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 (hence Tor)
- 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 was 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

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

Signing Pseudonymous Mail

- If you sign *all* of your email, it's easily attributable to you (whoever you are)
- But—is an unsigned email a forgery, or did you just forget?
- Also: there's a security risk to using your private key all the time; mailers aren't very secure

Certificates

- How do we associate a digital signature's public key with a person?
- We use a *certificate*—a digitally signed association between that person and their key
- But whose digital signature is on the certificate?
- A *certificate authority*—some party that you trust

Certificate Authorities

- Where do they come from? Why do you trust them?
- They're generally built into your OS, your browser, or your mailer
- There are serious security risks in the way they're used today—but that's a topic for COMS 4187

Pseudonymous Cryptographic Credentials

- Via cryptographic tricks, it’s possible for a certificate holder to create a pseudonymous sub-certificate
- Suppose that Columbia issued everyone here a certificate, which it signed
- You could issue a sub-certificate that was not linkable to you but was demonstrably derived from the Columbia University certificate
- It’s even possible to have unlinkable sub-certificates where you can blacklist the real owner *without* identifying the owner
- It’s also possible to have revocable anonymity

Attribute Certificates

- Sometimes, you want a certificate—age, employment, security clearance, etc.—but doesn't identify you
- An *attribute certificate* lists those values and a public key, but does not list a name
- (Again, signed by a certificate authority)
- Attribute certificates are used far less frequently than they should be
- Note: the public key itself (or even a bare public key) can also be a pseudonym

Social Networks: True Names

- Some social networks require use of real-world names
- “Facebook is a community where people use their authentic identities. We require people to provide the name they use in real life” (`https://www.facebook.com/help/112146705538576`)
- Google+: “we recommend using your first and last name on your profile” (`https://support.google.com/plus/answer/1228271?hl=en`)
- But: “there are no more restrictions on what name you can use” (`https://plus.google.com/+googleplus/posts/V5XkYQYYJqy`)

Why the Issue?

- Some people *need* anonymity
- LGBTQ individuals
- Sometimes people (especially women) who speak out on certain topics
- All of the reasons discussed at the beginning of the class

The Social Graph

- Whom you talk to leaks information
- Your set of Facebook friends, or who follows whom on Twitter leaks information
- Whom you call leaks information—no one else calls the same people as you do
- "If you had enough metadata—the pattern of how a communications device was used (whom did it call, who called it, when, for how long)—you could pretty much determine what the owner of a device was up to." (Michael Hayden)

Sexual Identity

- Many—but not all—LGB individuals self-identify that way on Facebook
- Hypothesis: LGB individuals have many more friends who are also LGB than do heterosexuals
- Question: is it possible to identify other LGB individuals on Facebook, simply according to their “friend” patterns?
- According to an MIT study, yes
- What else leaks that way?

No Good Defenses!

- The social graph is like the call graph—pure metadata
- Do we need a Tor equivalent for social network connectivity?
- How would that work? Who is the enemy, the social network operator or an outside observer?

Pseudonymous Social Networks

- Some social networks, e.g., Twitter, allow use of persistent pseudonyms
- I regularly deal with @thegrugq, @drunkenpredator, and more. . .
- But—uncommon pseudonyms can also be persistent
- Uncommon pseudonyms can be used on other networks, too—and loss of control of one can be serious
- Chasing a pseudonym off the net can be almost as devastating as doing it to a real name

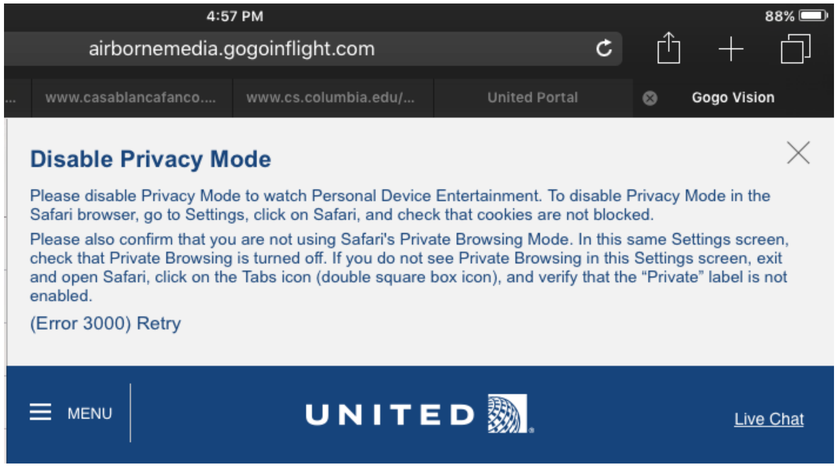
Application-Level Deanonymization

- Login names are a form of linkage
- Link that to other data
- Example: You log in to Facebook and view a product page
  Facebook records that
  You log in to some other site using Facebook Connect—and now Facebook knows you're on that site, too, and perhaps what your login name is there

Browser Privacy

- Turn on “private” or “incognito” modes
- That keeps browsers from retaining cookies or other local storage
- But—does nothing to protect against linkages while that data still exists

Some Sites Don’t Like That

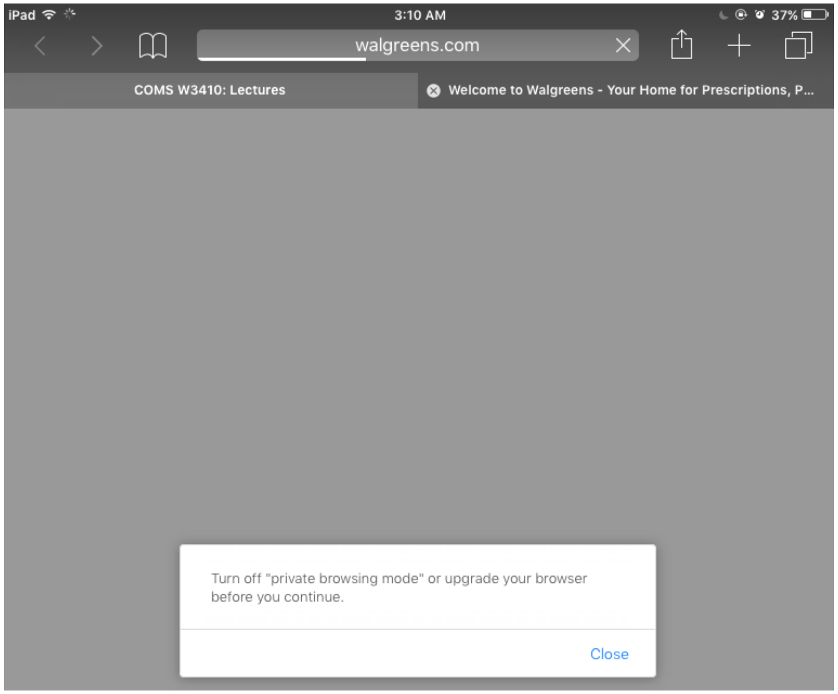


Disable Privacy Mode

Please disable Privacy Mode to watch Personal Device Entertainment. To disable Privacy Mode in the Safari browser, go to Settings, click on Safari, and check that cookies are not blocked.

Please also confirm that you are not using Safari's Private Browsing Mode. In this same Settings screen, check that Private Browsing is turned off. If you do not see Private Browsing in this Settings screen, exit and open Safari, click on the Tabs icon (double square box icon), and verify that the “Private” label is not enabled.

(Error 3000) Retry



Turn off "private browsing mode" or upgrade your browser before you continue.

Close

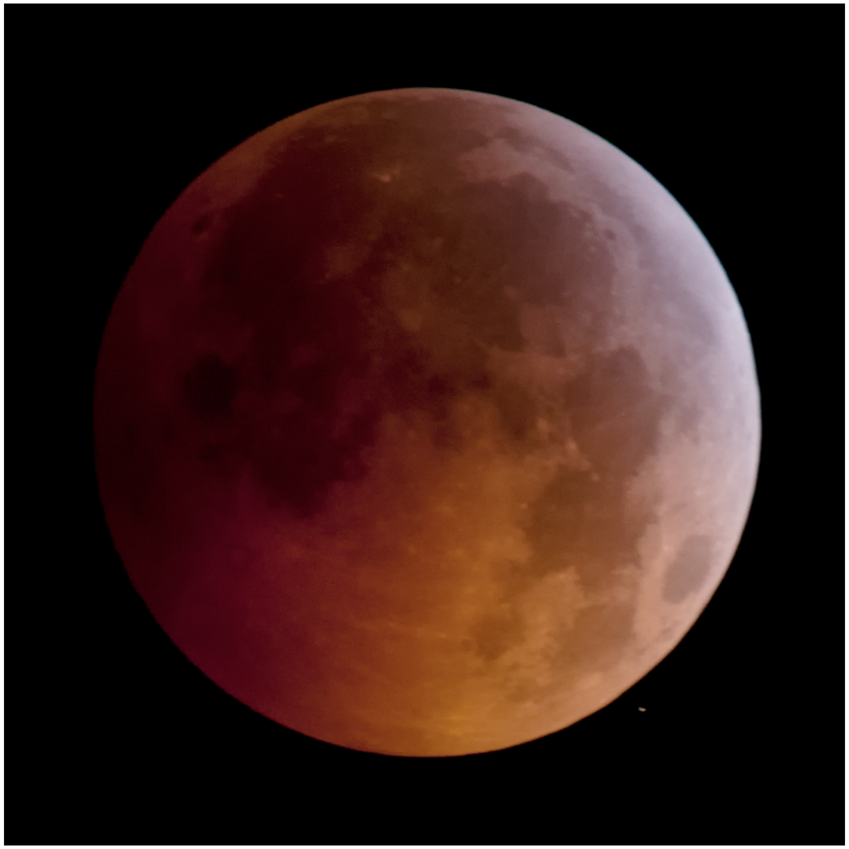
EXIF Tags

- JPGs contain a lot of metadata
- Some is about the photo:

```
Exposure Time: 1/60 sec
F-Number: f/4.5
Exposure Program: Not Defined
Exposure Bias: 0 EV
Metering Mode: Pattern
Light Source: Unknown
Flash: Flash, Auto, Return Detected
Focal Length: 44.00 mm
```

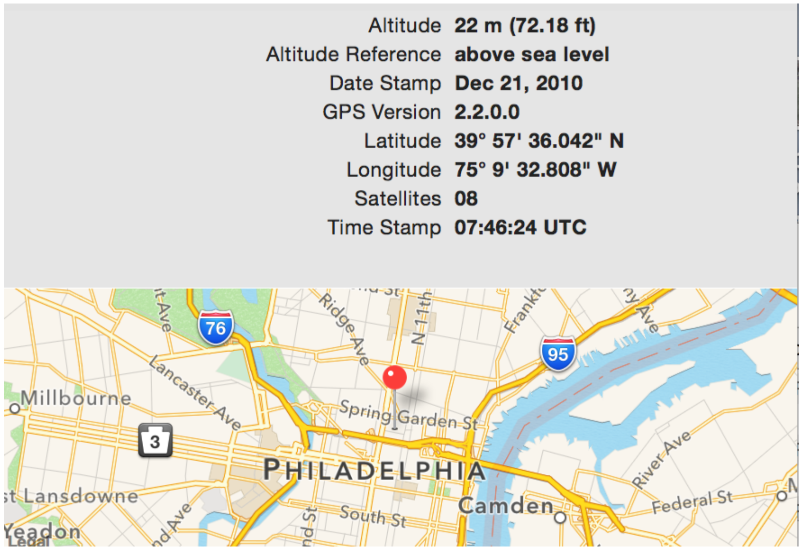
- But it can also include the camera model and serial number and GPS location

A Lunar Eclipse

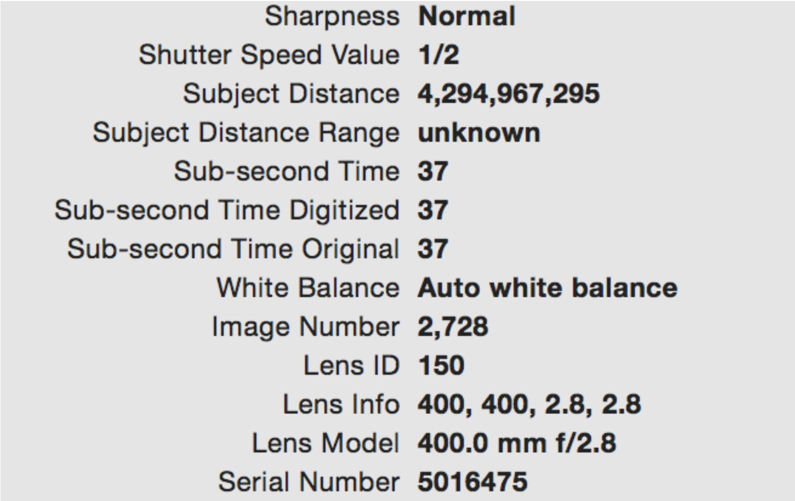


(Photo by Matt Blaze)

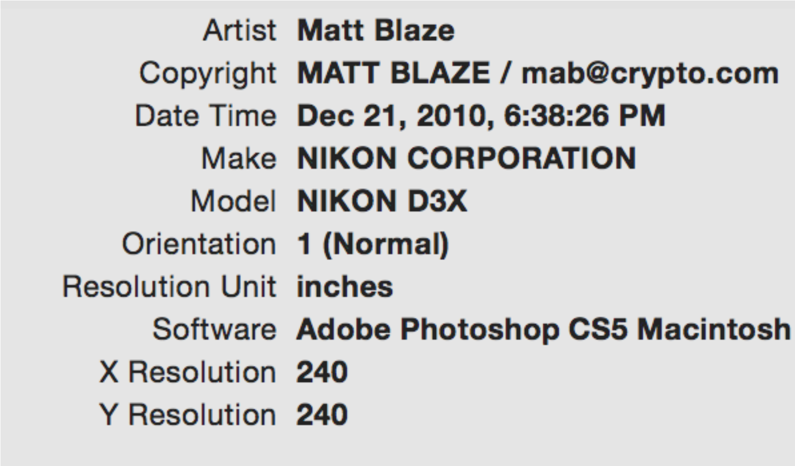
Some of the Metadata



Altitude	22 m (72.18 ft)
Altitude Reference	above sea level
Date Stamp	Dec 21, 2010
GPS Version	2.2.0.0
Latitude	39° 57' 36.042" N
Longitude	75° 9' 32.808" W
Satellites	08
Time Stamp	07:46:24 UTC



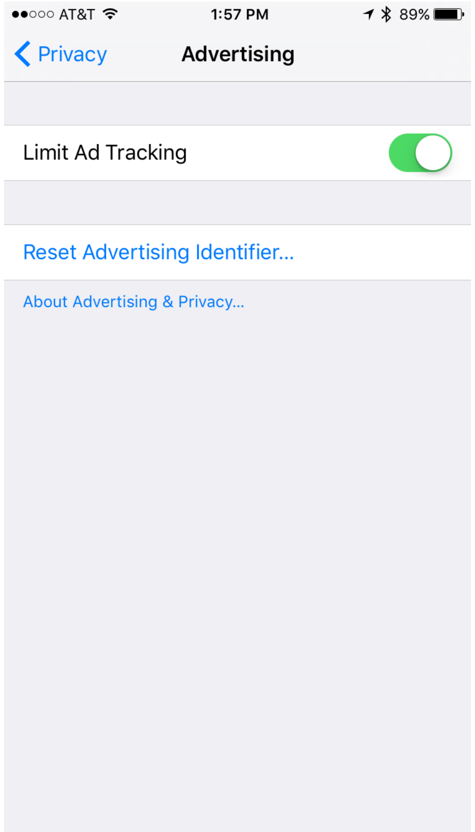
Sharpness	Normal
Shutter Speed Value	1/2
Subject Distance	4,294,967,295
Subject Distance Range	unknown
Sub-second Time	37
Sub-second Time Digitized	37
Sub-second Time Original	37
White Balance	Auto white balance
Image Number	2,728
Lens ID	150
Lens Info	400, 400, 2.8, 2.8
Lens Model	400.0 mm f/2.8
Serial Number	5016475



Artist	Matt Blaze
Copyright	MATT BLAZE / mab@crypto.com
Date Time	Dec 21, 2010, 6:38:26 PM
Make	NIKON CORPORATION
Model	NIKON D3X
Orientation	1 (Normal)
Resolution Unit	inches
Software	Adobe Photoshop CS5 Macintosh
X Resolution	240
Y Resolution	240

Eliminate Linkable Identifiers

- Don’t use the same login name on different sites
- Maybe even have multiple logins on the same site, though that’s often hard to manage
- Delete persistent identifiers when possible, e.g., from photos


●●○○○ AT&T
1:57 PM
89%
< Privacy
Advertising

Limit Ad Tracking ●

Reset Advertising Identifier...
About Advertising & Privacy...

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
- Most often, it's companies trying to make a profit
- Of course, most people are not aware of their data shadows. . .