Privacy at the Border

CS@CU

Searching Devices at the Border

Searching Devices at the Border

- What might the government want?
  - Only contraband?
  - Other information?
- What is technically feasible to collect?
  - What is feasible to collect under what circumstances?
- Device types: phones, tablets, computers, flash drives

Constraints

- Technical
- Time
  - Regular search or “enhanced” search?

Laptops

- Generally easy to boot a laptop from an external drive and run code to scan the file system
  - But: the laptop's disk may be encrypted
  - But: laptop disks are big
  - But: copying *everything* to an external drive is slow because USB ports and drives might be slower than the internal drive

External USB Drives

- Moderately easy to scan—not that big
- But—some people have many of them
- But—may be encrypted

Scanning Phones

- Much harder—phones are more closed, especially iPhones
- But—more data:
  - Location data
  - App-specific data
- Approach: use mobile device forensic tools (MDFT) or (for Android) privileged app
- Again: speed is an issue

Data of Interest

- Contacts
- Text messages*
- Phone call history*
- Email contacts and content*
- Photos? Videos? These are *very* large

*May be obtainable from provider via suitable legal process—but that isn't a border search

Scanning Photos for Contraband

- One major target for border scans: child sexual abuse material (CSAM)
- By definition these are contraband and illegal per se
- Can these be scanned at the border? How?

Scanning Photos for Contraband

- One major target for border scans: child sexual abuse material (CSAM)
- By definition these are contraband and illegal per se
- Can these be scanned at the border? How?
- Cryptographic hashes?
  - Don’t need to have copies of the CSAM at every border crossing

Three Images, Two Identical…

Three Images, Two Identical—and Their Hashes

```
4db948efe7218fa51969991d9dcbbb06
09678549929b2553b19032455ed946cf
```

```
708b5bd8320dca9b332cd60e3b1998e4
950fc7a4d542a15a7ad4285d82b5bd72
```

```
2ac91b4662b4ff68dc8d044ad57d5118
47b0797c838a225ca35cef4d96e34429
```

The images look identical—but the hashes are all different!

CS @CU

It’s the Metadata

```
20c20
< Modify Date                   : 2021:02:19 13:55:23
---
> Modify Date                   : 2021:02:19 13:55:35
134c134
< Metadata Date                 : 2021:02:19 13:55:23-05:00
---
> Metadata Date                 : 2021:02:19 13:55:35-05:00
142c142
< Document ID                   : xmp.did:0eb91a7c-f60a-4a51-88f4-3a402bcde621
---
> Document ID                   : xmp.did:cc9a1add-cbf9-4d63-9243-28585f24ffee
145c145
< Instance ID                   : xmp.iid:0eb91a7c-f60a-4a51-88f4-3a402bcde621
---
> Instance ID                   : xmp.iid:cc9a1add-cbf9-4d63-9243-28585f24ffee
```

It’s the Metadata — So Let’s Delete It!

```
20c20
< Modify Date                   : 2021:02:19 13:55:23
---
> Modify Date                   : 2021:02:19 13:55:35
134c134
< Metadata Date                 : 2021:02:19 13:55:23-05:00
---
> Metadata Date                 : 2021:02:19 13:55:35-05:00
142c142
< Document ID                   : xmp.did:0eb91a7c-f60a-4a51-88f4-3a402bcde621
---
> Document ID                   : xmp.did:cc9a1add-cbf9-4d63-9243-28585f24ffee
145c145
< Instance ID                   : xmp.iid:0eb91a7c-f60a-4a51-88f4-3a402bcde621
---
> Instance ID                   : xmp.iid:cc9a1add-cbf9-4d63-9243-28585f24ffee
```

```
$ dd if=s_DSC_3536.jpg bs=1k skip=50 | shasum -a 256
87+1 records in
87+1 records out
89560 bytes transferred in 0.047579 secs (1882341 bytes/sec)
948fe5063284bb2ded32bab6094bce1ed1a45875a8cb7a88bba2b1cc692d6050  -
$ dd if=s_DSC_3536-2.jpg bs=1k skip=50 | shasum -a 256
87+1 records in
87+1 records out
89560 bytes transferred in 0.026664 secs (3358834 bytes/sec)
948fe5063284bb2ded32bab6094bce1ed1a45875a8cb7a88bba2b1cc692d6050  -
```

Trivial Image Changes

- There are many image changes that don't materially affect perception but do change the hash: scaling, cropping, minor color tweaks, etc.
- Answer: a *semantic hash*, e.g., PhotoDNA
- But—there is no published data on how PhotoDNA works
  - Is it robust?
  - Does it produce false positives? False negatives? Under what conditions?
- And: who controls addition of photos to the database of contraband?

Arithmetic…

- For exact matches, hash every image or file in your database
  - How many are there? Assume 100,000,000
- Shorten the hashes to 64 bits—8 bytes
  - Odds of a false positive: 1 in 2^{64}, i.e., practically impossible
- Can you store 800,000,000 bytes at every custom's point? Of course! And searching that table is also very, very fast
- Scanning time: for 110K photos, about 200 GB, I hashed them all in ~8 minutes, and there are shortcuts, e.g., checking size first

Searching Devices

- Some searches are feasible
- Detailed ones take too long, and may require an enhanced search

Right to Be Forgotten

The Right to be Forgotten

- Sometimes, people don’t want items about them indexed by search engines
  - Example: minor crimes, long ago
- A matter of privacy: the right to control how much one is willing to “share its personal information with others”
- But—retaining and “speaking” such information is free speech
- Conflict of rights: free speech versus privacy

The EU

- Currently the law within the EU: EU citizens can request that information about them be removed from search engines
  - Does not apply to prominent people, major stories, etc.
- Enforceable *only* within the EU—information about such people need not be suppressed outside the EU
- Such orders are probably not enforceable in US courts
- How does Google know where you are?

The Economics of the Right to be Forgotten

- There is no requirement that the original data be removed, only that search engines delist it
- If you know where to look, you can still find it
- If you're well-resourced enough, you can build tools to search all likely places

Location and National Requirements

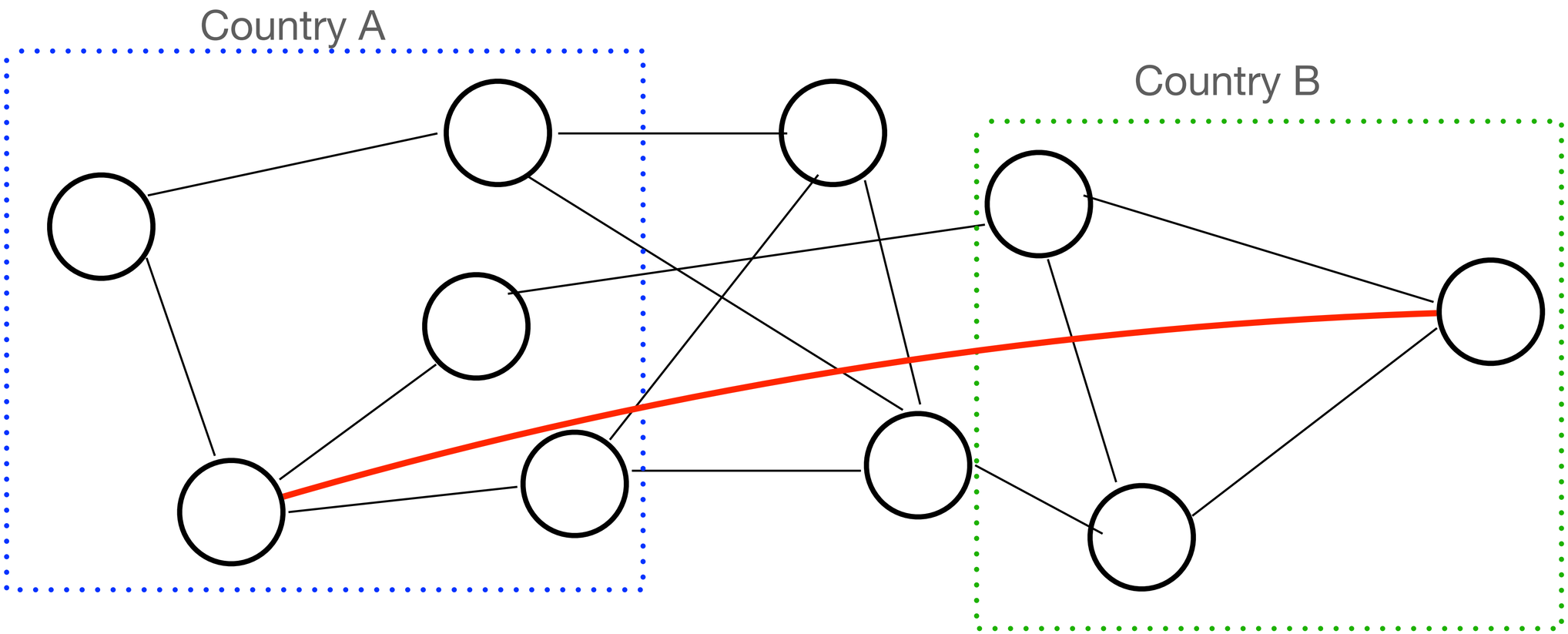
Location

- How does a website know where you are?
  - It can ask you, or it can do IP geolocation
- You can lie—and you can spoof IP geolocation
  - Apps? Are they hack-proof? How does a server know it's talking to the genuine app?
  - If it matters, stronger measures are needed—New Jersey's legal online gambling apps use WiFi geolocation

Virtual Private Networks

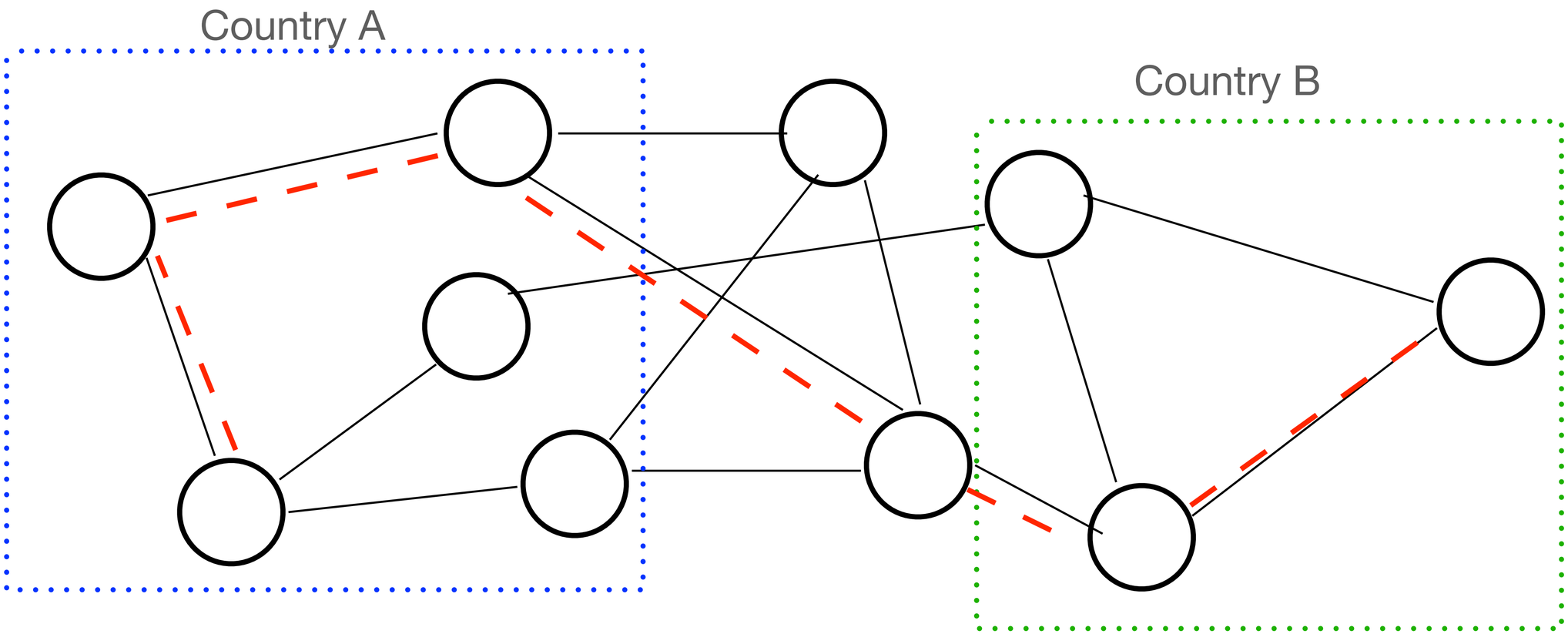
- IP addresses are customarily assigned to *interfaces*
- An interface is some way to speak IP to the outside world: Ethernet, WiFi, etc.
- You can create *virtual interfaces*—set up a software connection to some other host
- These virtual interfaces can acquire an IP address from the remote node—and appear there

A Virtual Link



These virtual links are often, but not always, encrypted

A Virtual Link: Hop by Hop



These links are just ordinary network connections

VPN Limits

- You have to trust the VPN provider
- Commercial VPN providers are well known, and are often blocked
  - Netflix et al. have geographic limits on their licensed content
- Slower

Content Censorship

- National censorship requirements
  - Lèse-majesté laws in Thailand
  - Content insulting Erdoğan in Turkey
  - Nazi content in most of Europe
  - Information on criminal trials in Canada
  - More…
- How is blocking done? How is scanning done?

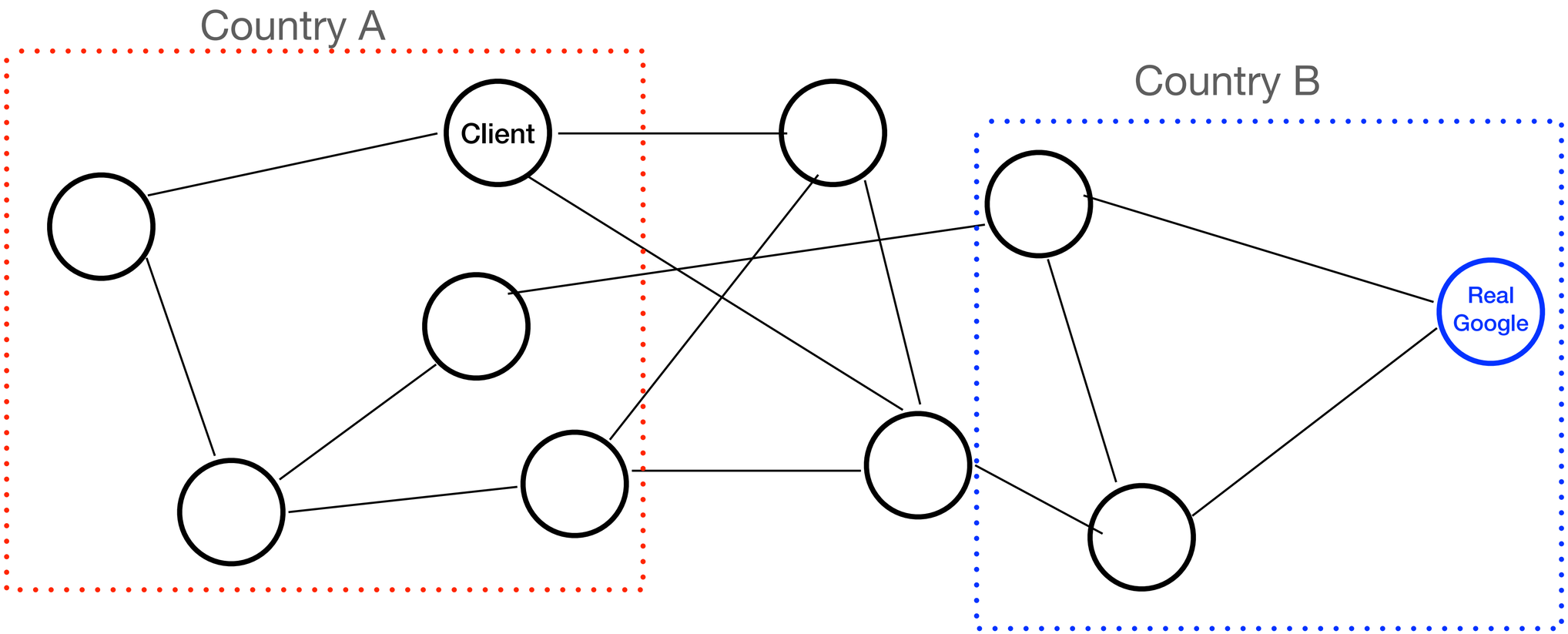
Routing

- Remember that ISPs route packets
- An ISP can refuse to send packets to particular destinations
- Example: Pakistan has blocked YouTube

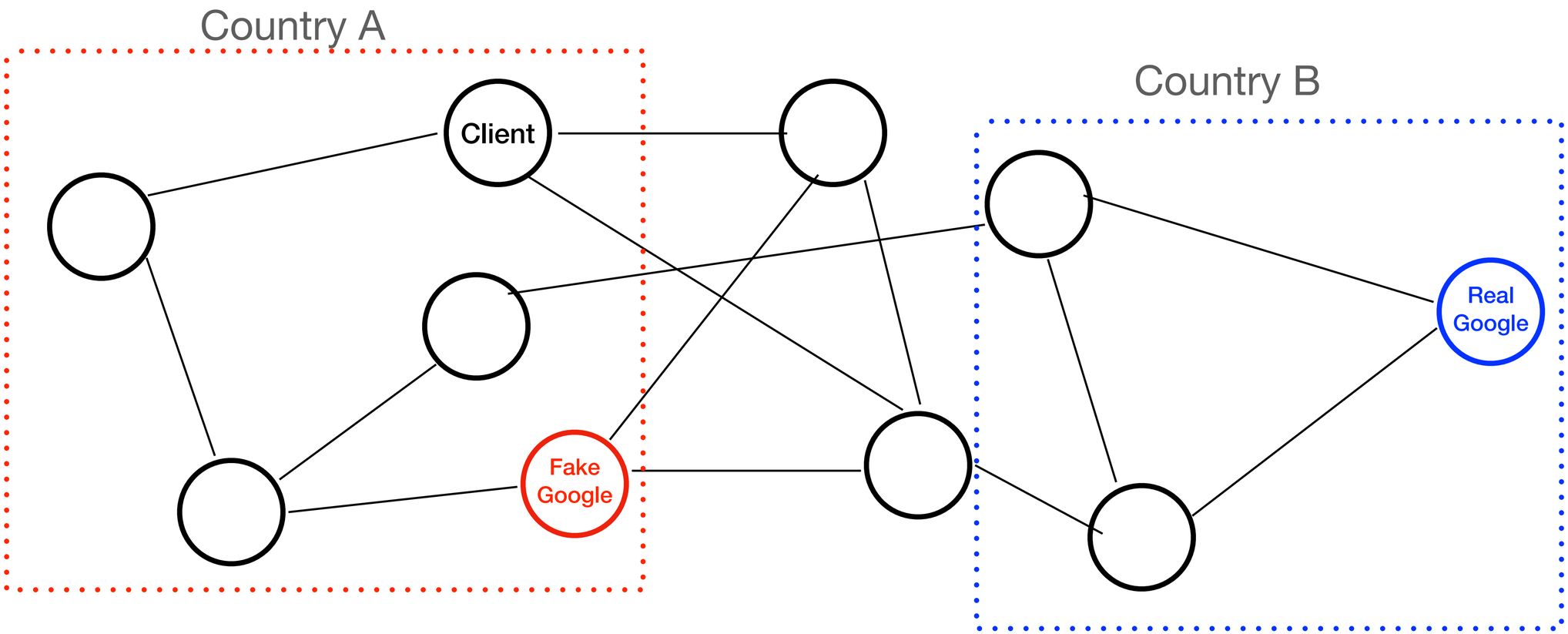
The Domain Name System

- The DNS maps hostnames to IP addresses
- If you can block or spoof DNS traffic, you can give the wrong address
- Send requests to forbidden websites to an error—or warning—page

Faking the Destination



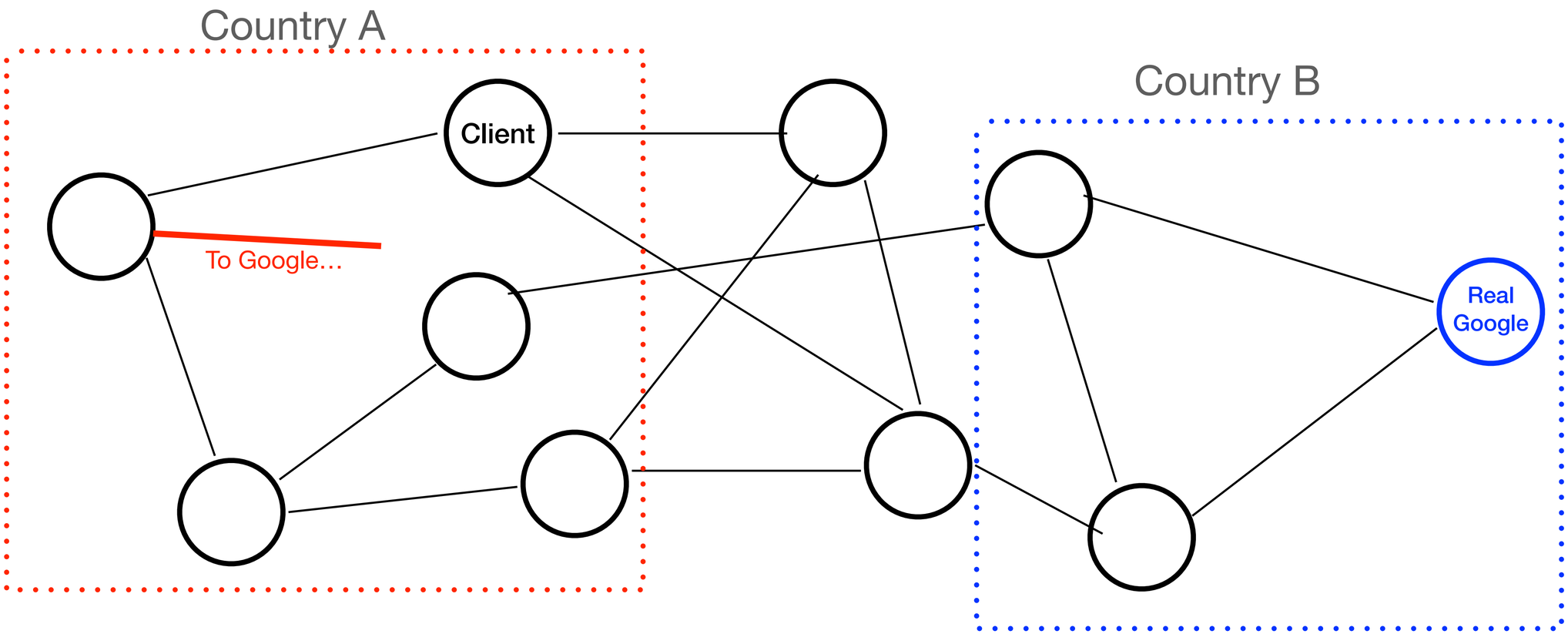
Faking the Destination



Certificates?

- In theory, certificate checks prevent spoofing
- However…
  - A government-mandated certificate authority (e.g., in Kazakhstan) can issue fake certificates
  - A hacked CA (Diginotar and Comodo) can issue fake certificates
  - Many users will click through the warnings
  - You don’t need a certificate to simply block the service

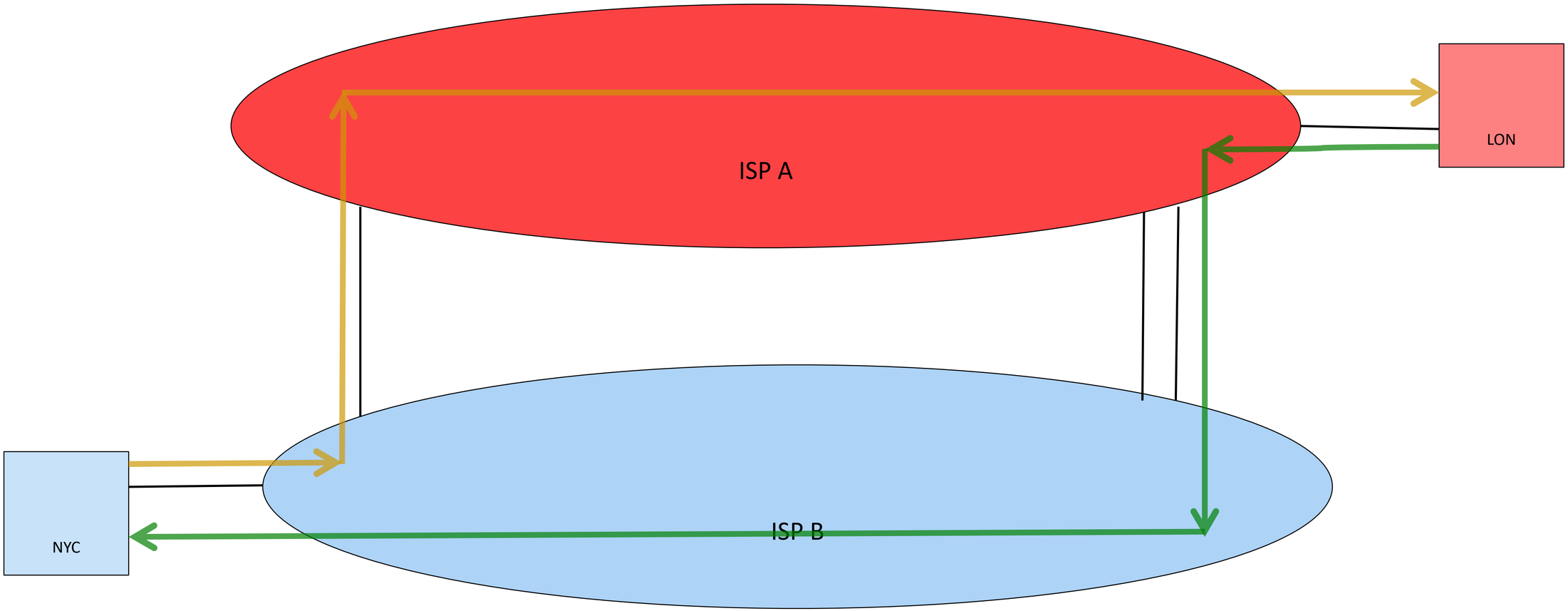
Discarding the Traffic



Filtering and Scanning

- It’s difficult, but possible, over the network
  - Link speeds are high
  - Messages are broken up into *packets*
  - It’s hard to know context
  - Asymmetric routing

Outbound and Inbound Paths Differ



ISP Topology

- International Internet connectivity from the US is based on the economic and engineering choices of the ISPs
  - Verizon Business considers North America a separate *autonomous system* (AS) (a concept for routing)
  - AT&T treats the whole planet as one AS—which makes it harder to do filtering
- Older technologies—telegraph, radio, telephone—were “facilities-based” and/or otherwise regulated, hence government permission was required for international gateways
- The early Internet grew up in a deregulatory era, and used circuits leased from telephone companies—no obvious buildings on the shoreline

Daily Bird

Black-capped chickadee, Central Park, February 21, 2021

CS @CU