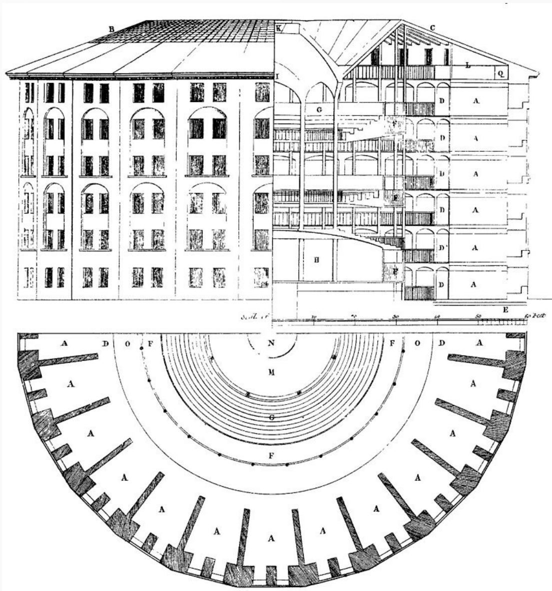
Distributed Computing

A-09. Tor

The Internet Panopticon

- Tracking is **ubiquitous** on the web: **more than 90%** of the websites do some form of tracking
  - Even after the GDPR became law and without user consent
  - Entities that track users share information, to get cross-website information
- Mobile apps probably track even more than websites
  - More difficult to investigate
  - More sensors
- Main player: the advertising industry
  - More covert ones: government agencies, malicious entities…



The Privacy Debate

- How much privacy should people have on the Internet?
- Main argument **against** complete privacy:
  - Crime and terrorism
- Some arguments **for** it:
  - Protecting whistleblowers, activists and journalists
  - Avoiding tracking by totalitarian governments and corporations
  - Avoiding psychological profiling and manipulation
  - Criminals already have access to anonymity
    - Stolen phones, compromised machines

Tor: The Onion Router

- A project initially funded by the USA government (Office of Naval Research & DARPA)
  - Purpose: protecting intelligence communication online
  - **Onion routing** idea published in mid-90s
- Picked up by the Electronic Frontier Foundation—a non-profit for digital rights—in 2004
  - References: Tor website & original design paper

How Tor Works

Onion Routing

- **Layers of encryption**, like those of an onion
- I send a message intended to a destination through three routers: A, B and C.
- Each router “peels” one layer of encryption and sends the rest to the next step
- The message finally gets sent by C to the destination, after removing all the crypto layers
- No router knows **both the source and the destination**

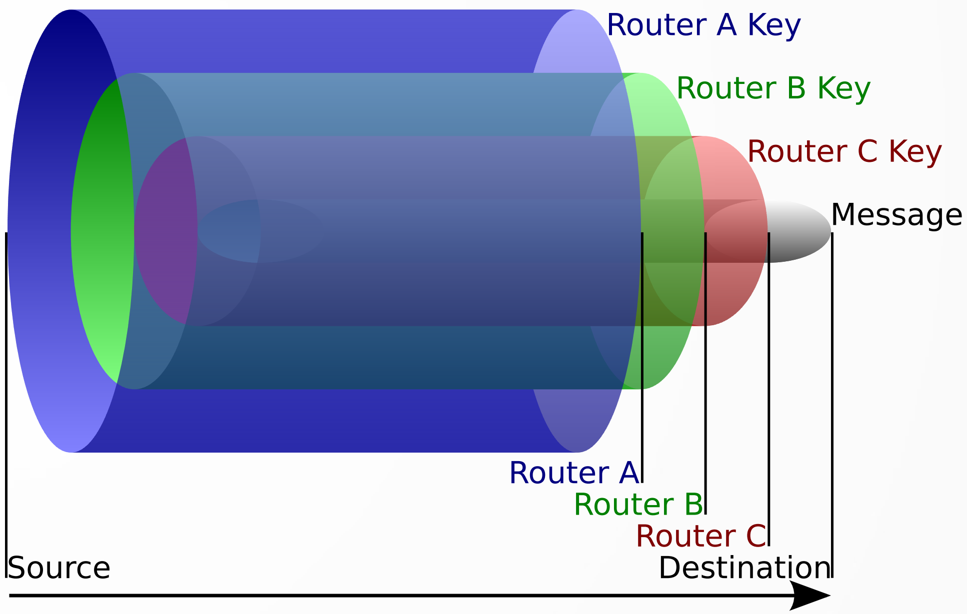
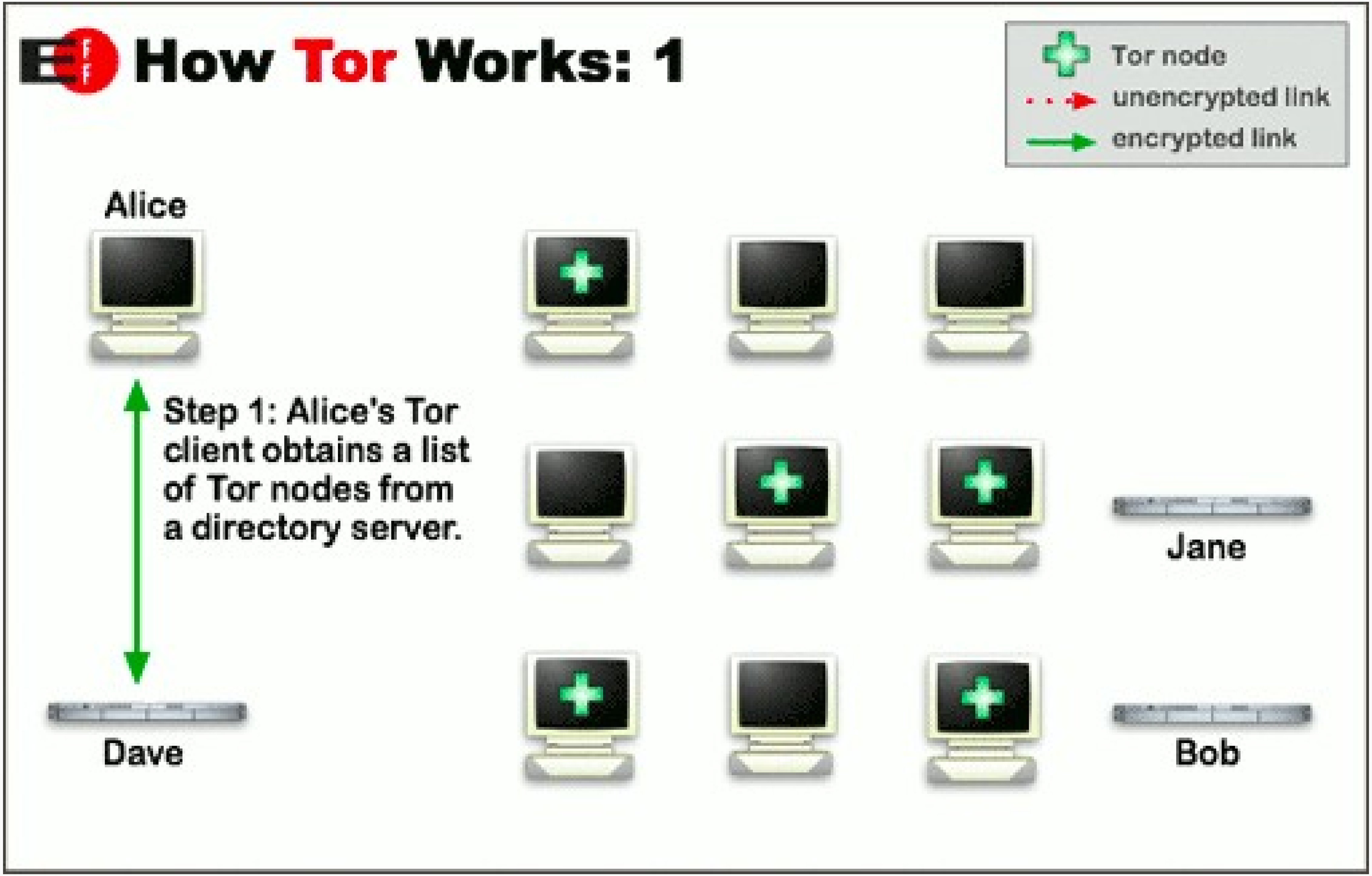
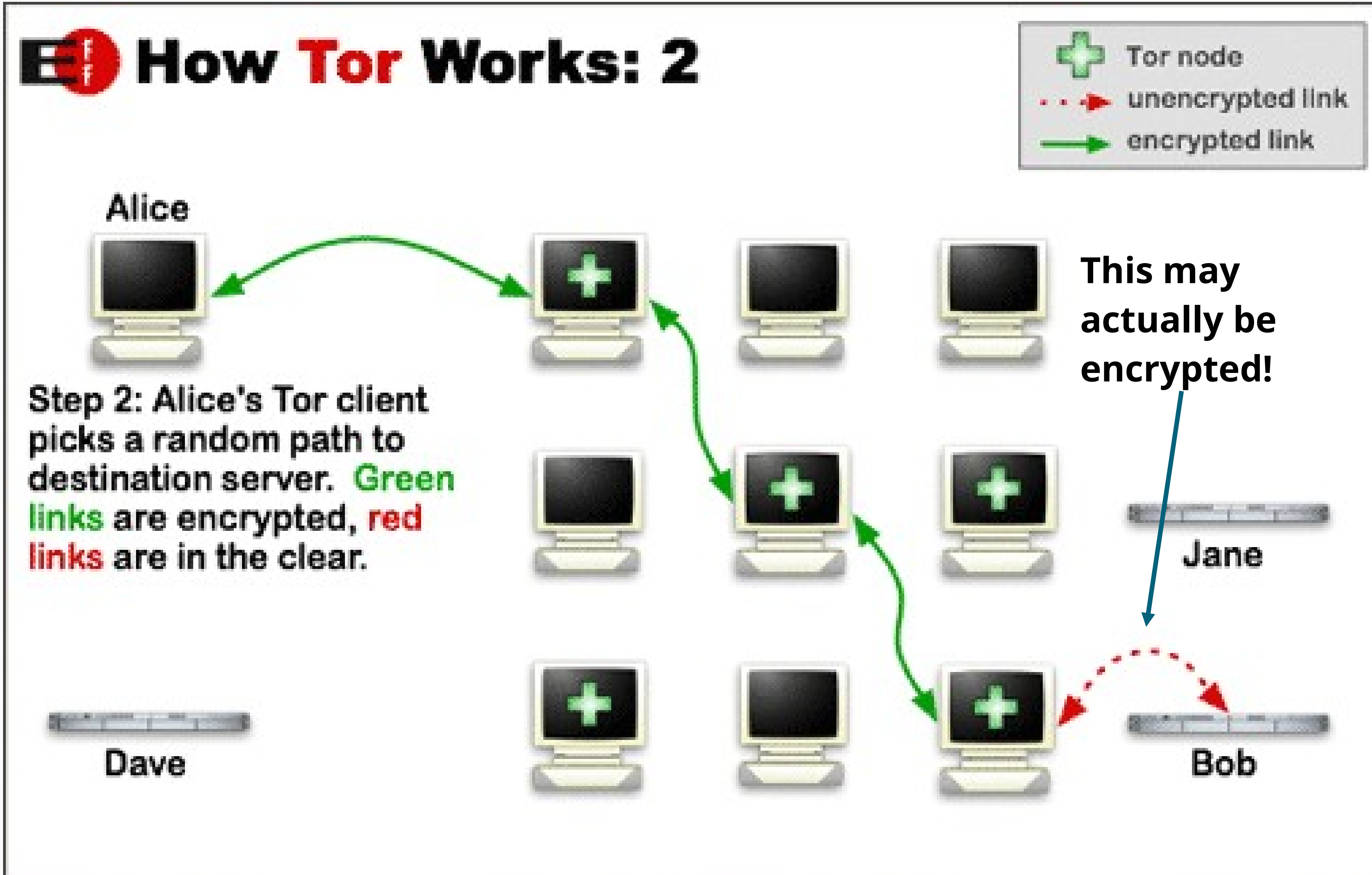
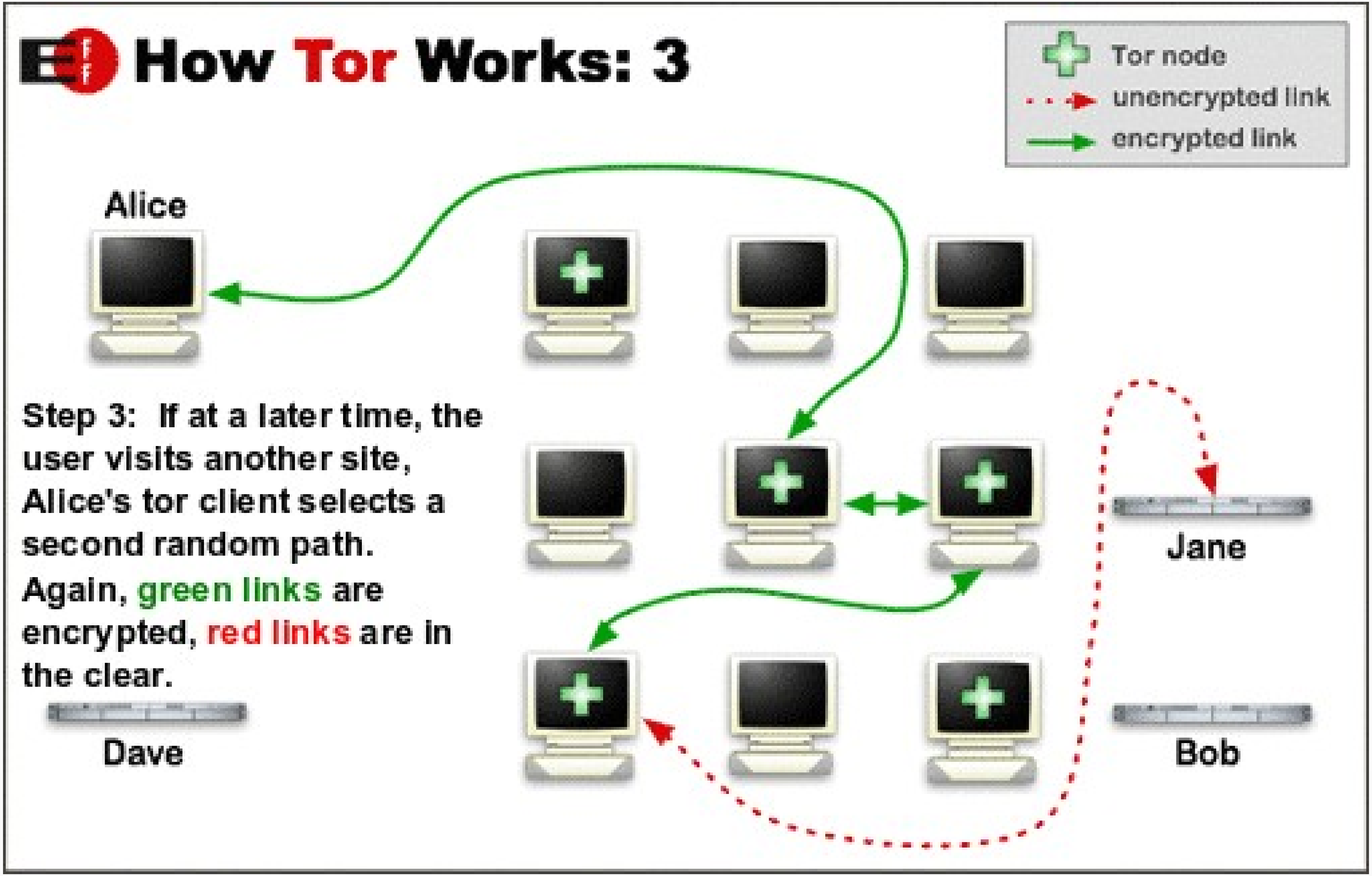


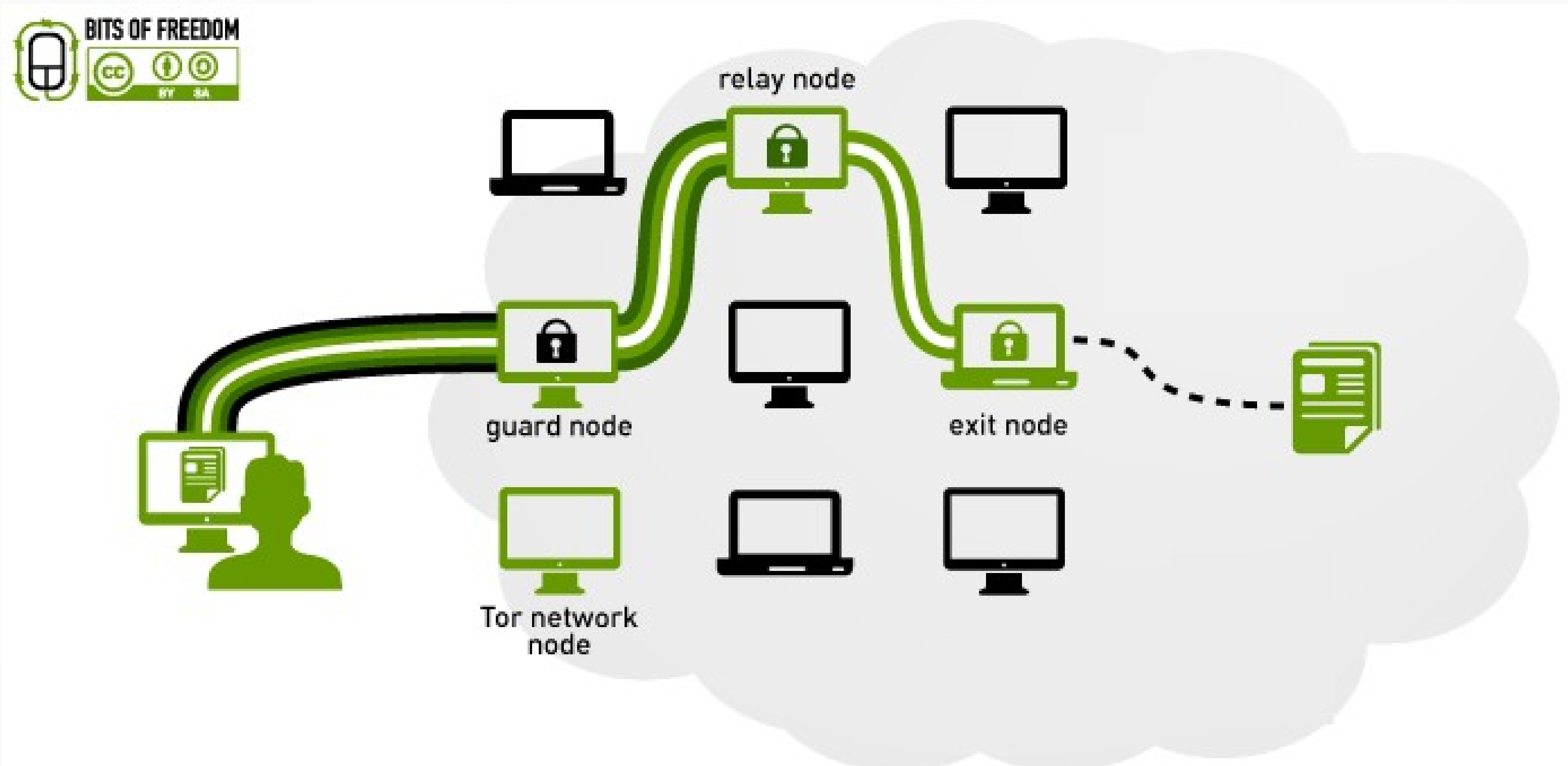
Image by Harrison Neal, CC-BY-SA 3.0


How Tor Works: 1
Tor node
unencrypted link
encrypted link
Alice
Step 1: Alice's Tor client obtains a list of Tor nodes from a directory server.
Jane
Dave
Bob


How Tor Works: 2
Tor node
unencrypted link
encrypted link
Alice
Step 2: Alice's Tor client picks a random path to destination server. Green links are encrypted, red links are in the clear.
This may actually be encrypted!
Jane
Dave
Bob


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

A Bit More Detail



Security Assumptions

- All the routers you choose shouldn't be **owned by the same attacker**
- Attackers can't see your traffic from **both guard and exit nodes**
  - Otherwise it's easy to correlate your traffic

Tor Is a SOCKS Proxy

- SOCKS is a protocol that allows encapsulating TCP connections
  - SOCKS support UDP since version 5, but Tor doesn’t support UDP connections
- You should only use apps that **will use that proxy server**
- Notably: Bittorrent over Tor isn't a good idea
  - It uses UDP

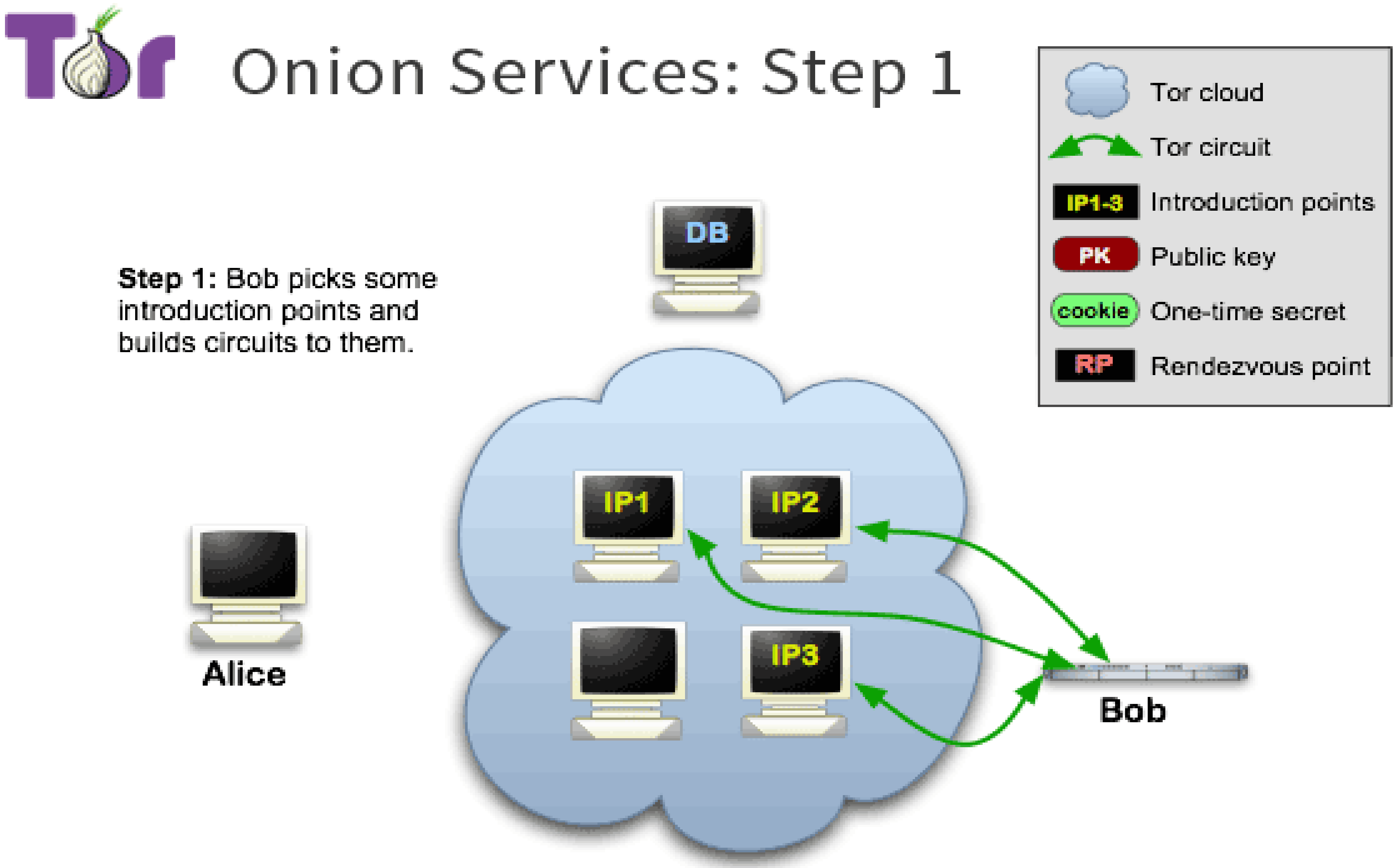
Tor Browser

- Using Tor on your everyday browser is not a great idea
  - **Cookies**: the way most websites track you everyday
    - Website can sync cookies to correlate your visits on different websites
  - **Fingerprinting**: specific information on your hardware/software
    - From OS & configuration information to specific characteristics of your hardware
- The Tor Browser is a hardened Firefox designed to look identical for all users and never exit from the proxy

Bridges

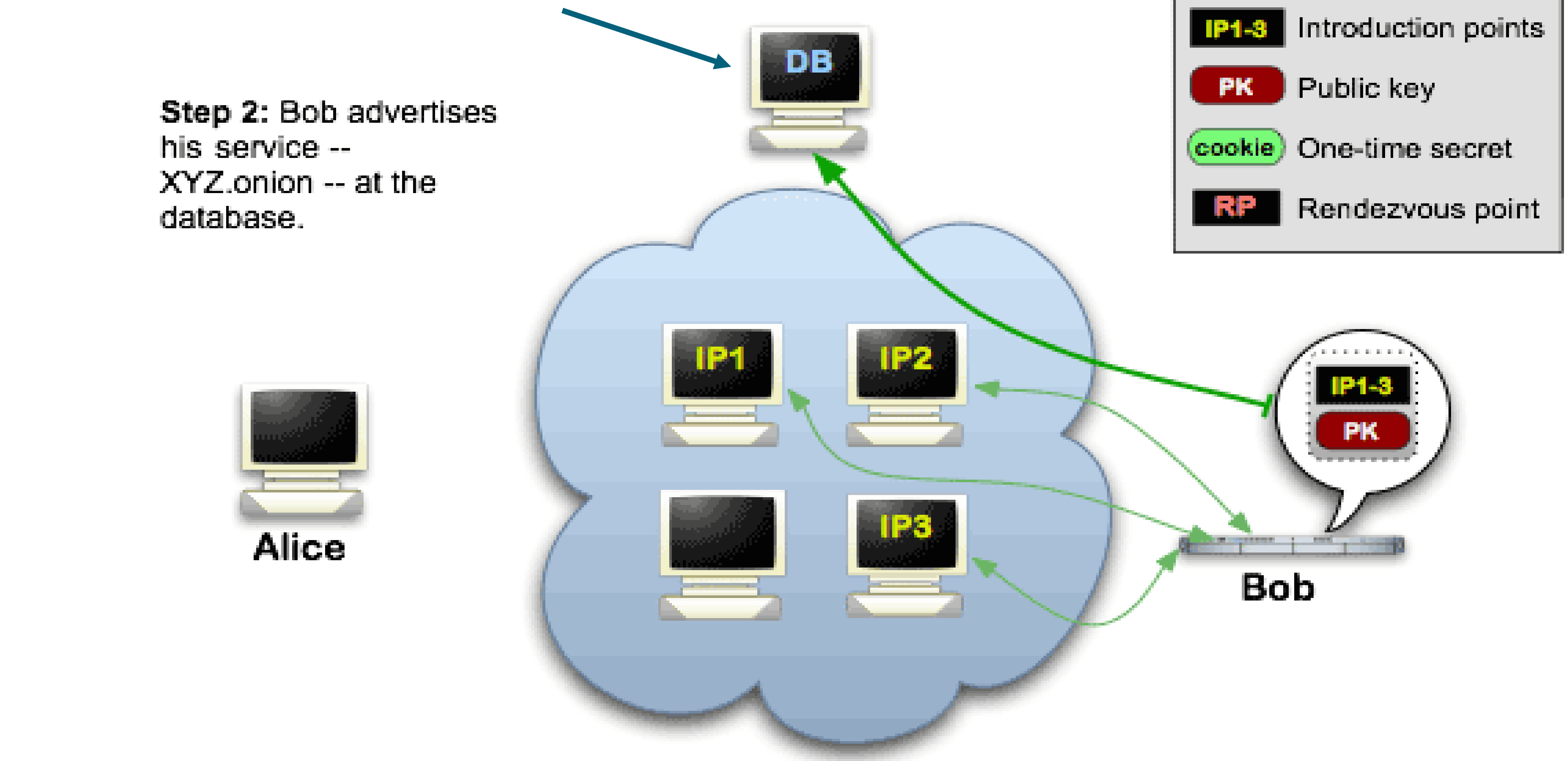
- Tor nodes are **publicly known**
  - Some countries **block access to known Tor nodes**
- **Bridges** exist to allow people to use Tor anyway
  - Non-public
  - People can ask for access to a bridge at a time

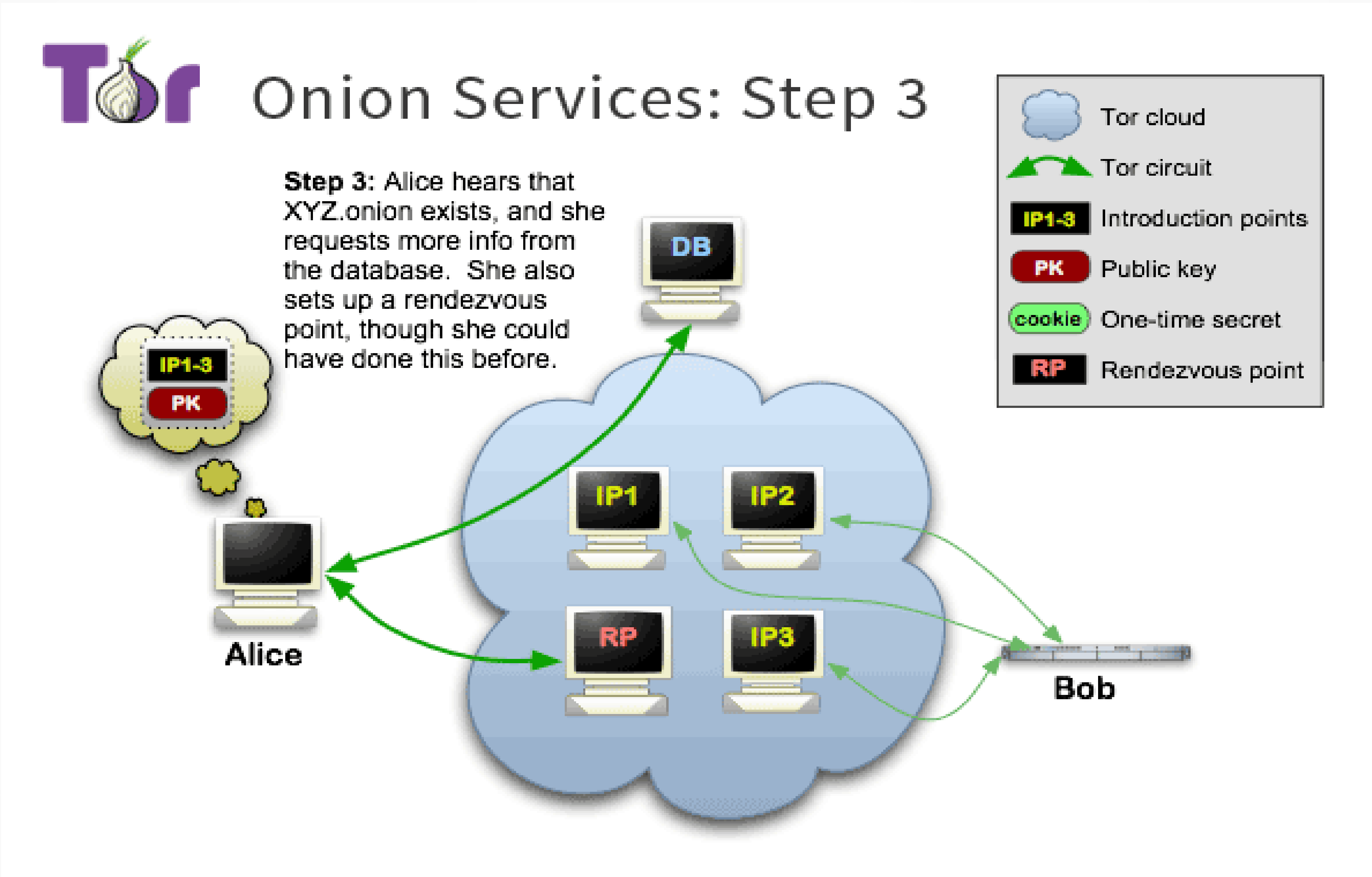
Onion Services (Darknet)


Onion Services: Step 1
Step 1: Bob picks some introduction points and builds circuits to them.
Tor cloud
Tor circuit
IP1-3
Introduction points
PK
Public key
cookie
One-time secret
RP
Rendezvous point
DB
IP1
IP2
IP3
Alice
Bob

Onion Services: Step 2

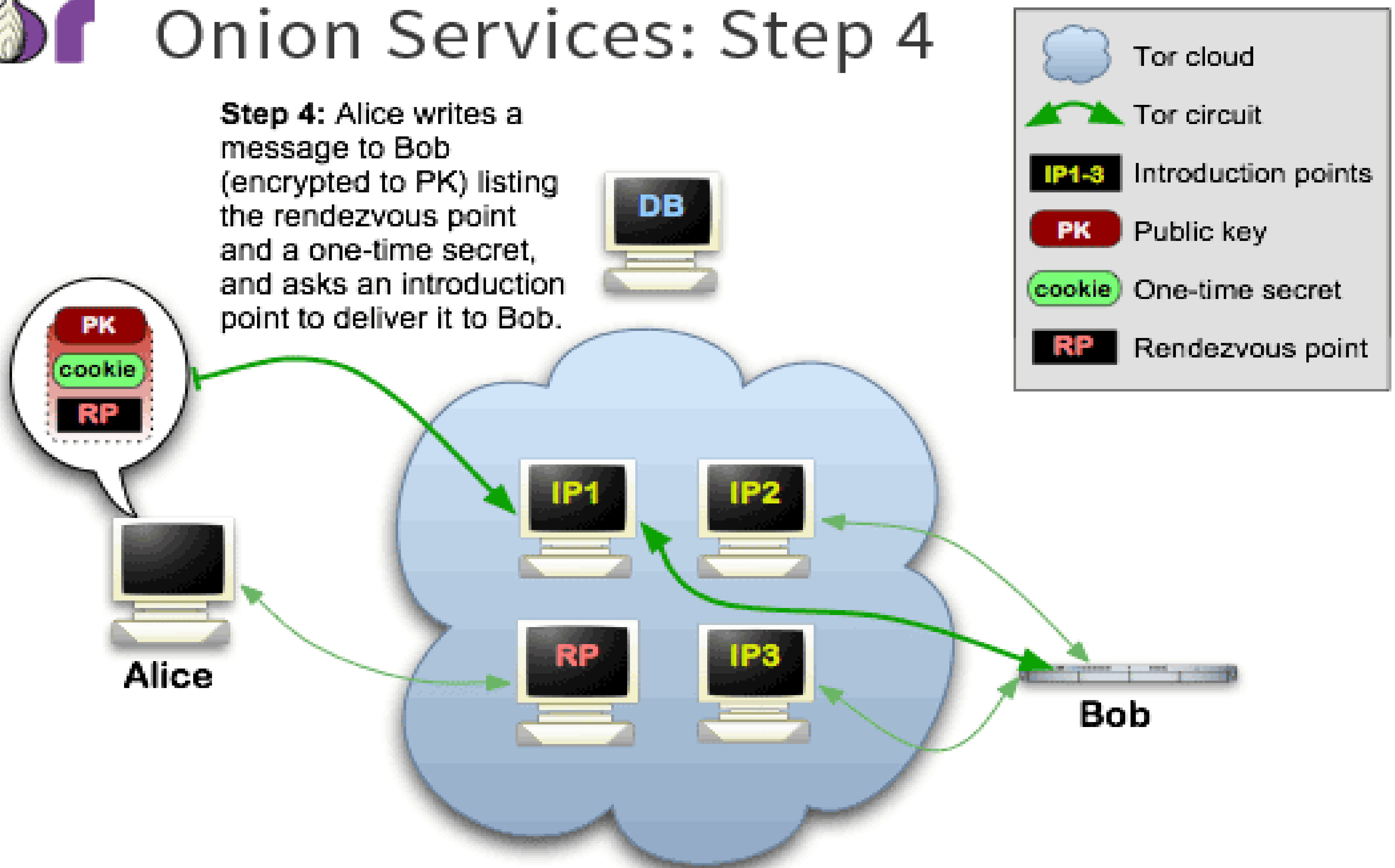
The DB is decentralized... remember the name: **DHT**




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
DB
IP1-3
PK
IP1
IP2
RP
IP3
Alice
Bob

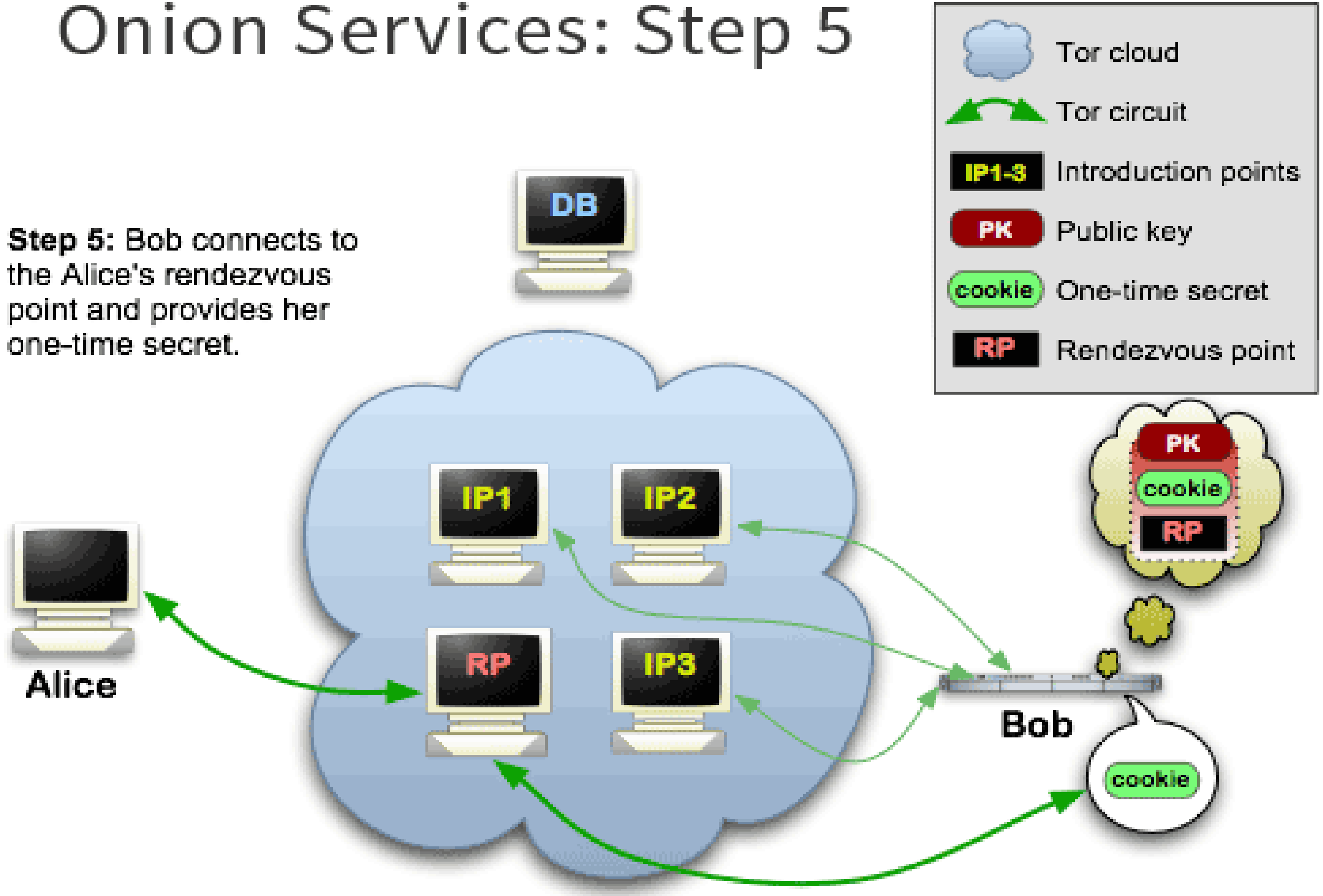
Onion Services: Step 4

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.

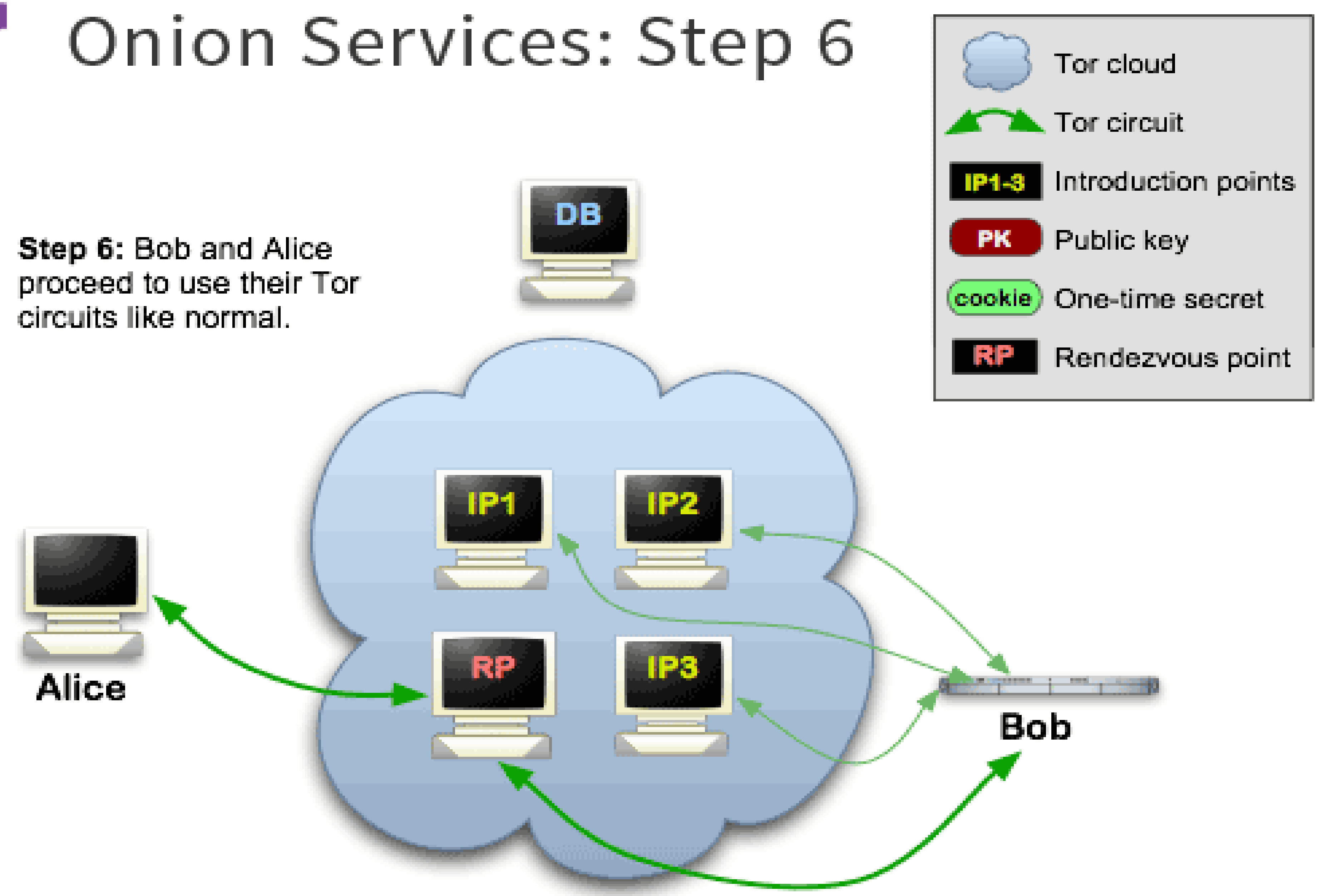


Onion Services: Step 5

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



Onion Services: Step 6



Numbers

How Big Is Tor?

Data from https://metrics.torproject.org/

- ~2M relay users (‘20-’22:2-2.5M, ‘23: 5M)
  - 19% USA, 10% Germany, 5% South Korea
- ~130k bridge users (‘20-’22: 40-80k, ‘23: 120k)
  - 46% Russia, 14% Iran, 9% USA
- ~8k relays (‘20-’23: 6-8k)
- ~2k bridges (‘20-’23: 1.8-2.3k)
- ~300 GB/s GB/s aggregate bandwidth (‘20-23: 250-300)
- ~875 GB/s available (‘20: 500, ‘21: 500-1000, ‘22: 600, ‘23: 900)

What Are Onion Services Used For?

- Source: *"Cryptopolitik and the Darknet"* (Moore & Rid, 2016)

Category	Websites
None	2,482
Other	1,021
Drugs	423
Finance	327
Other illicit	198
Unknown	155
Extremism	140
Illegitimate pornography	122
Nexus	118
Hacking	96
Social	64
Arms	42
Violence	17
Total	5,205
Total active	2,723
Total illicit	1,547

Attacks & Defenses

Tor History and Research

- The security of Tor is **not perfect**
  - We've seen that, by design, powerful attackers can discover information about users
- However, history tells us it's **good enough** in most cases
  - Research looking for its weaknesses
  - Even big agencies like the NSA
  - People were caught because of **mistakes**, not attacking Tor

Discovering Bridges

- Bridges can be discovered (and censored)
- With a full scan of all the IPv4 addresses
  - in 2013, Durumeric et al. (Zmap) discovered 86% of the Tor bridges
- With deep packet inspection (DPI)
  - E.g., the Great Firewall of China recognizes traffic protocols
- Countermeasure: **obfuscation** (pluggable transports)
  - Together with the bridge address you get a secret; protocols like obfs4 and ScrambleSuit hide your protocol to DPI

Website Fingerprinting

- A technique to identify which website a user is looking at by looking at the sizes and timing of encrypted packages
- Tor uses messages of a fixed 512 byte size (“cells”)
  - Together with higher latencies, this makes fingerprinting less efficient
- Many works use a “closed world” hypothesis
  - “Out of these X websites, which one am I visiting”?
  - The real-world fingerprinting problem is more difficult because websites are a lot and change frequently
  - On the other hand, darknet sites are less: attacks to fingerprint them may actually be more relevant

NSA: “Tor Stinks”

- A 2012 presentation
  - Revealed in 2013 among the Snowden documents
- Limited success in attacking it, through
  - Controlling nodes
  - Vulnerabilities
  - Exploiting errors
- *“We will never be able to de-anonymize all Tor users all the time”*

Operation Bayonet

- Suggested reading/listening
  - From *Darknet Diaries*, a podcast about computer security
- The story of two darknet services selling illegal goods, seized by the police of two different countries

Sybil Attack

- Name from a book about a woman with 16 personalities
- In P2P: an attacker creates a **very large number of nodes** to **subvert the system**
- Here, it runs many relays, increasing **likelihood of correlating traffic**
- Countermeasure: **fingerprint** node behavior (joining, uptime, …)
- 2021: a large attack (probably state-sponsored) was discovered