The Dark Web

Tor and the Dark Web

- There are ways to use the Internet (almost) untraceably
- This can be used for good purposes or bad purposes
- Two technologies are necessary, communication and and payment
- Let’s look at the technologies first

Payment System Requirements

- Untraceable (i.e., cash, not credit cards)
- Decentralized, to avoid a single point of monitoring and control

Digital Cash Requirements

- Prevent counterfeiting
- Prevent double-spending
- If these can’t be prevented, make sure that violators are identified
- There have been schemes for this since the 1980s, but they all required a “bank”

Digital Currencies

Morningside Park, April 29, 2019

(photo from Wikipedia)

Bitcoin

- Bitcoin was the first truly decentralized digital cash system
- It was invented by the pseudonymous “Satoshi Nakamoto”
- Bitcoins are “mined” by solving a hard cryptographic problem
- ☞ The only way to “solve” it is brute force
- Double-spending is prevented by recording transactions in the “block chain”, a distributed database updated via a peer-to-peer network
- People generally buy and store Bitcoin via exchanges

The Block Chain

- A distributed, publicly verifiable, ledger
- Bitcoin transactions are sent to special “mining” block chain nodes
- They work (i.e., do lots of calculations) to solve a cryptographic problem; when they succeed, they add a new block to the chain
- (Each block points to the previous block)
- In case of ties, the longest block wins
- Eventually, all nodes agree on the sequence of transactions
- Since each transaction is recorded, you can’t double-spend a Bitcoin
- Important note: block chains have other uses besides cryptocurrencies

Status of Bitcoin

- Bitcoin had been used legitimately but its volatility and high transaction costs have made that much less common
- Its value fluctuates quite a bit, and there have been scandals in the Bitcoin industry
- Mining has become uneconomical; it costs more in electricity than the resulting Bitcoins are worth
- Bitcoin doesn't scale well—it was never properly engineered for high-volume use

Bitcoin and Electricity

- Bitcoin uses a *lot* of power to run the mining nodes
- Investors are buying old, obsolete, dirty power plants to power these computers
- Is this a responsible, ethical thing to do during an era of CO_2-induced climate change?

It Isn’t Completely Anonymous

- Because all transactions are publicly recorded in the block chain, it’s possible to link together various transactions
- If the user’s private files are compromised, all transactions by that user are revealed
- There are variants of Bitcoin that solve that
- There are also “tumblers”—sites that mix together various Bitcoin transactions, to disguise who has received what from whom (though they don’t work all that well)

Ethereum

- Ethereum: embed programs—“smart contracts”—into the coins
- Will those programs be correct?
- The DAO incident: a bug in an Ethereum contract let someone loot a lot of Ethers
- (Other bugs, too)

NFTs—Non-Fungible Tokens

- Take a digital item, e.g., a JPG, and digitally sign it
- Put this on a block chain; record sales of this signature on the block chain, too
- At its best, the digital equivalent of a certificate of authenticity
- But—NFTs do *not* convey copyright or ownership of the signed image!
- I could sell you an NFT of this slide—but I'd still own the copyright, not you; you'd merely own the NFT
- NFTs—Non-flowering tulips?

Initial Coin Offerings

- Instead of selling stock, sell cryptocoins—usually by payment in Bitcoin or Ethers
- In theory–or at least, according to some theories—these are currency, not stocks, and therefore not subject to Securities and Exchange Commission regulations
- Some can, in principle, be used later to buy services from the companies
- The SEC doesn't always agree with the issuers. . .

Current Status

- There are very many cryptocurrencies
- Most have relatively low acceptability, i.e., you can't easily sell or spend them
- Most economists think that Bitcoin is a bubble (but that's been said for years)
- A lot of work to improve scalability
- Many legal concerns: fraud by the exchanges, hacking of them, money laundering, use by criminals, and more
- But—major financial firms are seriously interested
- El Salvador has declared Bitcoin legal tender

Tracking

- 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 harass users
- The IP addresses are public, for governments that want to block certain sites
- Is there a solution?

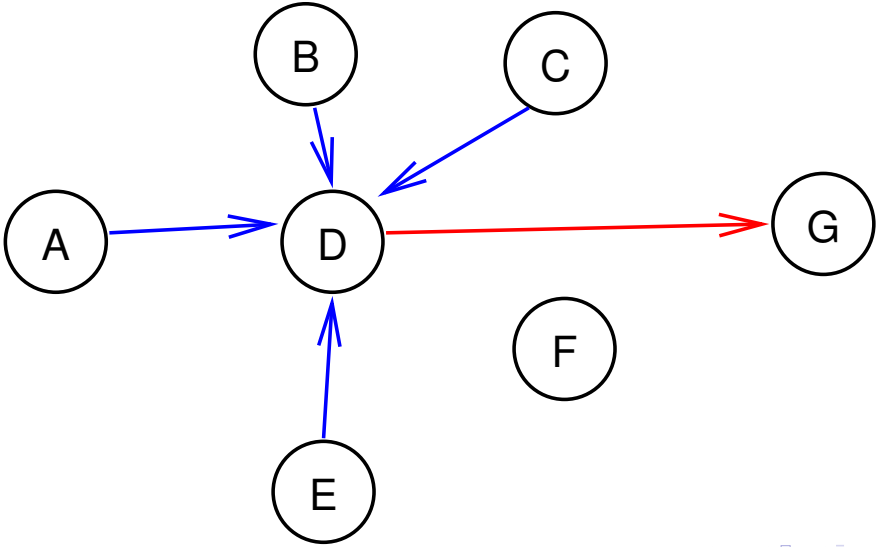
Solutions

- Forward traffic through some other node
- Will the trace stop at the forwarding node?
- What are the attackers' powers?

Virtual Private Network

- All nodes forward traffic to a VPN node
- It sends the traffic to desired destinations
- Destination nodes see the IP address of the VPN node, not that of the actual clients
- However. . .

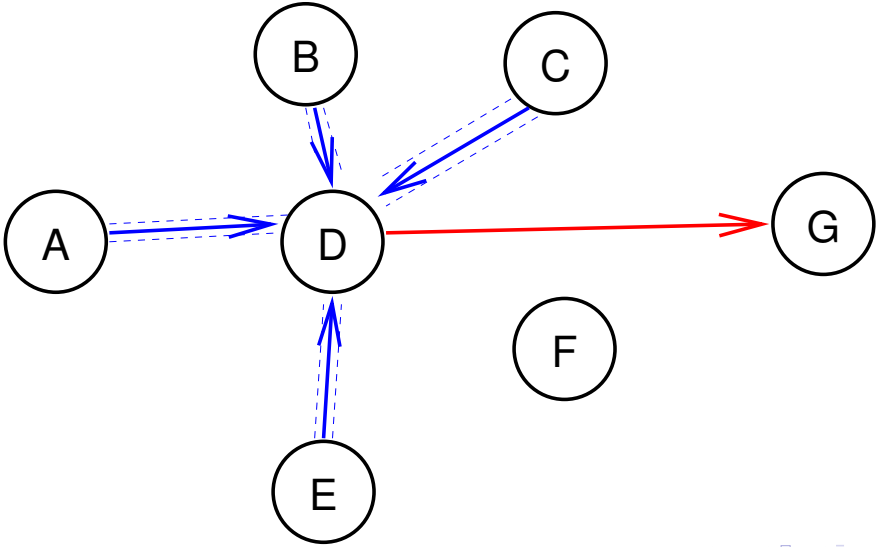
Hiding the Source Address with a VPN



Powerful Attackers

- Suppose the attacker can wiretap the link to the VPN
- Match the input packets with the output packets
- Who are the attackers? Destination nodes can’t identify the source, but a government can
- And what if the VPN provider is corrupt?

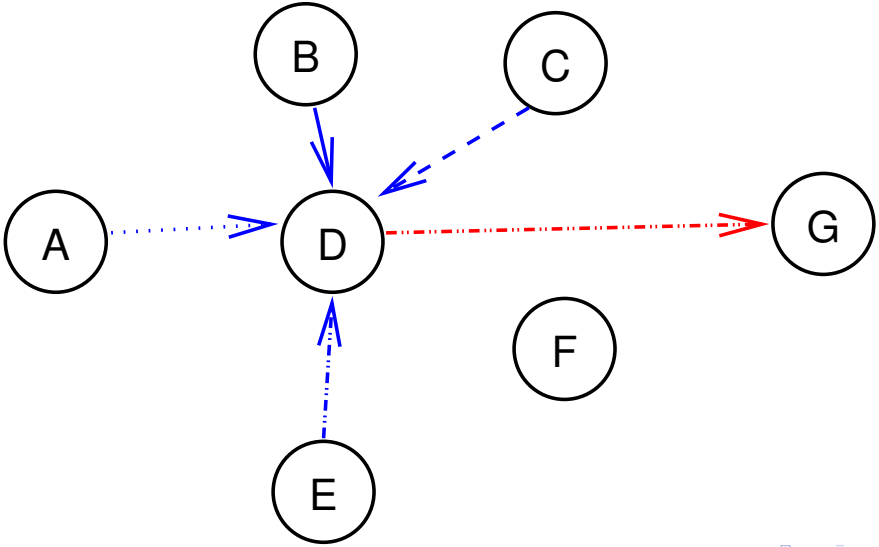
An Encrypted VPN



Metadata

- Using the IPsec VPN protocol, the length of encrypted packets is (usually) $42 + |P| + ((14 - |P|) \bmod 16)$, where $|P|$ is the length of the plaintext packet
- An eavesdropper who monitors packet lengths can still identify the real source
- It's possible to add padding—but that doesn't hide timing: who sent a packet just before one was forwarded?
- Or—look for packet size *distribution*

Monitoring Packet Size Patterns



Requirements for a Solution

- Mix traffic from multiple sources
- Defend against local eavesdroppers
- Avoid easy-to-detect patterns

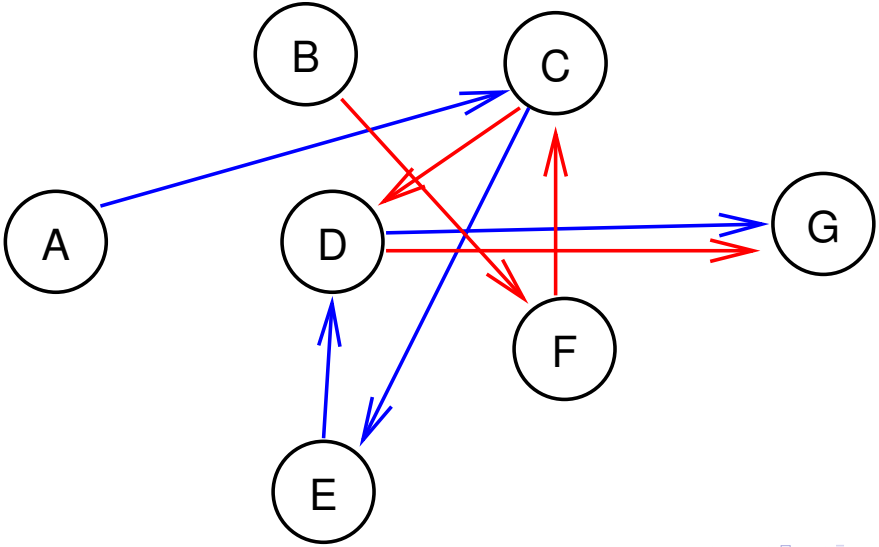
Tor: The Onion Router

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

How it Works

- 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 used, including the exit node, is changed frequently
- In other words, the source IP address is short-lived

Multiple Hops



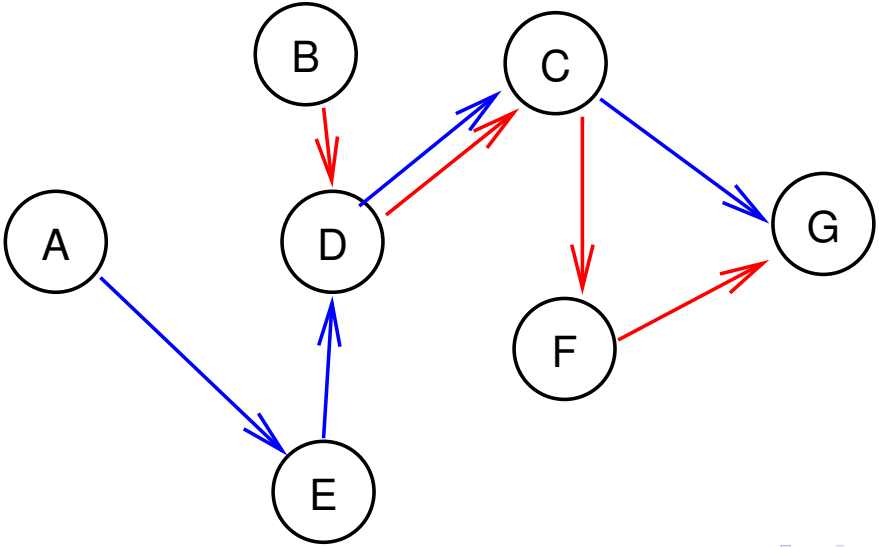
Onion Routing

- **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

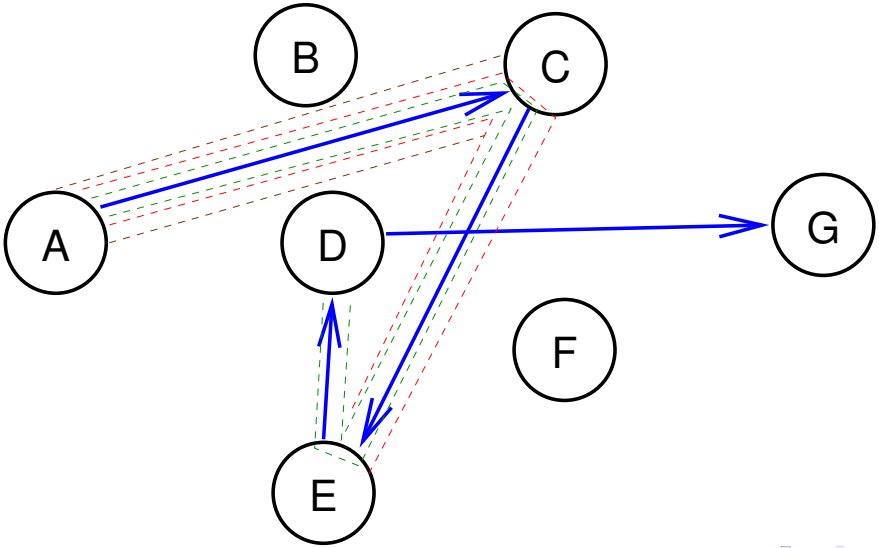
Why Multiple Hops?

- 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.
- (Tor was invented at the Naval Research Lab—the military understands the need for this sort of anonymity for their communications)
- No accountability. . .

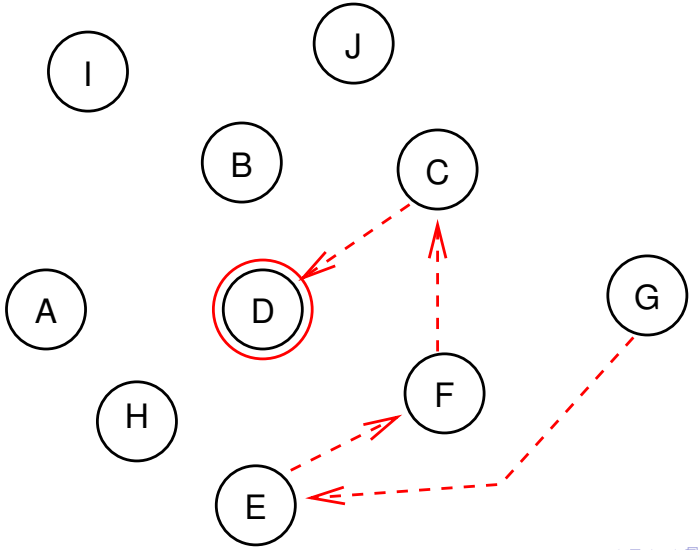
What About Servers?

- 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?

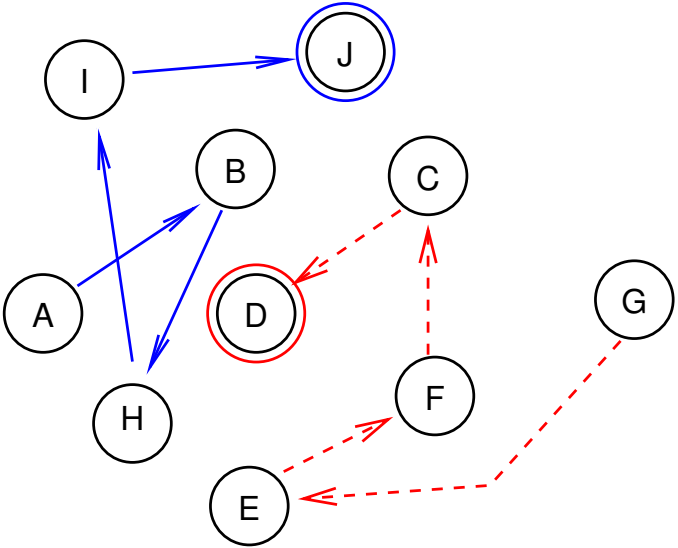
Tor Hidden Services

- 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 rendezous point
- The rendezvous node forwards traffic between the two Tor services

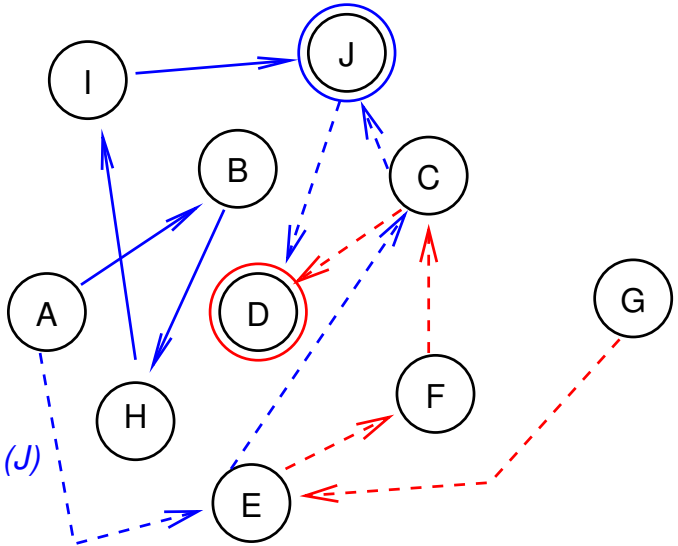
Creating an Introduction Point



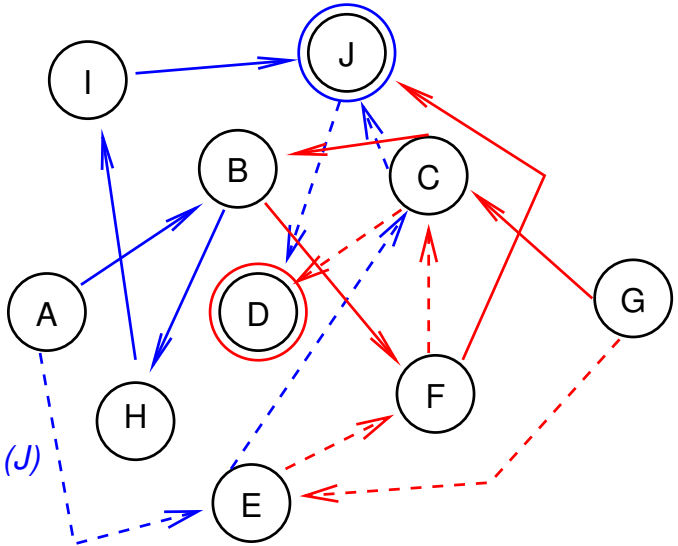
Creating an Rendezvous Point



Notifying the Introduction Point



Traffic Can Flow



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./postalC
ode=94025, C=US, ST=CA, L=Menlo Park, O=Facebook, Inc., CN=*.facebookcorewwwi.onion
        Subject Public Key Info:
```

How Did They Get facebookcorewwwi.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—"corewwwi"—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

Hacking Tor

- 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?

SecureDrop

- 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

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 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. . .

User Errors

- 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?

Questions?

(Two red-tailed hawks, possibly in a courtship flight, Morningside Drive, October 19, 2020)