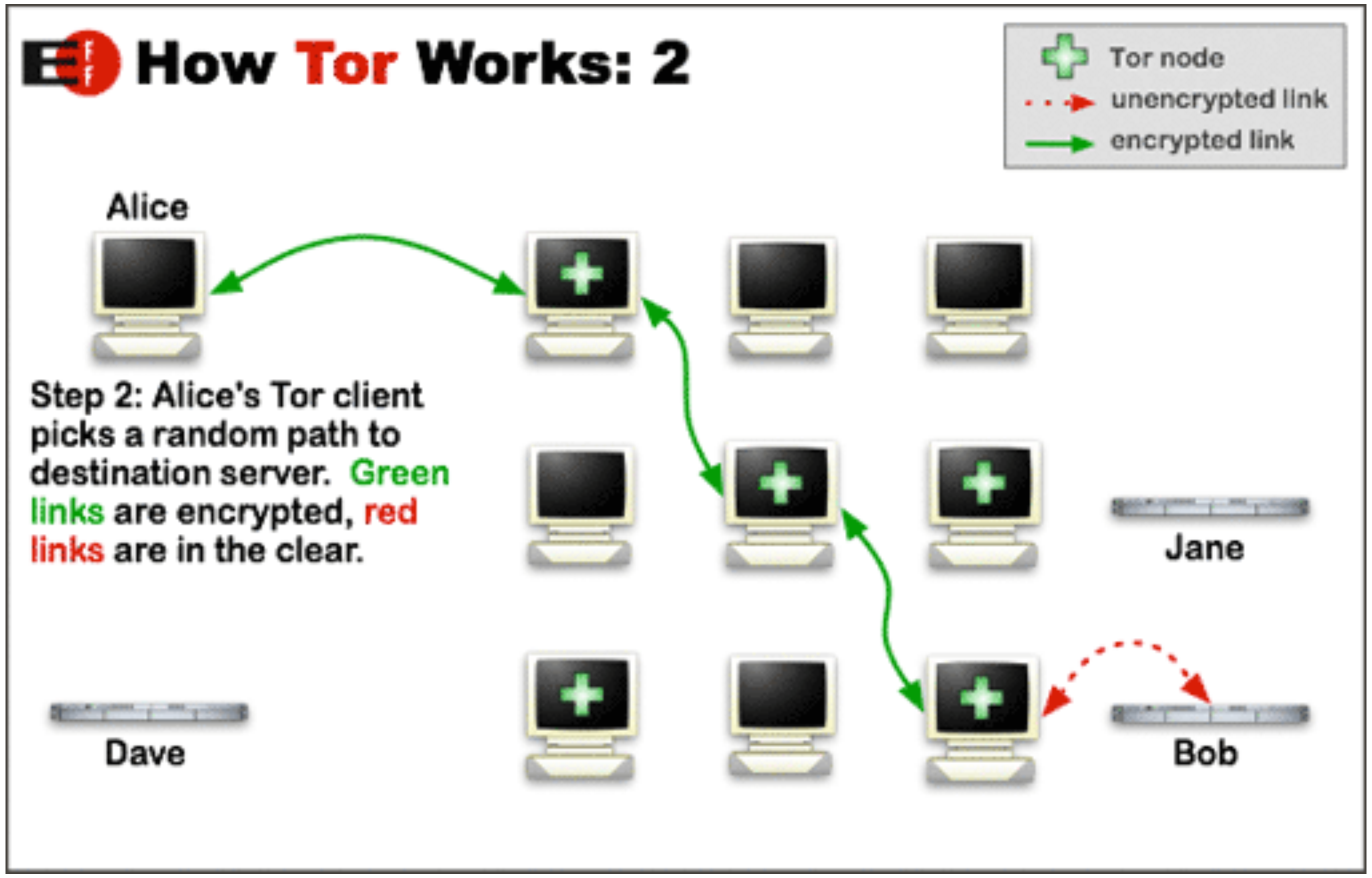
- Different nodes
  - Middle relays
  - Exit Relays
  - Bridges



Step 1: Alice's Tor client obtains a list of Tor nodes from a directory server.

Tor Network

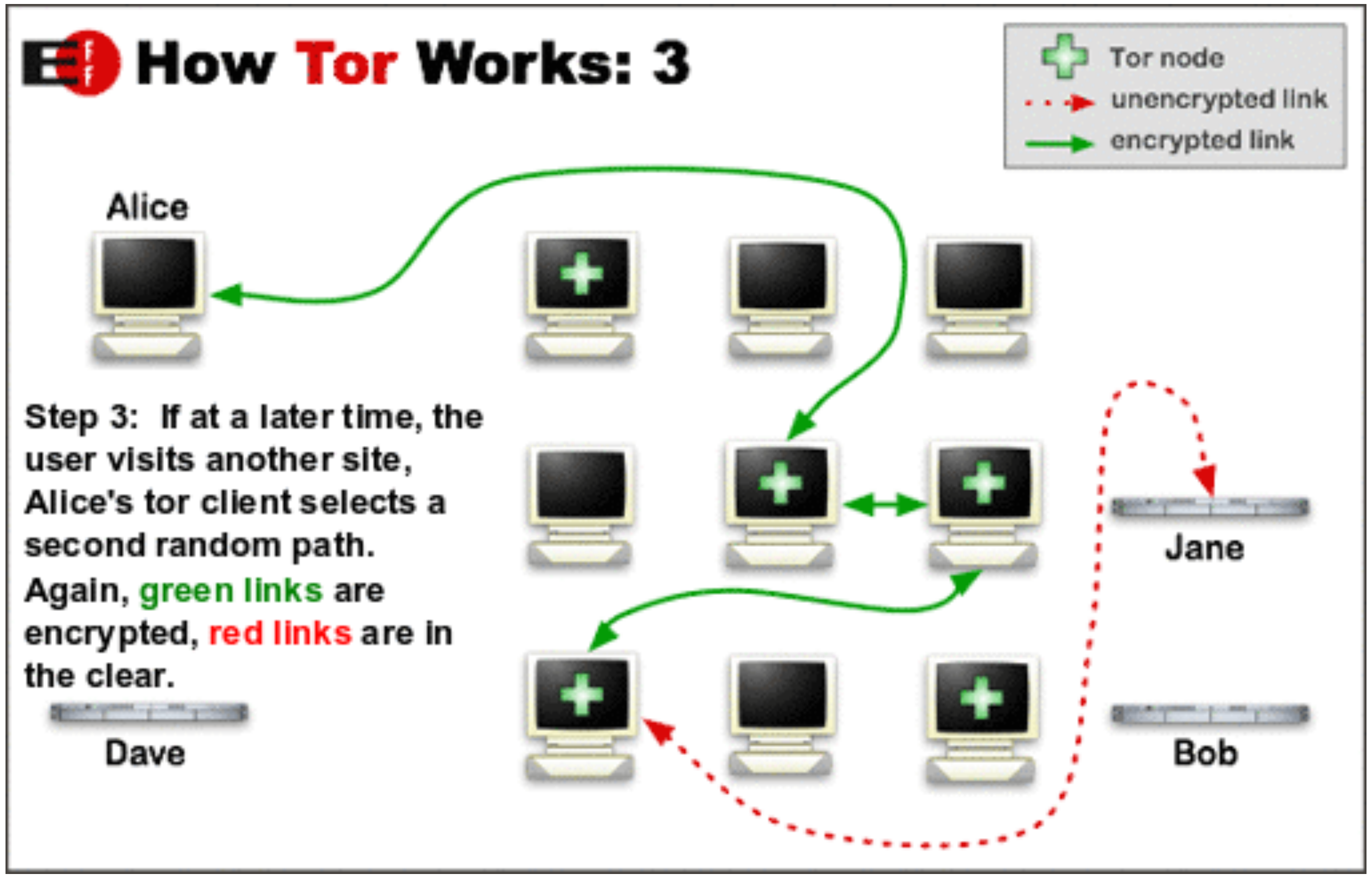
- Circuit
  - Two middle relays and one exit relay
- Circuits are selected by clients
  - Randomized selection algorithm
- Circuit establishment
  - Relays establish peer-to-peer (TLS) connections



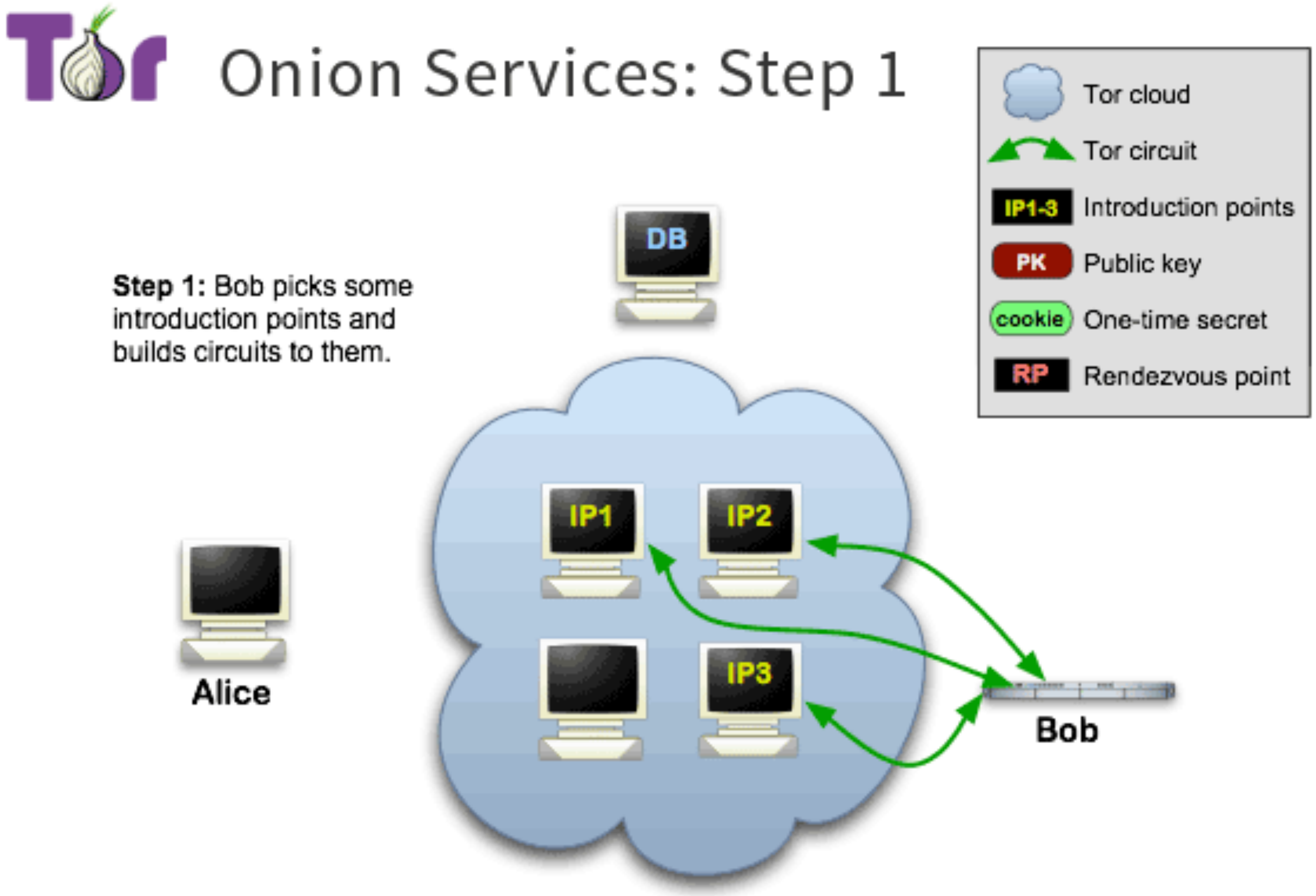
Step 2: Alice's Tor client picks a random path to destination server. Green links are encrypted, red links are in the clear.

Tor Network

- A new circuit can be established for every new website
  - Why needed?


How Tor Works: 3
Tor node
unencrypted link
encrypted link
Alice
Step 3: If at a later time, the user visits another site, Alice's tor client selects a second random path. Again, green links are encrypted, red links are in the clear.
Jane
Dave
Bob

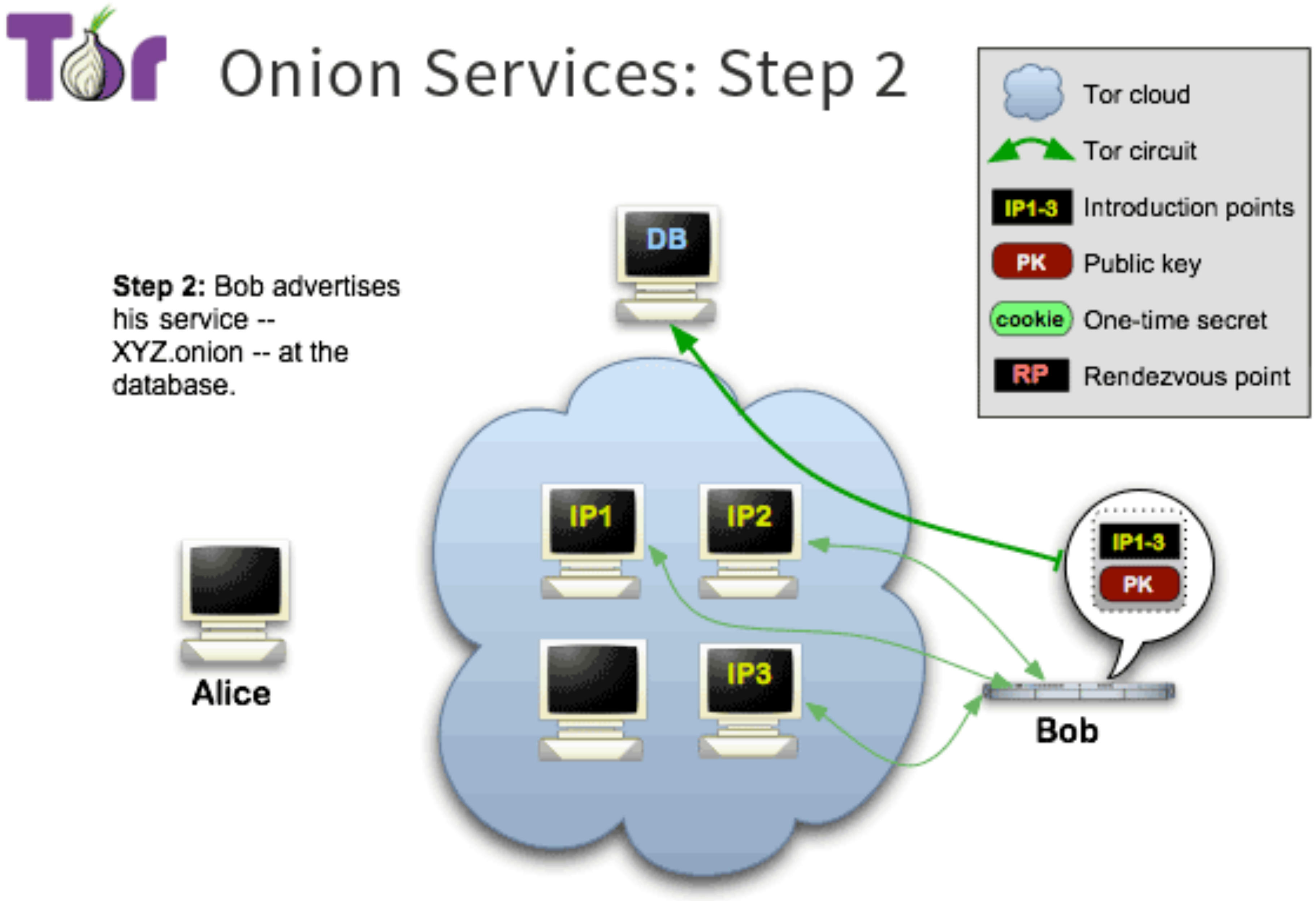
Tor Hidden Services



Onion Services: Step 1

Step 1: Bob picks some introduction points and builds circuits to them.

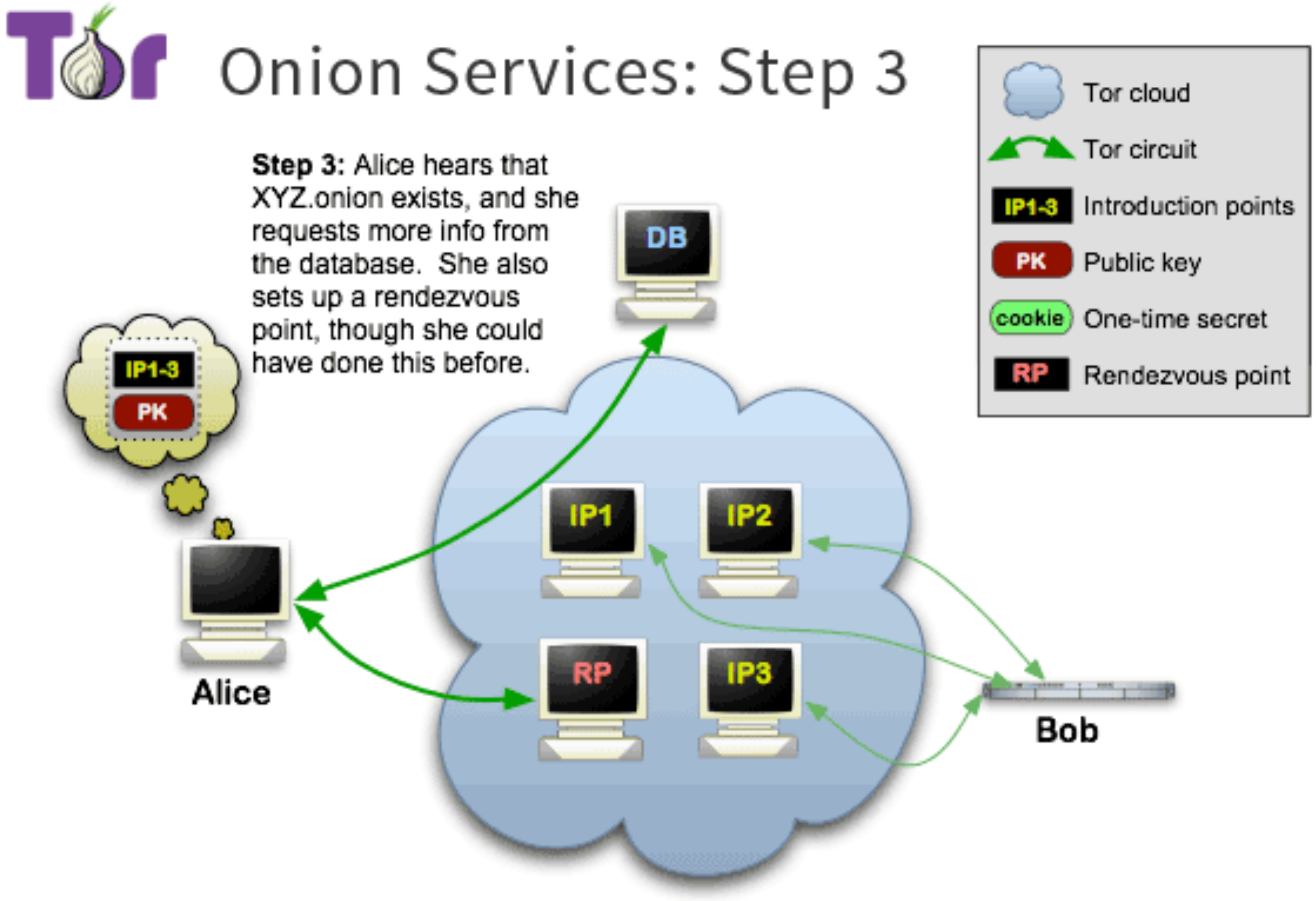
Tor Hidden Services



Onion Services: Step 2

Step 2: Bob advertises his service -- XYZ.onion -- at the database.

Tor Hidden Services

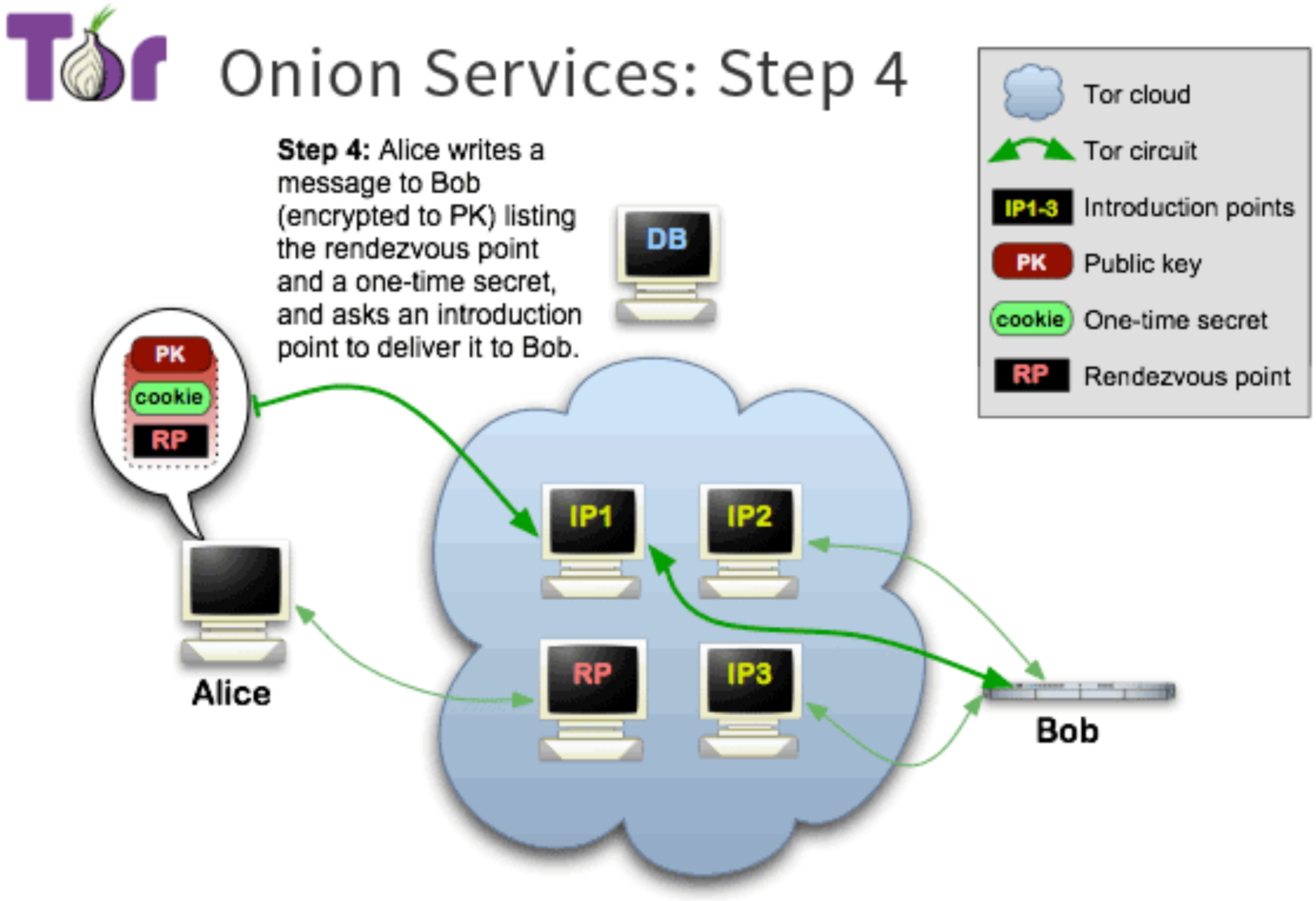


Tor Onion Services: Step 3

Step 3: Alice hears that XYZ.onion exists, and she requests more info from the database. She also sets up a rendezvous point, though she could have done this before.

☁	Tor cloud
↔	Tor circuit
IP1-3	Introduction points
PK	Public key
cookie	One-time secret
RP	Rendezvous point

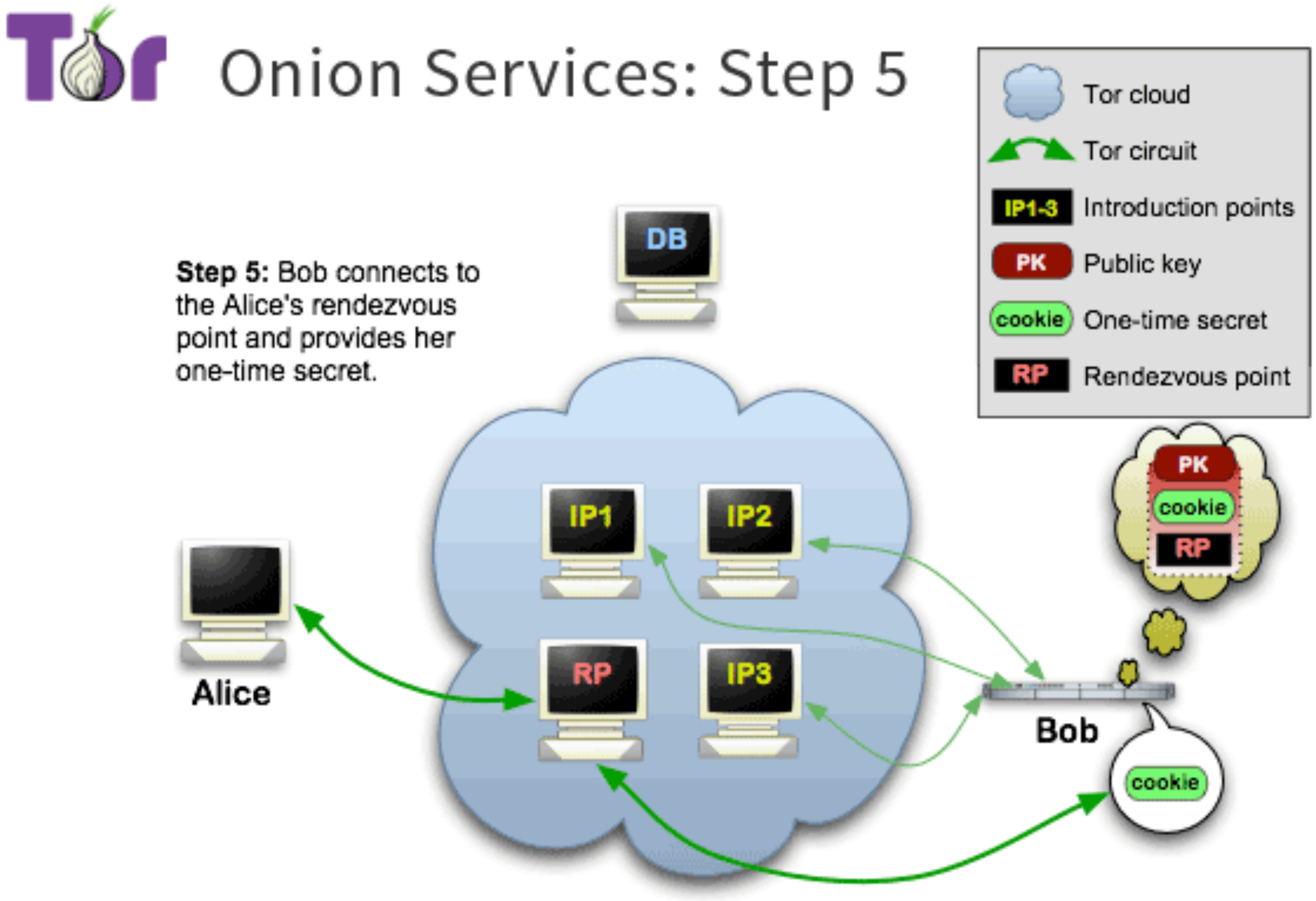
Tor Hidden Services



Onion Services: Step 4

Step 4: Alice writes a message to Bob (encrypted to PK) listing the rendezvous point and a one-time secret, and asks an introduction point to deliver it to Bob.

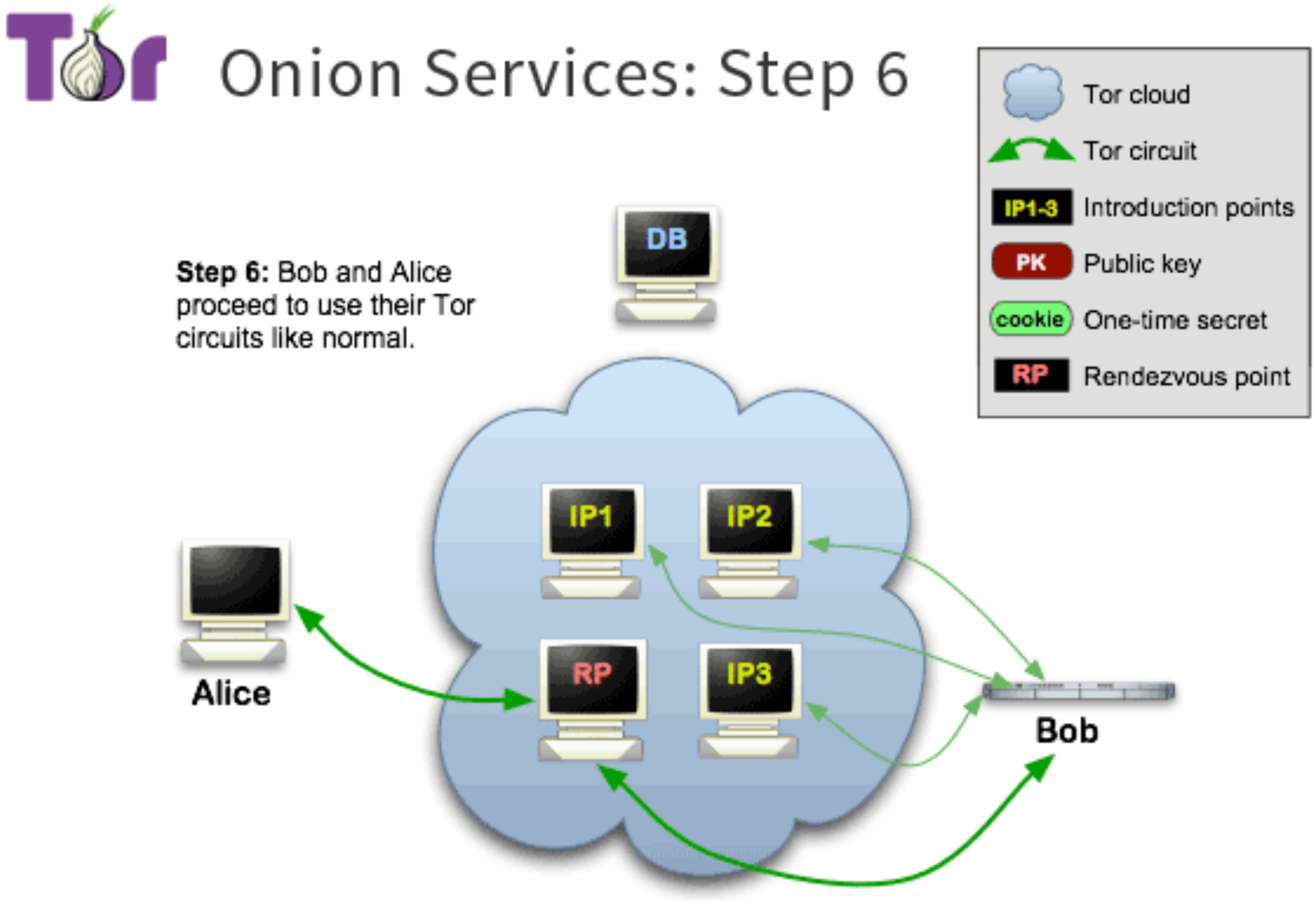
Tor Hidden Services



Onion Services: Step 5

Step 5: Bob connects to the Alice's rendezvous point and provides her one-time secret.

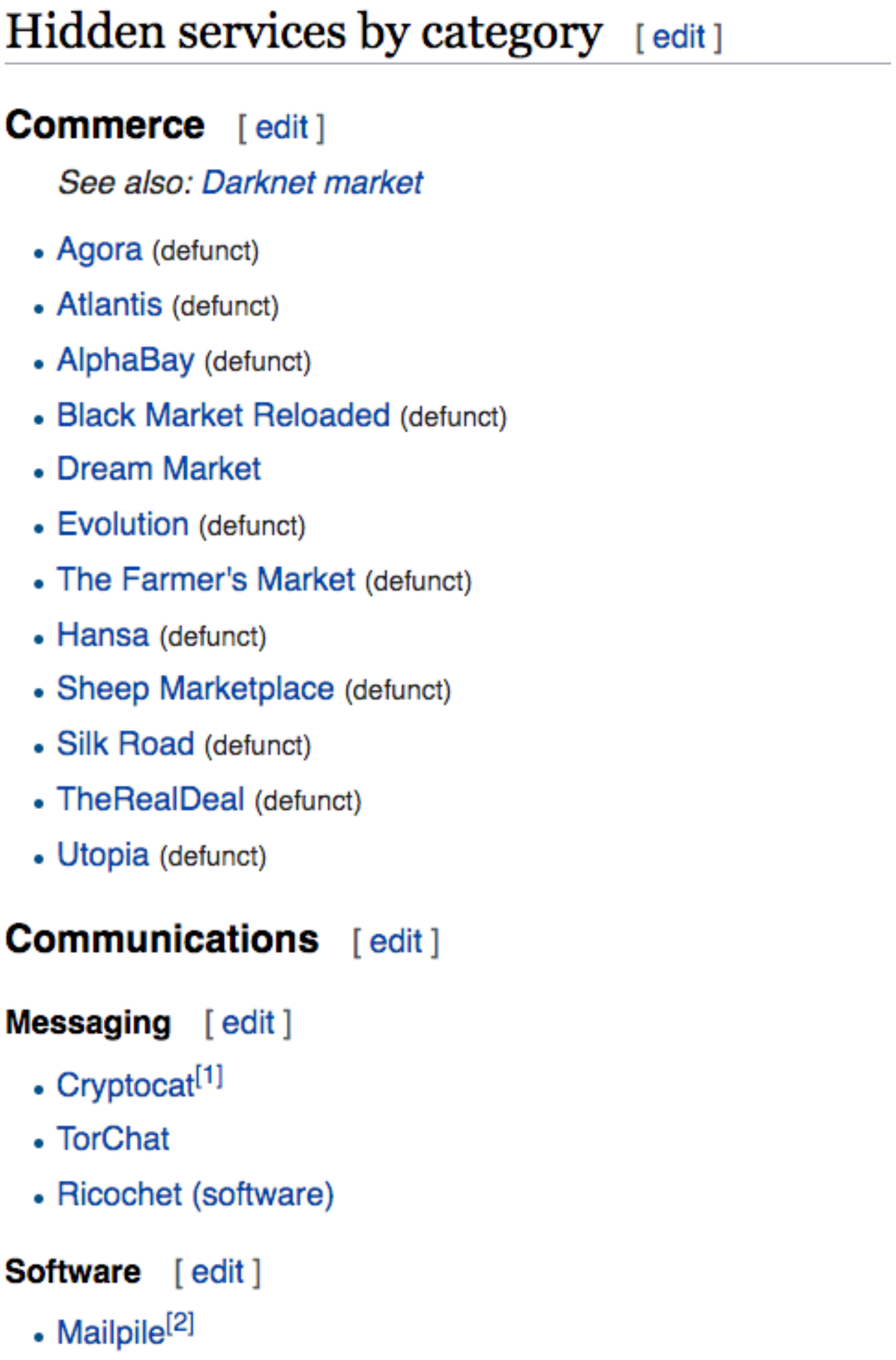
Tor Hidden Services



Onion Services: Step 6

Step 6: Bob and Alice proceed to use their Tor circuits like normal.

Tor Hidden Services



Hidden services by category [edit]

Commerce [edit]

See also: Darknet market

- Agora (defunct)
- Atlantis (defunct)
- AlphaBay (defunct)
- Black Market Reloaded (defunct)
- Dream Market
- Evolution (defunct)
- The Farmer's Market (defunct)
- Hansa (defunct)
- Sheep Marketplace (defunct)
- Silk Road (defunct)
- TheRealDeal (defunct)
- Utopia (defunct)

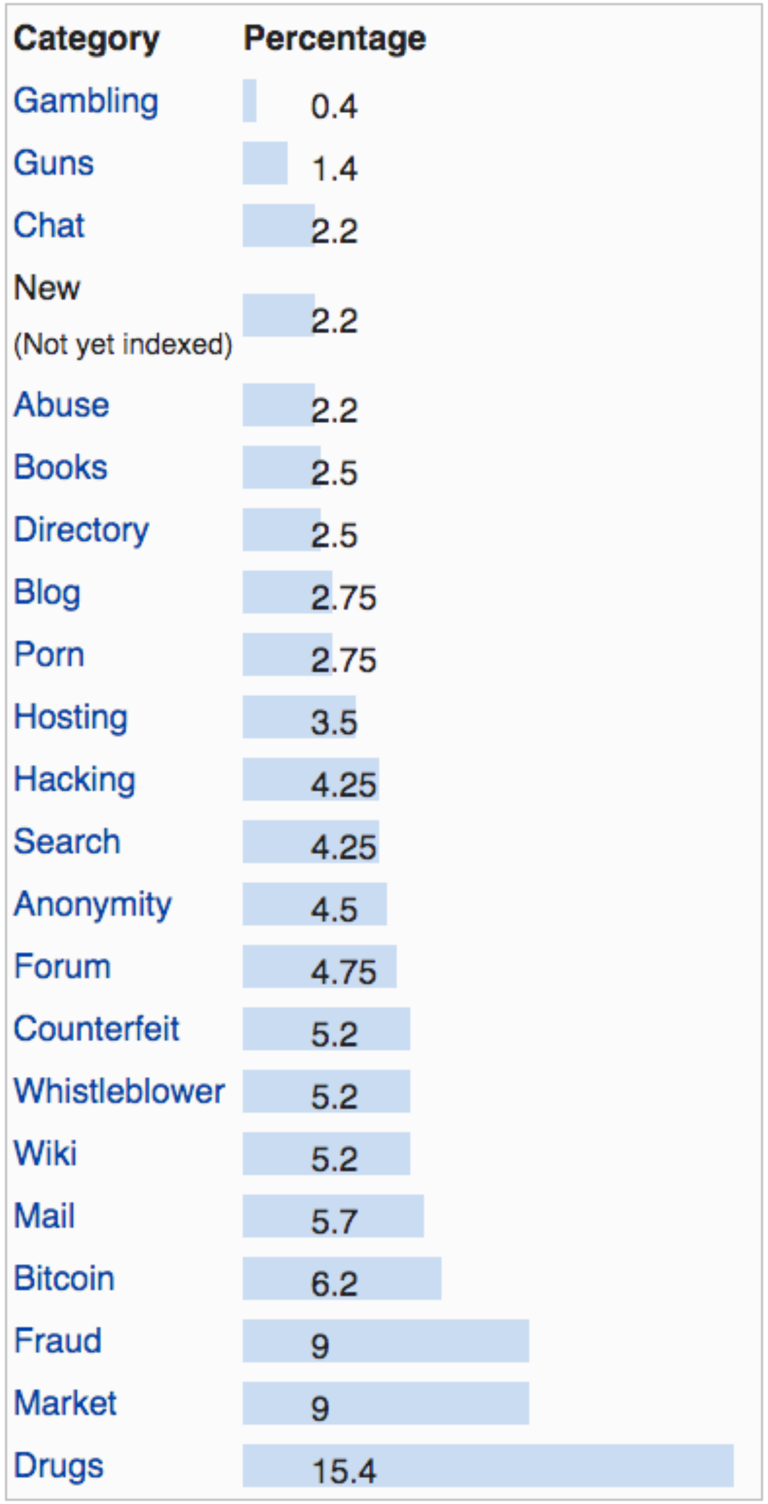
Communications [edit]

Messaging [edit]

- Cryptocat[1]
- TorChat
- Ricochet (software)

Software [edit]

- Mailpile[2]



Category	Percentage
Gambling	0.4
Guns	1.4
Chat	2.2
New (Not yet indexed)	2.2
Abuse	2.2
Books	2.5
Directory	2.5
Blog	2.75
Porn	2.75
Hosting	3.5
Hacking	4.25
Search	4.25
Anonymity	4.5
Forum	4.75
Counterfeit	5.2
Whistleblower	5.2
Wiki	5.2
Mail	5.7
Bitcoin	6.2
Fraud	9
Market	9
Drugs	15.4

Tor Issues

- Performance: latency, bandwidth, …
- Node operators can be enforced by Govs
    - Makes sense to use nodes from different counties
        - Performance?
- Many attacks
    - Malicious/colluding nodes
        - Exit nodes are particularly interesting
    - Timing information between Alice sending and Bob receiving
        - Delay helps to hide it
    - Global adversary observing input and output of the Tor network
        - Tor will not help with that

Private Web Browsing

- Tor provides its own (Firefox-based) browser
- Many tracking methods (besides IP/TCP)
  - JavaScript (I/O, mouse movements, windows layout, …)
  - Cookies, DOM storage, …
  - Headers, credentials, client certificates, …
  - Browser Extensions and Plugins
- Incognito modes, header randomization, JS disabled, Isolating tabs/browsers, clearing cookies and storage, w/o client certificates

Questions ?