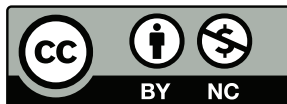


Tor: The Onion Router



- Data that is “about data”
- Many different forms
- Today (and some of next week) are about communications metadata

Communications Metadata

- Who is talking to whom, when, and for how long
- Visible even if the content is encrypted
- Originally developed during World War I as an adjunct to cryptanalysis
- Later, as *traffic analysis*, it became very important as an intelligence field of its own
- Michael Hayden: “We kill people based on metadata”

The EFF's Examples

- They know you rang a phone sex service at 2:24 am and spoke for 18 minutes.
- They know you called the suicide prevention hotline from the Golden Gate Bridge.
- They know you spoke with an HIV testing service, then your doctor, then your health insurance company in the same hour.
- They know you received a call from the local NRA office while it was having a campaign against gun legislation, and then called your senators and congressional representatives immediately after.
- They know you called a gynecologist, spoke for a half hour, and then called the local Planned Parenthood's number later that day.

(From <https://www.eff.org/deeplinks/2013/06/why-metadata-matters>)

Traffic Analysis

- A person from a suspicious country has a pattern: a short Skype call at the same time every week
- One week, there's a long call
- The next week, there's nothing...

Traffic Analysis

- A person from a suspicious country has a pattern: a short Skype call at the same time every week
- One week, there's a long call
- The next week, there's nothing...
- An agent checking in, a planning call, then the plan is in motion?

Traffic Analysis

- A person from a suspicious country has a pattern: a short Skype call at the same time every week
- One week, there's a long call
- The next week, there's nothing. . .
- An agent checking in, a planning call, then the plan is in motion?
- Or a student hearing from home, learning of an illness, and then hurriedly returning?

Mixnets (Chaum, 1981)

- Assume a group of computers A, B, C, D, E, F, G, H
- Every five minutes, the following communications take place: A sends 10KB of encrypted data to B ; B then sends 10KB of ciphertext to C , etc., ending in H sending to A
- Who is talking to whom? Was it real data or dummy padding data?
- If the lengths and timing are the same, you can't tell
- But: it's inefficient: a lot of dummy traffic, and slow actual transmission

Threat Models

- Who is your enemy?
- Can they actually monitor all of the links for those eight communications paths?
- Could they if we replaced the deterministic pattern with a random one?
- What is the actual threat model?

Tor: The Onion Router

- Assume a *limited* adversary
- The adversary can listen to *some* links, but not all links, and not all the time
- Answer: Tor
- (Tor was invented at the Naval Research Lab—the military understands the need to foil traffic analysis)

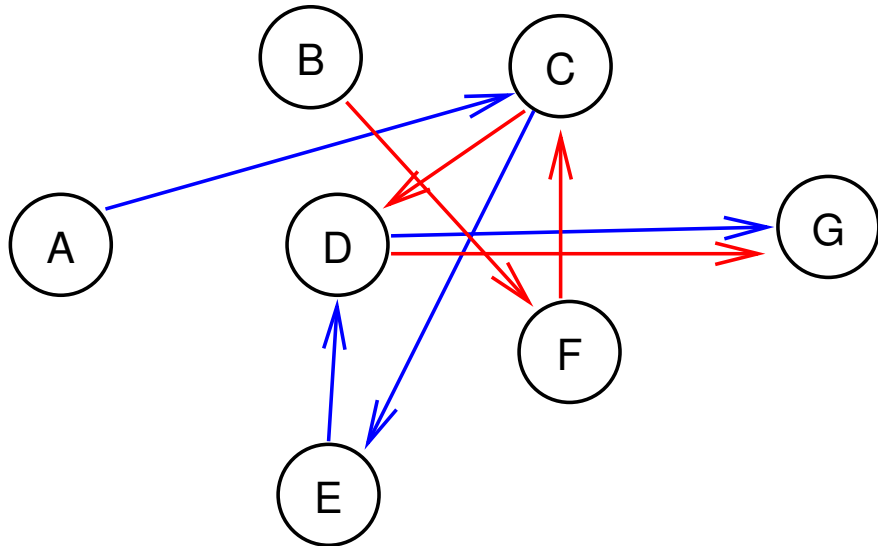
- There are many ways to identify users and servers on the Internet
- One is IP address—every computer that talks on the net needs an IP address
- ISPs know who owns an IP address at a given time
- Governments can obtain that information if they wish to surveil or harass users
- The IP addresses are public, for governments that want to block certain sites

- Assumption: many clients
- Assumption: eavesdropping possible
- But—the adversary isn't *global*
- That is, it can monitor many links but not *all*

How it Works

- A client computer picks a set of “relay nodes” and an “exit node”
- (All of these nodes are volunteers)
- The client sends the traffic to the first node, which sends it to the second, etc.; the exit node forwards it to the real destination
- (Often, only one relay node is used)
- The set of Tor nodes used, including the exit node, is changed frequently
- In other words, the source IP address is short-lived

Multiple Hops



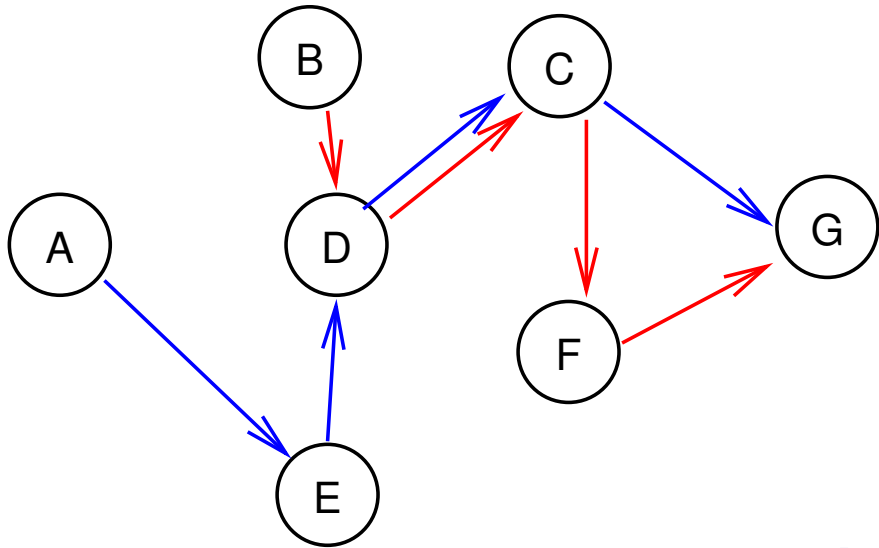
Onion Routing

- **G** thinks that both connections are coming from **D**
- The real sources—**A** and **B** — are hidden
- On subsequent visits, **C** and **Z** may be the exit nodes
- Intuitive understanding: nested envelopes

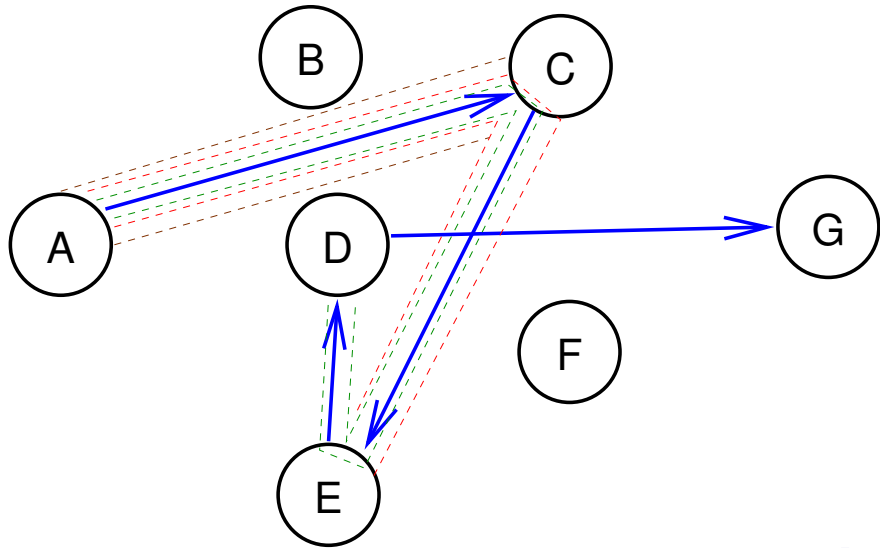
Why Multiple Hops?

- If someone is spying on **D** or its links, they'll see where traffic is coming from
- Here, though, traffic is coming from **E** and **C** — which is which?
- Can the same attacker spy on **E** and **C**?
- Remember that the path will switch soon

Change Paths Frequently



Using Cryptography



What is Known

- Each node knows only the previous and next hops
- Nodes do not know where on the path they are
- Only the exit nodes knows the destination
- Only the entrance node knows the source
- Intuitive understanding: nested sealed envelopes; each hop adds its own return address

Anonymous Browsing

- With Tor, it is possible to browse the web without being identified
- It's great for dissidents in oppressive countries
- It's also great for spies, law enforcement investigations, etc.
- No accountability...

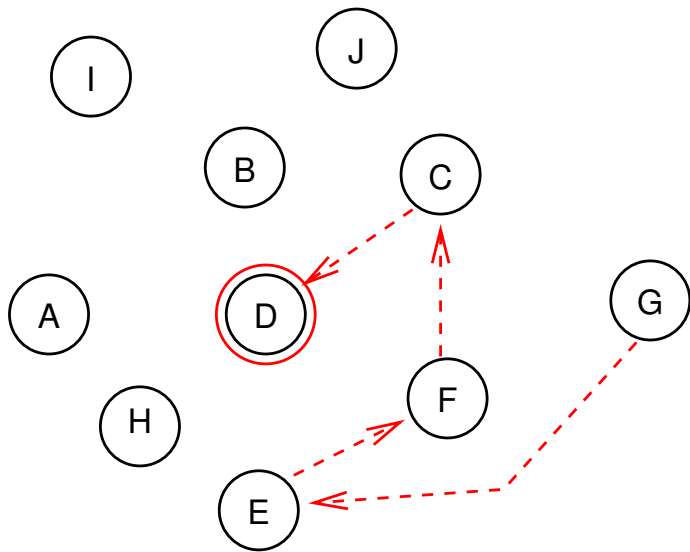
What About Servers?

- Servers traditionally live at a known IP address
- But Tor is designed to hide IP addresses—even the exit nodes don't know the user's real IP address
- Even if we solve that problem, what about authenticity? How does the Tor network know which is the real claimant to some service?

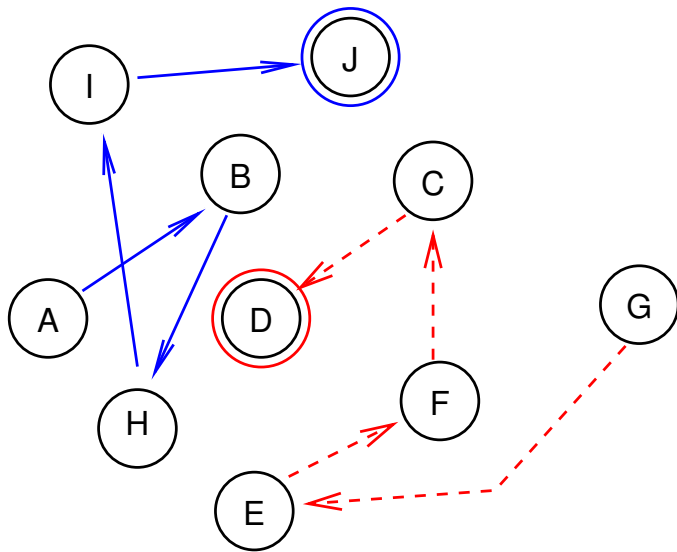
Tor Hidden Services

- The server operator picks some set of Tor nodes as *introduction points*
- These nodes are registered in a distributed directory
- A client node opens a Tor service to some random Tor node, and uses it as a *rendezvous point*
- The client sends the address of its rendezvous point to the server's introduction point
- The server opens a Tor circuit to the rendezvous point
- The rendezvous node forwards traffic between the two Tor services

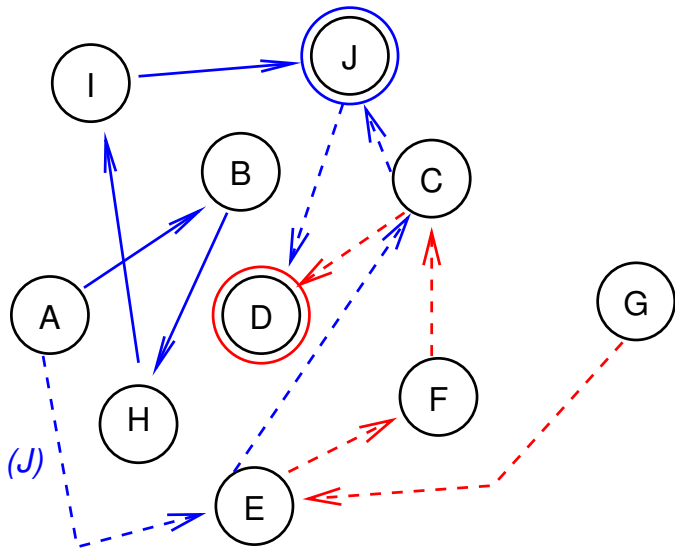
Creating an Introduction Point



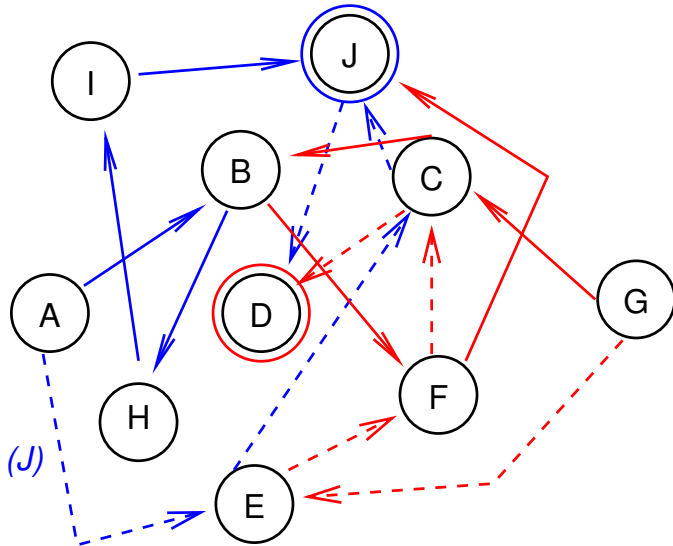
Creating an Rendezvous Point



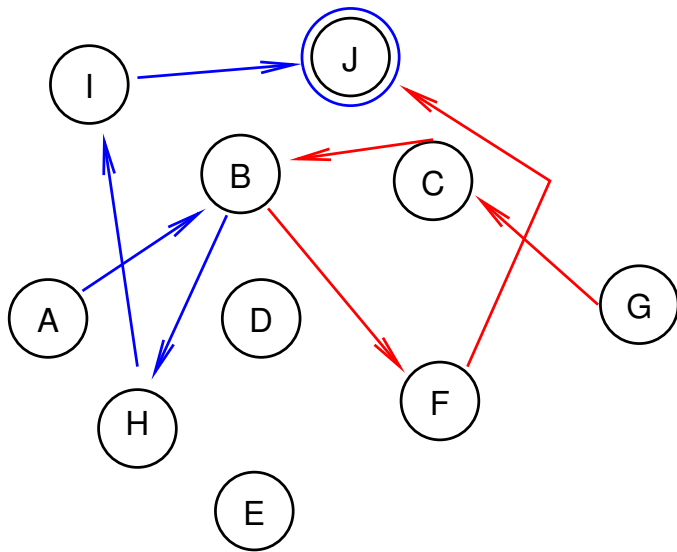
Notifying the Introduction Point



Traffic Can Flow



With the Setup Messages Deleted...



Authenticating Hidden Services

- The server generates a key pair
- The private key is used to sign all of its announcements, e.g., of the introduction points
- The server's name is formed from a hash of the public key
- In other words, you cannot have arbitrary “.onion” names—but you can keep generating keys until you get one you like

Generating .Onion Names

- Generate a key pair
- Take the SHA-1 hash of the public key, and truncate it to 80 bits
- Represent the truncated hash in base 32, using 26 letters and 6 digits
- If you don't like the result, try again

Facebook's Hidden Service Certificate

Certificate:

Data:

Version: 3 (0x2)

Serial Number:

0e:87:85:21:62:33:85:ea:90:2d:16:5d:81:7f:37:1b

Signature Algorithm: sha256WithRSAEncryption

Issuer: C=US, O=DigiCert Inc, OU=www.digicert.com, CN=DigiCert SHA2 Extended

Validation Server CA

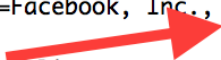
Validity

Not Before: Sep 25 00:00:00 2015 GMT

Not After : Nov 28 12:00:00 2016 GMT

Subject: businessCategory=Private Organization/1.3.6.1.4.1.311.60.2.1.3=US/1.3.6.1.4.1.311.60.2.1.2=Delaware/serialNumber=3835815/street=1601 Willow Rd./postalCode=94025, C=US, ST=CA, L=Menlo Park, O=Facebook, Inc., CN=*.facebookcorewwwi.onion

Subject Public Key Info:



How Did They Get facebookcorewwi.onion?

- The prefix 'facebook' is 8 characters—40 bits, if a base 32 number
- Generating a key whose hash has the first 40 bits of that string takes $O(2^{40})$ tries
- They then looked at the candidate names for one that had a suffix—"corewwi"—for which they could construct a plausible story
- Facebook has *lots* of computers. . .

Why Does Facebook Use Tor?

- Facebook, of course, wants to learn lots about its users
- Why should it like Tor?
- Some countries, notably Iran, were blocking Facebook—but not Tor
- They noticed that many of their Iranian users were connecting over Tor, so they decided to make it work properly

The Dark Web

- There are other services that use Tor hidden services as well
- Some of them are rather less benign than Facebook

The Silk Road

- An online drug, etc., market place
- Created by “Dread Pirate Roberts” (DPR), later shown to be Ross Ulbricht
- More of an EBay than an Amazon—the site hosted independent sellers
- Payment was in Bitcoin; delivery was by UPS, FedEx, etc.
- DPR also solicited murders of former lieutenants he thought had betrayed him

The Fall of the Silk Road

- The FBI—somehow!—located the physical server, in Iceland
- Assorted Federal agents wormed their way into DPR's confidence—after all, it was all online, anonymous activity—and became assistant site admins
- Early on, Ulbricht had posted a query to Stack Overflow on setting up Tor services—and he used his own name
- He was arrested in a San Francisco library, while online as DPR
- To add to the fun, two of the Federal agents investigating the Silk Road were themselves corrupt. . .

Child Pornography

- Child pornography is also popular on the Dark Web
- It's a natural fit—it's all information-based; there's no need to ship anything physical
- The FBI has had some success here, too

- Suppose you control a Tor hidden server
- Maybe you've found it and done something physical—or maybe you've hacked into it
- Plant malware on that server—and when other Tor users visit it, infect their machines
- All that software has to do is send the FBI the machine's real IP address
- The FBI has done exactly that

Legal and Ethical Issues!

- Is it proper for the FBI to hack computers? There's no explicit statutory authority, but most lawyers say it's OK if they have a search warrant
- Do judges understand the warrants they're signing?
- Is it OK for the FBI to run a child porn server for a while?
- Is it OK to hack a machine in another country, or one where you don't even know what country it's in?
- Is it OK to hack hundreds or thousands of machines with a single warrant?
- Do judges understand those warrants?

- A Tor hidden service for whistleblowers
- News organizations run Tor SecureDrop services—to send information anonymously to such a organization, connect via Tor
- (See <https://theintercept.com/securedrop/> or <https://securedrop.propublica.org/>)
- Note well: procedural security matters, too

- Exit nodes have been seized or searched by the police
- What if the exit node is corrupt? That has happened.
- There are various statistical attacks on Tor links
- (The FBI apparently subpoenaed the results of some experiments at CMU)

The Bomb Threat During Finals

- At one school, a bomb threat was email in during finals
- It was sent over Tor
- The network folks found that only one person at that school was using Tor at that time. . .

- Tor protects the IP address, but not anything else
- Higher-level data is not anonymized—it can often reveal identity or at least continuity (e.g., login names or tracking cookies)
- If you don't patch your system, you can be hacked
- Never use Tor *except* through the official Tor Browser Bundle or the Tails bootable USB stick

Is Tor Worth It?

- Evading censorship is good
- Talking freely to news agencies is good
- Child pornography is not good
- Soliciting murders for hire is even worse
- Should Tor exist? What about Tor hidden services?

Bird of the Day



(Red-tailed hawk, Central Park, July 16, 2019)