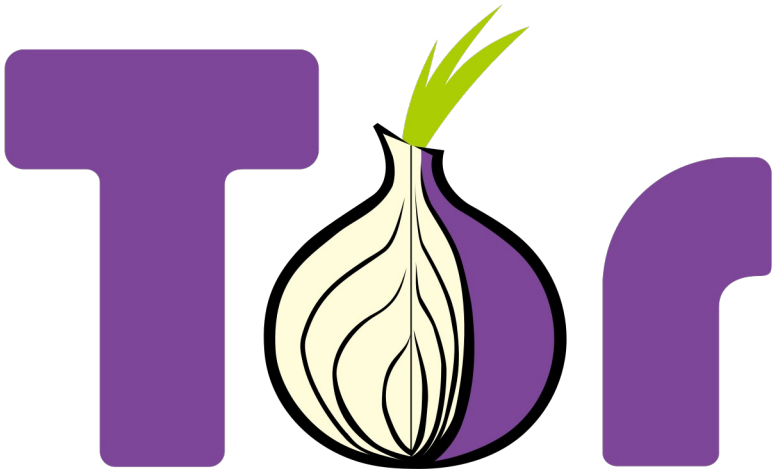
Introduction to Tor Project

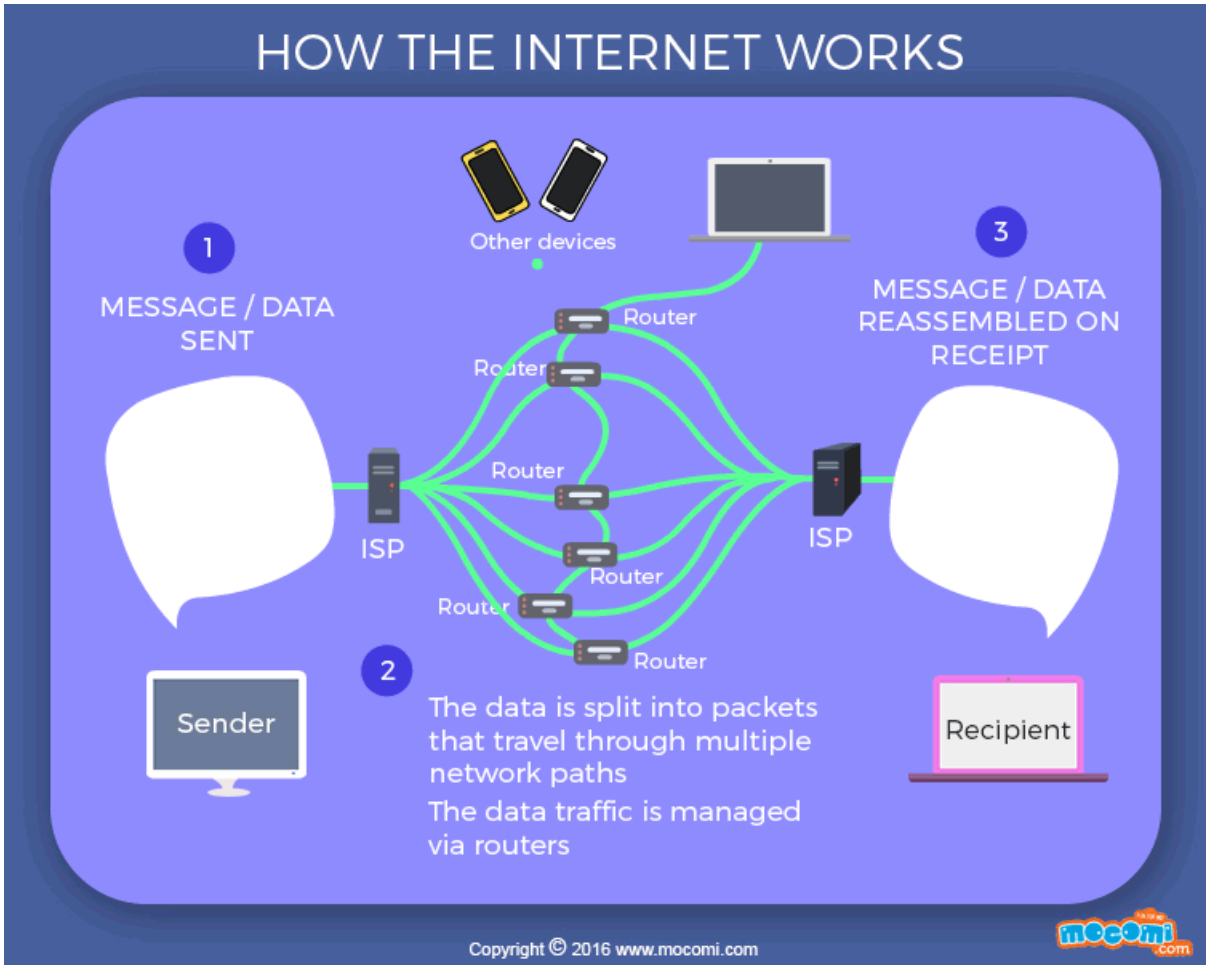


"On the Internet, nobody knows you're a dog."

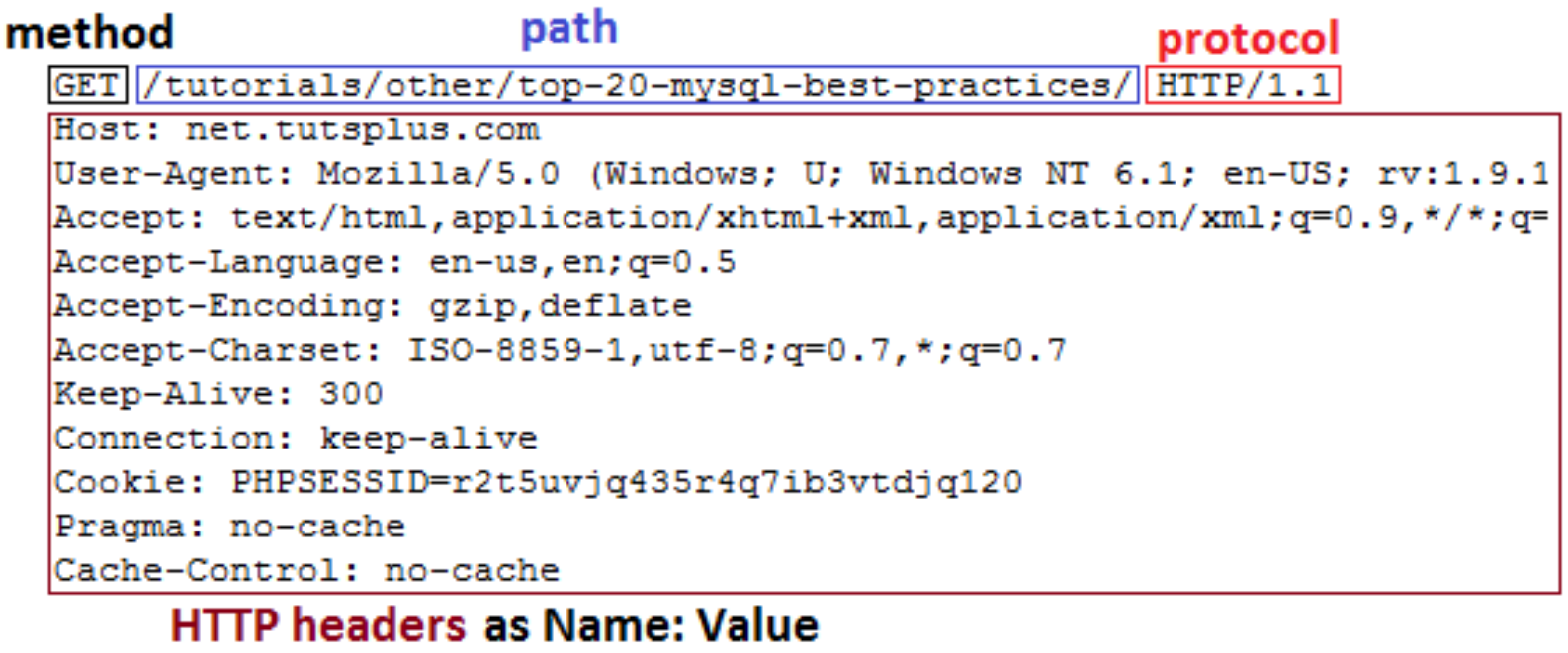
Outline

- How the Internet works
- Packet headers
- Problem
- How Tor works?
- Example of using Tor in Web scrapping
- Questions/Comments

How the Internet works?



Packets


method
path
protocol
GET /tutorials/other/top-20-mysql-best-practices/ HTTP/1.1
Host: net.tutsplus.com
User-Agent: Mozilla/5.0 (Windows; U; Windows NT 6.1; en-US; rv:1.9.1
Accept: text/html,application/xhtml+xml,application/xml;q=0.9,*/*;q=
Accept-Language: en-us,en;q=0.5
Accept-Encoding: gzip,deflate
Accept-Charset: ISO-8859-1,utf-8;q=0.7,*;q=0.7
Keep-Alive: 300
Connection: keep-alive
Cookie: PHPSESSID=r2t5uvjq435r4q7ib3vtdjq120
Pragma: no-cache
Cache-Control: no-cache
HTTP headers as Name: Value

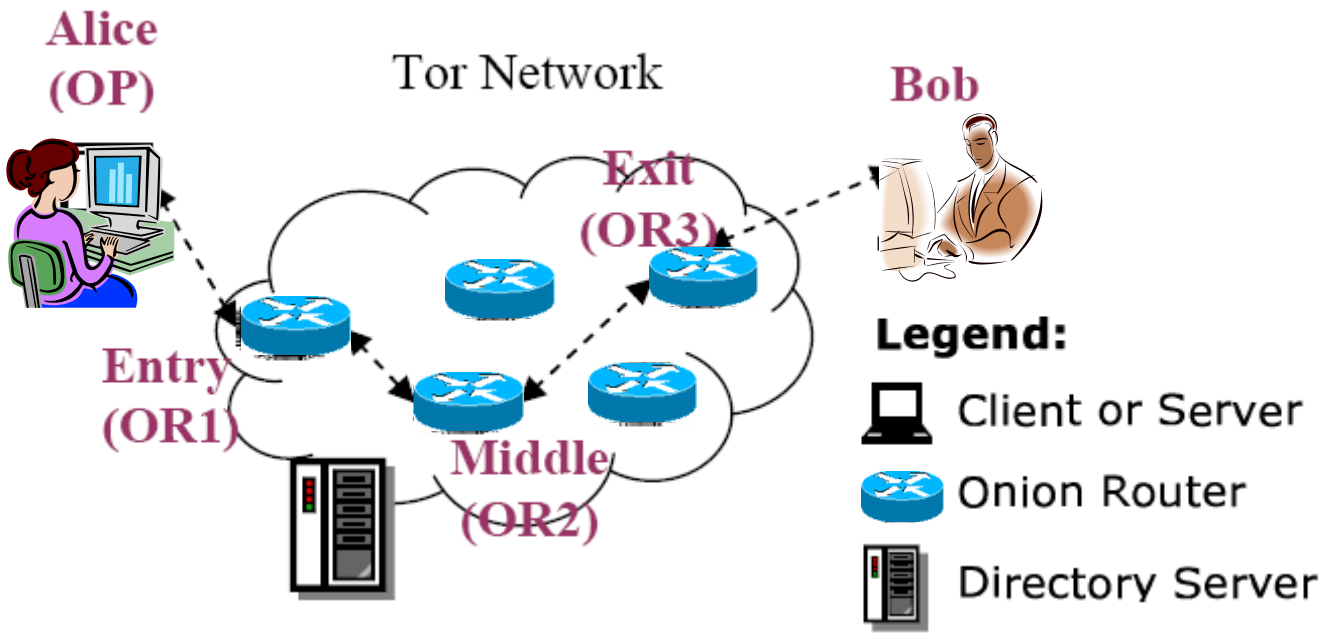
Problem

- Internet surveillance, like traffic analysis, reveals users privacy.
- Encryption does not work, since packet headers still reveal a great deal about users.
- End-to-end anonymity is needed.
- Solution: a distributed, anonymous network

What is Tor?

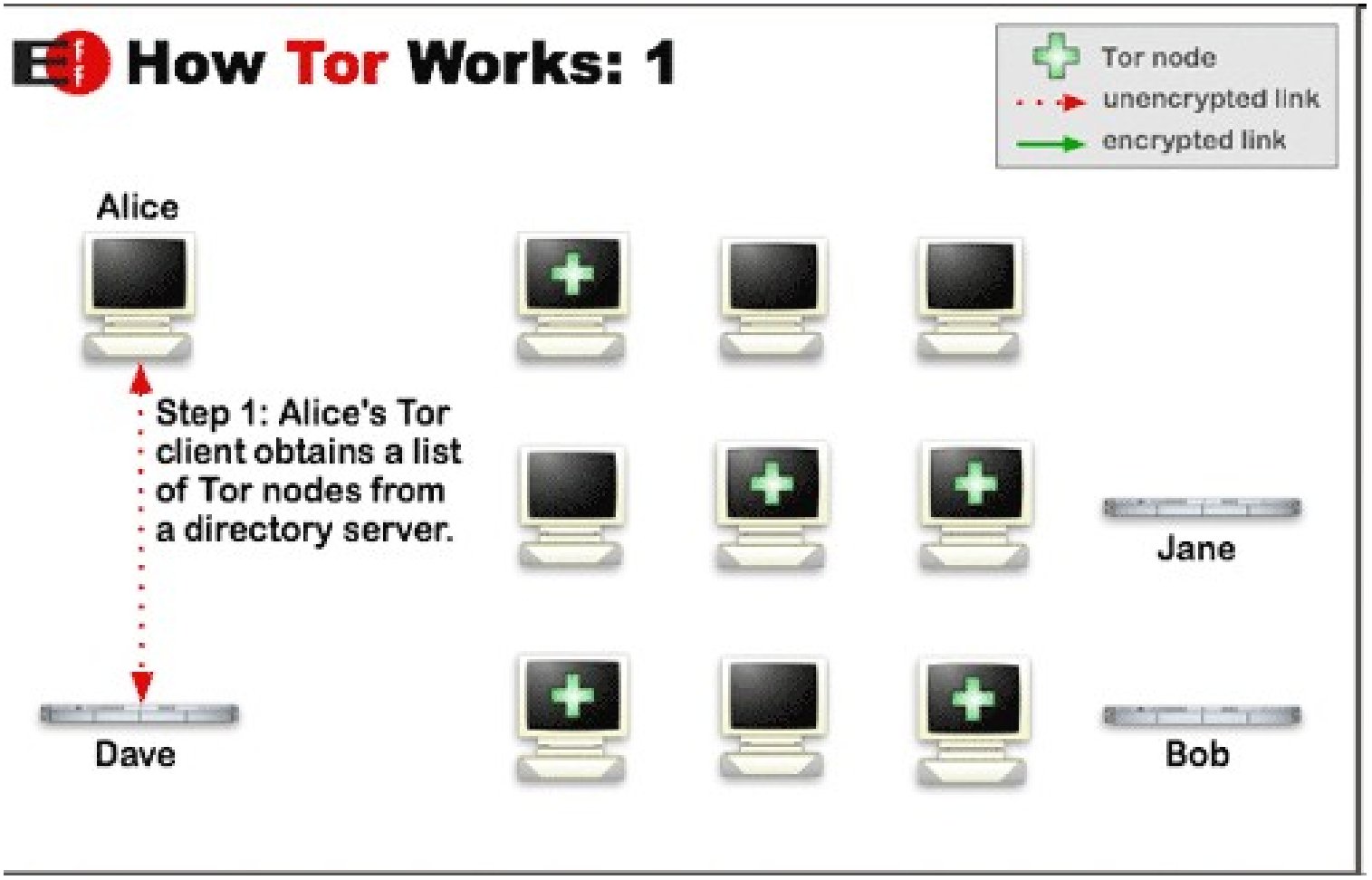
- Tor is comprised of two parts: software you can download that allows you to use the Internet anonymously, and the volunteer network of computers that makes it possible for that software to work..
- Individuals use Tor to keep websites from tracking them, or to connect to those internet services blocked by their local Internet providers.
- Tor's hidden services let users publish web sites and other services without needing to reveal the location of the site.

Components of Tor

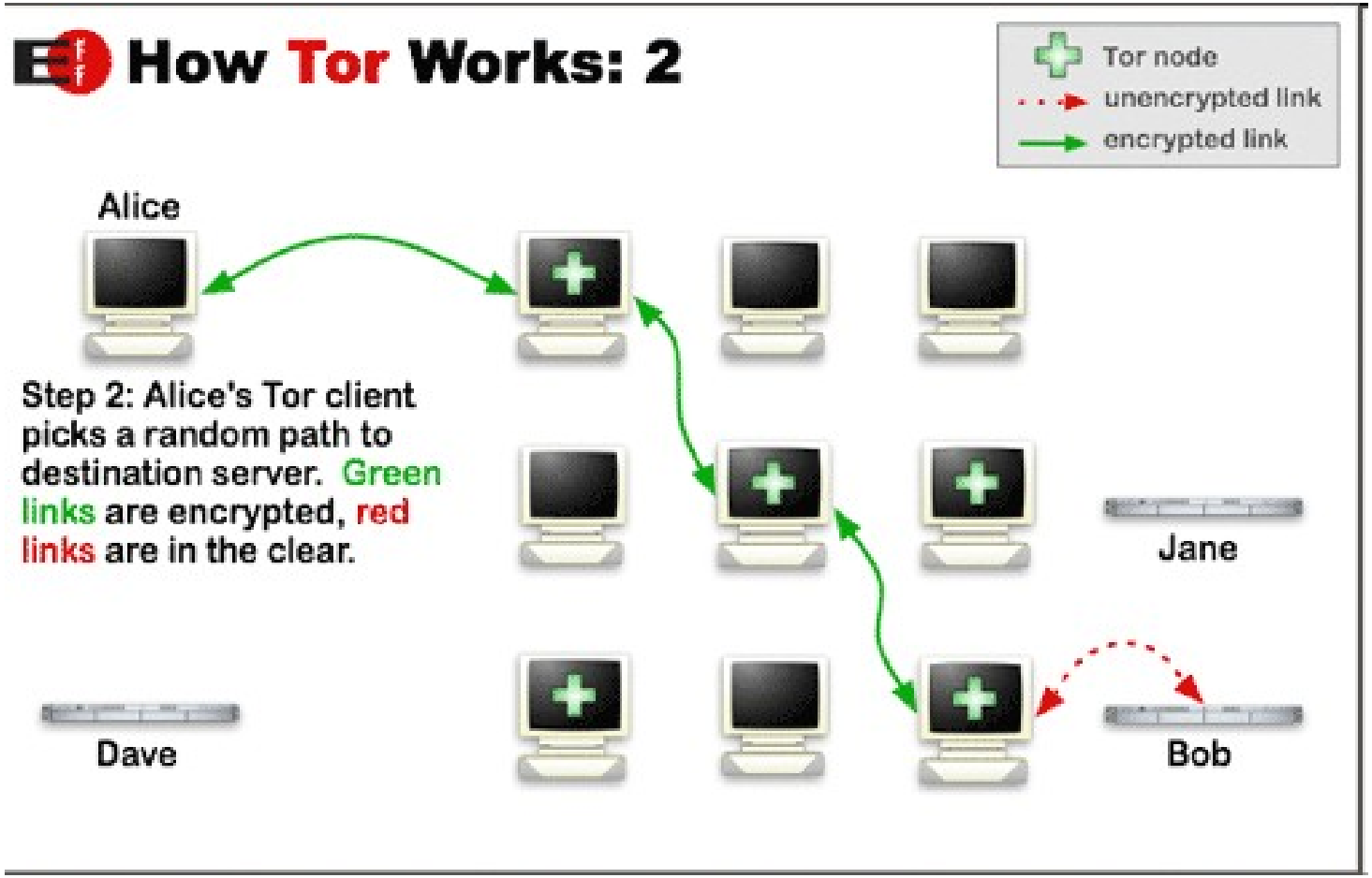


- **Client**: the user of the Tor network
- **Server**: the target TCP applications such as web servers
- **Tor (onion) router**: the special proxy relays the application data
- **Directory server**: servers holding Tor router information

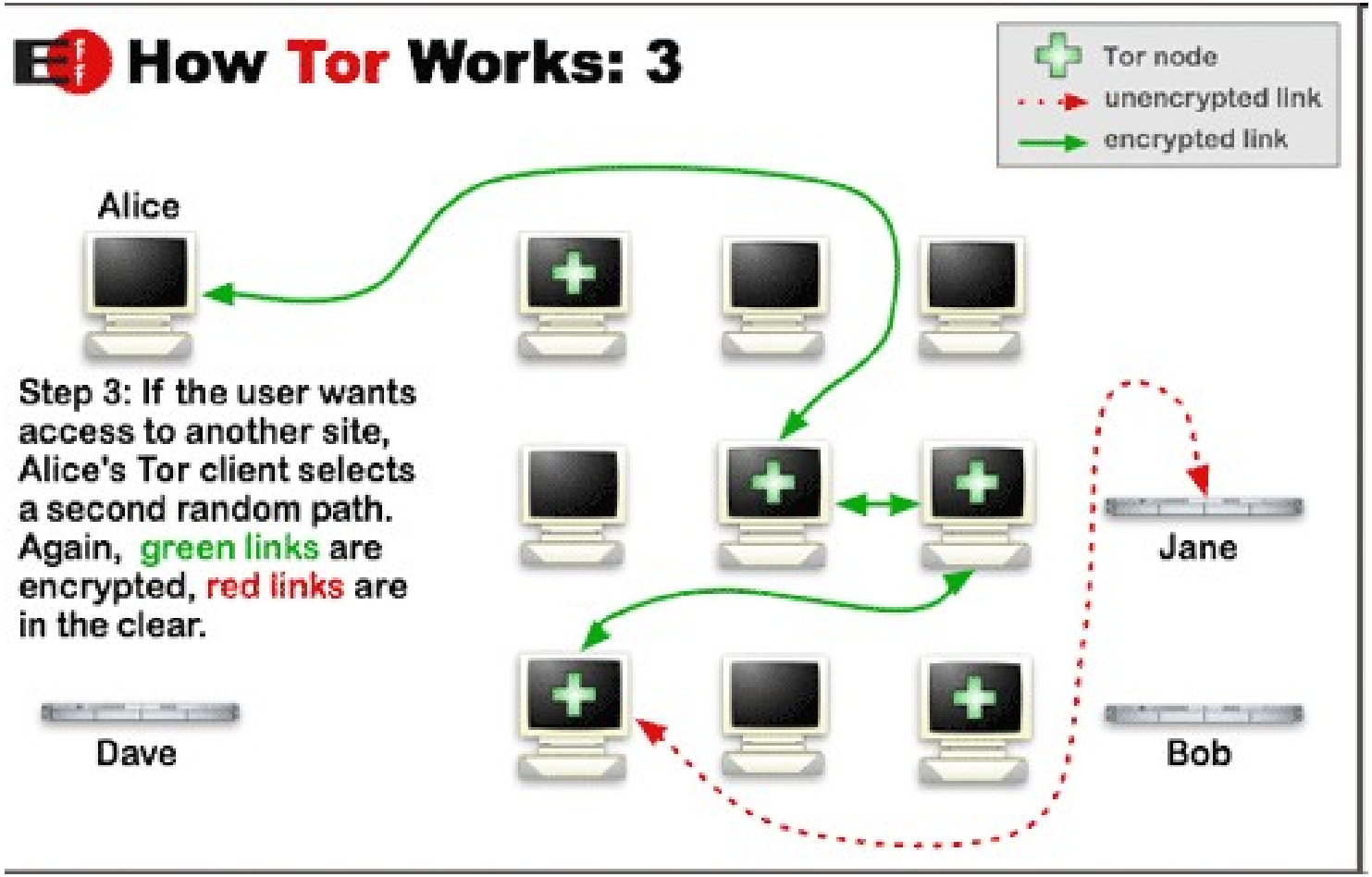
How does Tor work?



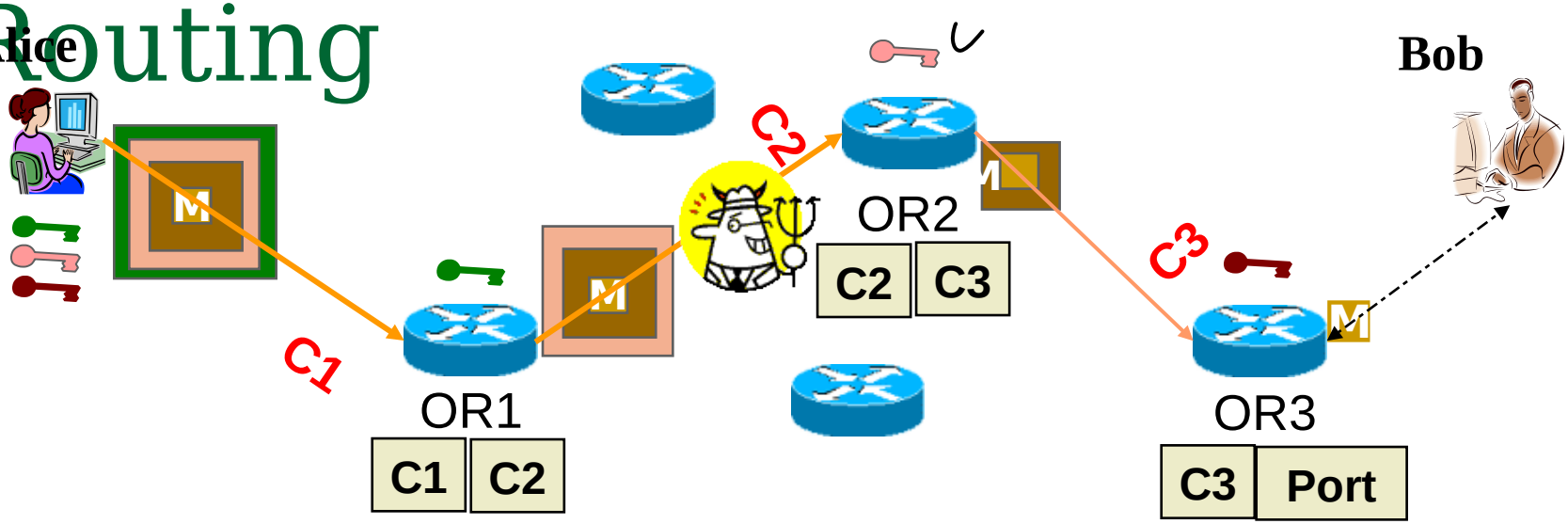
How does Tor work?



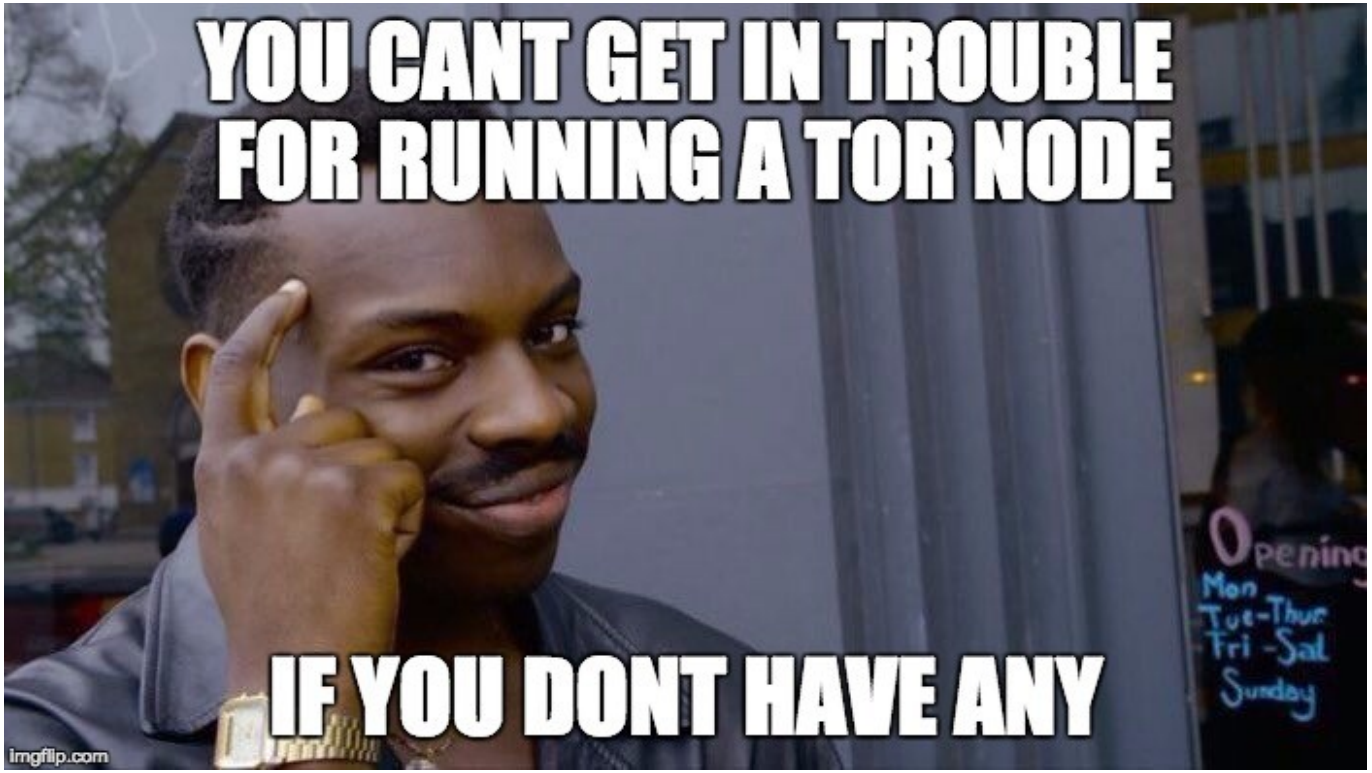
How does Tor work?



How Tor Works? --- Onion Routing



- A circuit is built incrementally one hop by one hop
- Onion-like encryption
  - Alice negotiates an AES key with each router
  - Messages are divided into equal sized cells
  - Each router knows only its predecessor and successor
  - Only the Exit router (OR3) can see the message, however it does not know where the message is from



To Web scrapping using Tor!

Questions/Comments?

- Credit and resources:
  - This presentation is a modification of https://www.cse.unr.edu/~mgunes/cpe401/cpe401sp11/student/tor.ppt
  - https://www.eff.org/torchallenge/what-is-tor.html
  - https://en.wikipedia.org/wiki/Nothing_to_hide_argument
  - https://scholarship.law.gwu.edu/cgi/viewcontent.cgi?article=1159&context=faculty_publications
  - https://support.mozilla.org/en-US/kb/connection-settings-firefox?as=u&utm_source=inproduct
  - https://en.wikipedia.org/wiki/Facebookcorewwwi.onion
  - https://en.wikipedia.org/wiki/On_the_Internet,_nobody_knows_you%27re_a_dog