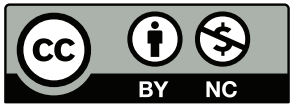
Freedom of Speech: Anonymity



Why Anonymity?

- Free speech can be unpopular
- Threats of physical harm
- Threats of job loss or other forms of financial coercion
- Social shame—unpopular lifestyles, embarrassment, etc,
- *Often, anonymity is necessary for truly free speech*

Long History of Anonymous Political Speech

- The *Federalist Papers* were nominally written by “Publius”
- There were many examples in British history of reprisals against authors—and of others writing anonymously to avoid such fates (i.e., the “Letters of Junius”)
- “There can be no doubt that such an identification requirement would tend to restrict freedom to distribute information and thereby freedom of expression.” Talley v. California, 362 U.S. 60 (1960).

But—What About Accountability?

- Sometimes, we want to hold people accountable for what they say
- We vote by secret ballot, but the legislators we elect (usually) vote publicly
- The Supreme Court has closed deliberations, but its votes and the rationale for them are very public
- "during election campaigns . . . false statements, if credited, may have serious adverse consequences for the public at large." McIntyre v. Ohio Elections Commission, 514 U.S. 334 (1995).
- More on accountability next class

It's Not Just Political Speech

- The Court has held that all speech can benefit from anonmity: "The decision in favor of anonymity may be motivated by fear of economic or official retaliation, by concern about social ostracism, or merely by a desire to preserve as much of one's privacy as possible. Whatever the motivation may be, at least in the field of literary endeavor, the interest in having anonymous works enter the marketplace of ideas unquestionably outweighs any public interest in requiring disclosure as a condition of entry. Accordingly, an author's decision to remain anonymous, like other decisions concerning omissions or additions to the content of a publication, is an aspect of the freedom of speech protected by the First Amendment. " (McIntyre)
- The Court also noted that law school exams are graded without knowing the students' names

Anonymous Speech in Cyberspace

- Assume that we do need anonymous speech. How do we do it online?
- Virtually all Internet traffic requires an accurate source IP address to be useful
- Can we achieve *useful* anonymity?

Anonymity versus Pseudonymity

- *Anonymity*: “The quality or state of being unknown or unacknowledged.” (`www.dictionary.com`)
- *Pseudonymity*: “Use of a [fictitious name, especially a pen name]”
- Which do we need, anonymity or pseudonymity?

Anonymity

- We cannot have true anonymity at the IP layer
- We can have it at the application layer—we often do, for web sites we visit
- Is there a linkage? Most applications keep log files showing what IP address performed various actions
- How long is this log file kept? Who can access it? Under what conditions?

Pseudonymity

- Pseudonymity is extremely common on the Internet
- Login names, screen names, etc., are all forms of pseudonyms
- Some people have many different ones—and occasionally with markedly different apparent attributes
- We can do something similar at the IP layer, by having another node carry our traffic

IP Addresses

- Recall that IP addresses are assigned topologically
- There are technical benefits to clustering IP addresses assigned to particular locations by ISPs
- This means that IP addresses can reflect geographic location

Where's Steve?

- When I first created this lecture, my IP address was 206.117.31.142
- Per `http://www.ip2location.com/206.117.31.142` I am indeed in Los Angeles
- IP addresses are often leaked by mailers...

Mail Headers

```
Received: by machshav.com (Postfix, from userid 512)
    id 46A9252D5D7; Wed, 17 Feb 2010 11:56:42 -0500 (EST)
Received: from tarap.cc.columbia.edu (tarap.cc.columbia.edu
    [128.59.29.7]) by machshav.com (Postfix) with ESMTP
    id 6CFB652D496 for <smb@machshav.com>;
    Wed, 17 Feb 2010 11:56:37 -0500 (EST)
Received: from [147.28.2.10] ([147.28.2.10]) (user=smb2132
    mech=PLAIN bits=0) by tarap.cc.columbia.edu (8.14.3/8.14.
    with ESMTP id o1HGtpvm003198 (version=TLSv1/SSLv3
    cipher=AES128-SHA bits=128 verify=NOT)
    for <smb@machshav.com>; Wed, 17 Feb 2010 11:56:31 -0500
```

Note that it thinks I was at 147.28.2.10, not 206.117.31.142

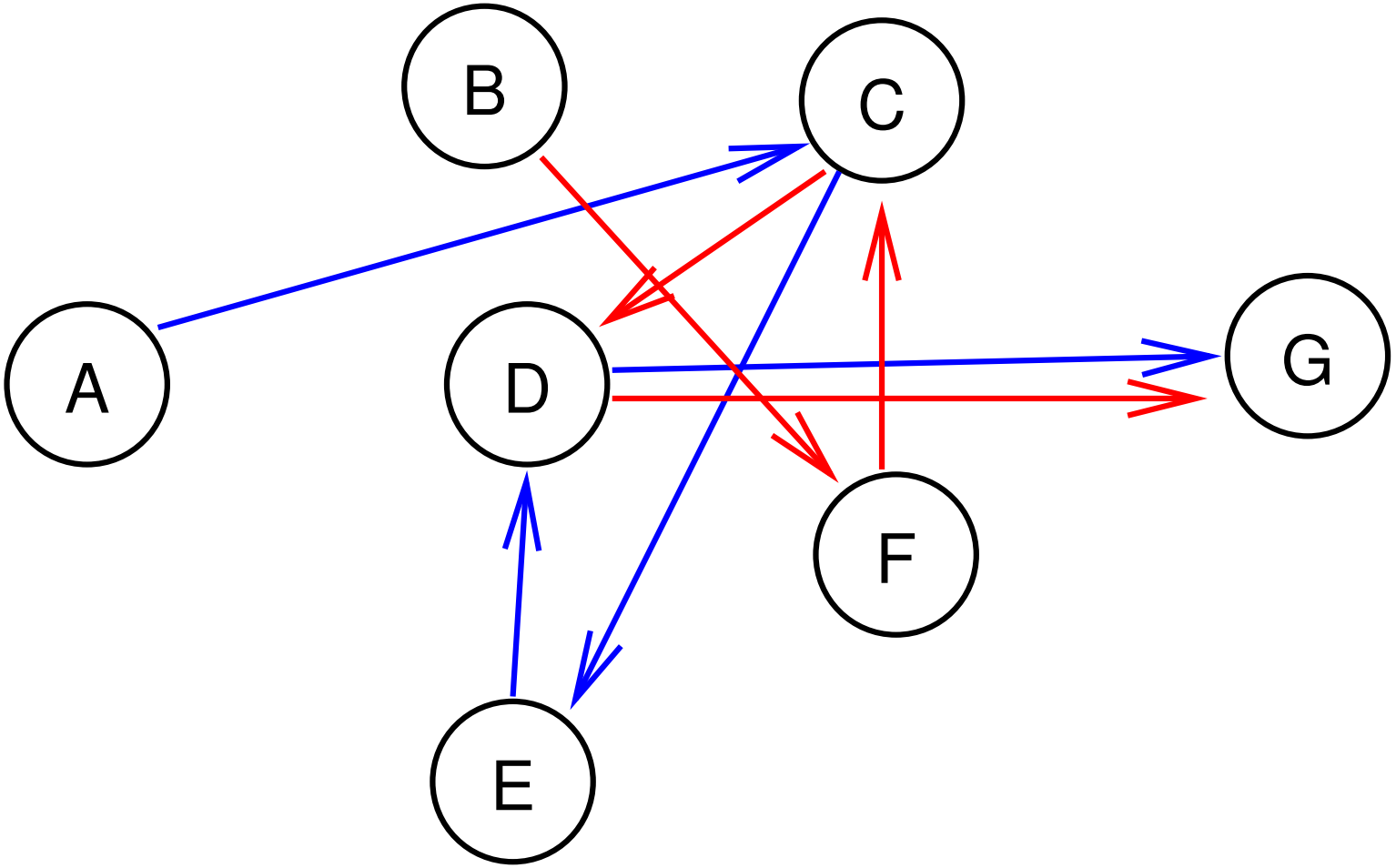
What Is Learnable?

- 147.28.2.10 appears because I was using a VPN hosted in Seattle
- IP geolocation thinks it's in Tokyo—that site used registration data to determine physical location
- (GMail generally doesn't show the sender's IP address)
- Note the "smb2132"—even if I'd changed my **`From:`** address, CUIT lists my UNI

Enter Onion Routing

- A client computer picks a set of 3–4 “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
- The set of Tor nodes, including the exit node, is changed frequently
- In other words, the pseudonym is short-lived

Hiding the Source Address



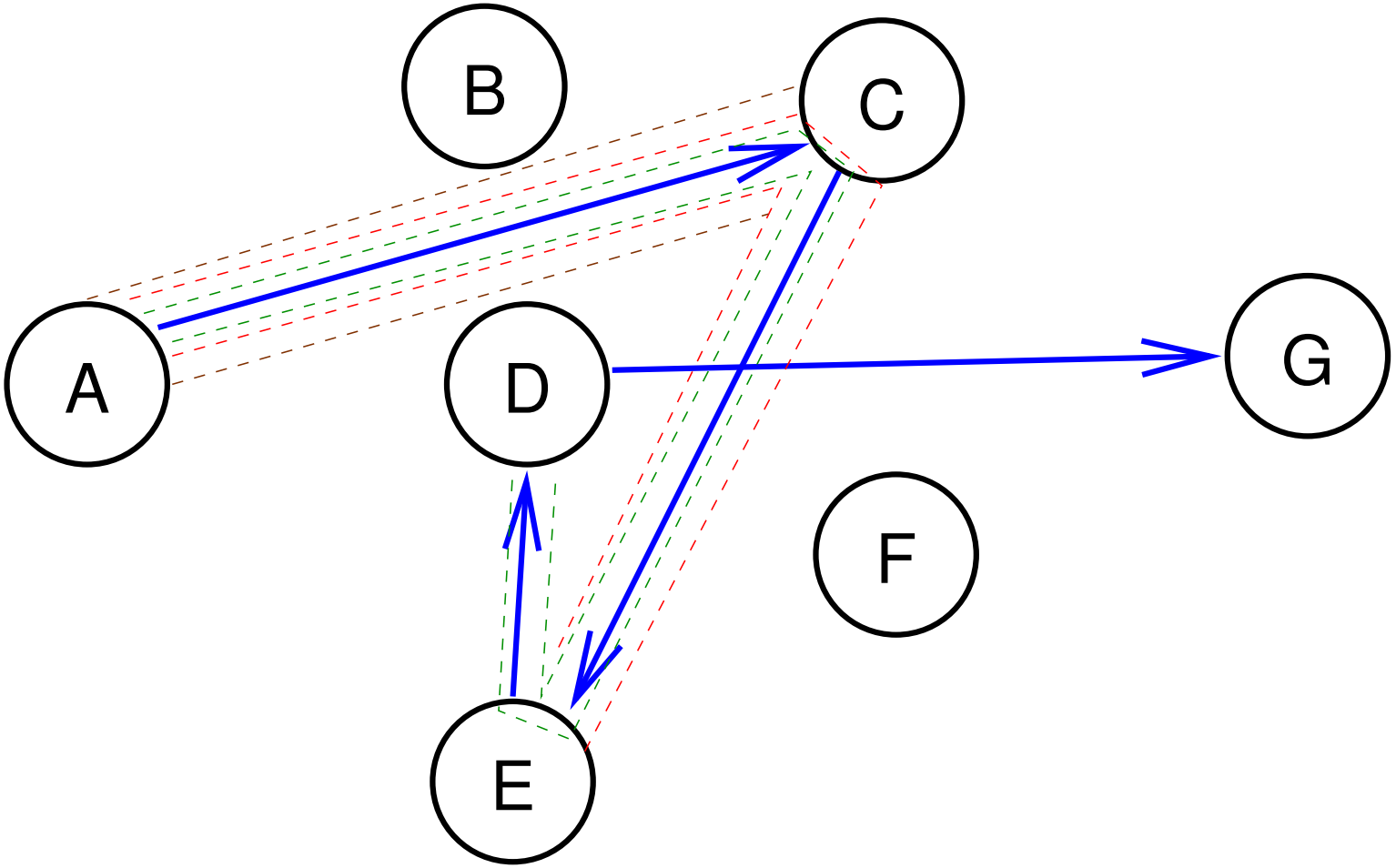
Onion Routing

- **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

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

Risks of Tor

- Exit nodes have been seized or searched by the police
- What if the exit node is corrupt? That has happened.
- There are varioius statistical attacks on Tor links
- Higher-level data is not anonymized—it can often reveal identity or at least continuity (e.g., login names or tracking cookies)
- The Silk Road server was somehow found despite using Tor

Pseudonymity

- Create an alias that can receive services
- Cryptographically-protected—and changing — path to the real service
- With Tor, they're called "hidden servers"

Remailers

- Simplest form: mailer has an alias that forwards to you
- More popular before Gmail/Hotmail/YahooMail, etc.
- More complex ones use cryptography, in ways (roughly) similar to Tor
- Recipients identified by a public key and a first-hop email address

An Early Remailer: `anon.penet.fi`

- Simplest form; also supported Usenet posting
- Targeted by “subpoena attack” by the Church of Scientology
- Creator pulled the plug after the second such incident

Who Are You if You're Anonymous?

- How do people know whether or not they can trust your emails if you're anonymous or pseudonymous?
- Reputation—have your mails over time been trustworthy?
- How do they know the same person sent the 100th message as sent the first 99?
- Messages can be *digitally signed*
- Note that that problem exists for ordinary email, too!

Digital Signatures

- Related to public key cryptography
- Again, everyone has a private key and a public key
- (With public key cryptography, encrypt with the public key and decrypt with the private key)
- For digital signatures, sign (which is really encrypt!) with the private key; anyone can verify this with the publicly-known public key

Anonymous Payments

- How do you buy something online anonymously?
- In person, we can use cash
- Online, we generally use credit cards

What's the Problem?

- Apart from anonymity, many people have security concerns—is it safe to use a credit card number on the Internet?
- A digital equivalent of cash seems impossible—files (or bits) can be copied, so how can you prevent double-spending?
- Nevertheless, it's possible

Digital Cash

- Invented by David Chaum; many more schemes since then
- Get “coins” from bank; engage in interactive protocol with recipient
- Neither the bank nor the recipient learns the spender’s identity
- The merchant can confirm that the coins are valid
- Does not prevent double-spending; however, the identity of the double-spender is revealed to the bank

Over-Simplified Intuitive Version of the Protocol

- When Alice withdraws an electronic "coin" from the bank, she receives a set of digitally-signed "identity halves", with her name
- An identity half reveals *nothing*; you need both halves to learn anything about the name
- To accept the coin, the merchant asks for a random subset of 50% of the identity halves. These are verifiably from the bank because of the digital signatures. The halves are sent to the bank to deposit the coin
- If Alice "spends" the same coin again, probablistically at least one of the identity halves will match one from the previous instance. With two halves, the bank can recover Alice's identity
- The details and the math are a lot more complex. . .

Bitcoin

- Invented in 2008 by “Satoshi Nakamoto”
- A distributed digital currency—no central bank
- To prevent double-spending, transactions are recorded in the *blockchain*—an Internet-wide, peer-to-peer, distributed database
- Bitcoins “mined” by solving a hard cryptographic problem
- Payments are made from one bitcoin address to another

Bitcoin Isn't Anonymous

- All transactions are recorded in the public block chain
- This permits traffic analysis of transactions
- Identities can sometimes be linked to real parties, e.g., by engaging in a transaction with someone
- It's a pseudonymous currency

Pseudonymous Payment

- There are a number of widely-used pseudonymous payment schemes in use, most notably Paypal
- Some “storefronts”—Amazon.com, Google Checkout, others—are effectively pseudonymous with respect to the merchants
- Primary issue has been security, rather than privacy
- Delivery of physical objects remains problematic

Real-World Anonymity

- Thus far, light-weight pseudonymity has sufficed
- Tor is popular among a group of enthusiasts, and has the EFF coordinating the project—but a commercial analog failed
- Acceptance depends on the threat model

Threat Models

- Who are the adversaries, and what are their capabilities and motives?
- Sometimes, the problem is hacking (i.e., Google versus China, or some cases of harrassment or stalking)
- More often, it's court orders; most people are not afraid of that threat
- Of course, most people are not aware of their data shadows. . .