It's Too Complicated: How the Internet Upends *Katz*, *Smith*, and Electronic Surveillance Law

Steven M. Bellovin
https://www.cs.columbia.edu/~smb

Joint work with Matt Blaze,
Susan Landau, and Stephanie Pell

Content versus Metadata

- Under US law, phone calls are strongly protected
- However, the phone numbers you dial—the metadata—are only weakly protected
- Does the same distinction apply to the Internet? How? Why? What is content? What is metadata?

The Internet: It's Not the Phone System

- More services than "make or answer a call"
- Fundamentally different architecture:
  - Packet-switched, not circuit-switched
  - Smart hosts, dumb network
- Historically, not regulated
- Highly innovative applications

Why should the legal rules be the same? They can't be…

US Legal Basics

- The US, of course, has a formal, written constitution
  - It was written by well-educated men who were very familiar with 500 years of British history
- Laws and government practices must be consistent with the constitution
- The courts in general, and the Supreme Court in particular, can declare something unconstitutional
  - Oddly enough, that power is not stated in the constitution
  - In 1803, the Supreme Court asserted it had that power—but that was really a political maneuver…

The Fourth Amendment

"The right of the people to be secure in their persons, houses, papers, and effects, against unreasonable searches and seizures, shall not be violated, and no warrants shall issue, but upon probable cause, supported by oath or affirmation, and particularly describing the place to be searched, and the persons or things to be seized."

Searches do not always require a warrant, but they have to be *reasonable*

Are Wiretaps "Searches"?

- *Olmstead v. United States* (1928): No—conversations are not "persons, houses, papers, [or] effects"
- *Katz v. United States* (1967): Yes—wiretaps are an invasion of privacy
  - Assertion by the justices: the *purpose* of the Fourth Amendment was to protect *privacy*, not *objects*
  - "[T]he Fourth Amendment protects people, not places"
- Ergo, a search warrant based on probable cause is needed to tap a phone

Are Pen Registers Searches?

- *Smith v. Maryland* (1979): No—you have "given" the numbers you dial to the phone company
- The "3rd Party Doctrine"
  - Individuals have no privacy interest in information they have voluntarily shared with the someone else
  - No warrant—and hence no "probable cause"—is needed
- "All telephone users realize that they must 'convey' phone numbers to the telephone company, since it is through telephone company switching equipment that their calls are completed."

Pen Registers?

- A “pen register” is a device that records the numbers you dial
- A “trap-and-trace” device records who calls you

The Wiretap and Pen/Trap Acts

- In 1968 and 1986, Congress passed laws implementing the *Katz* and *Smith* rulings. (There have been later amendments.)
- Wiretaps, which need "superwarrants", are for content: "the substance, purport, or meaning of [a] conversation"
  - Police need probable cause and other factors
- Pen registers/trap-and-trace orders are easy to get, and can obtain "dialing, routing, addressing, and signaling" (DRAS)
  - Police must certify that the information "likely to be obtained is relevant to an ongoing criminal investigation"

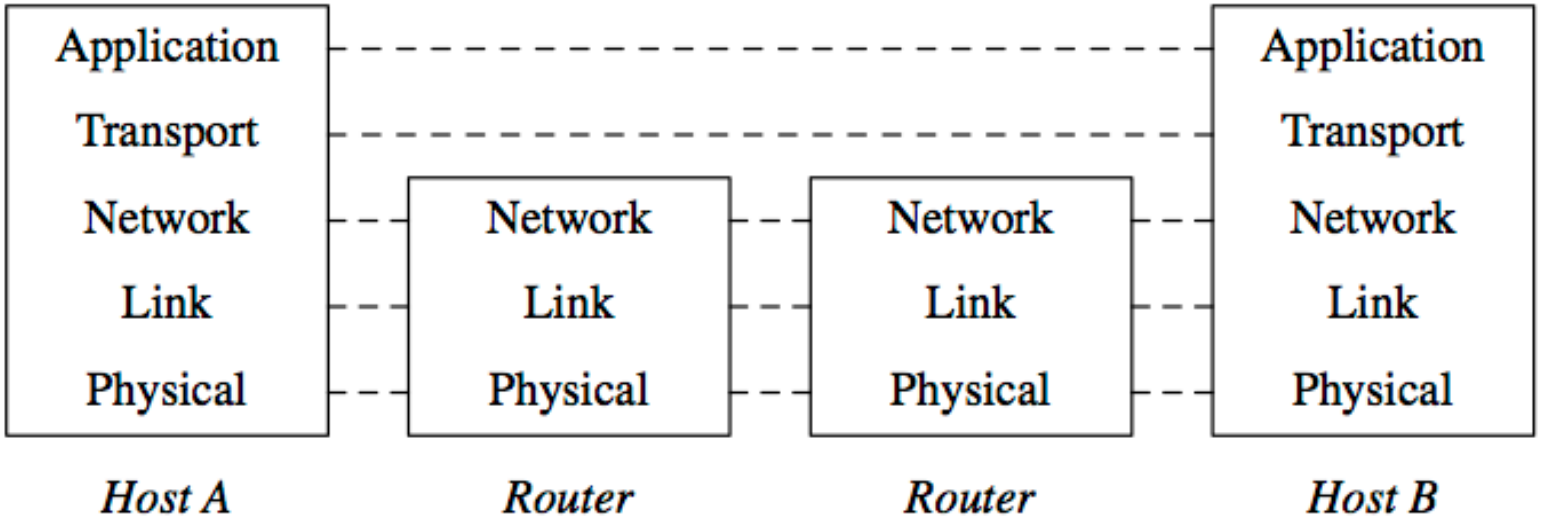
Legality of Data Collection

- For collection to be legal, it must (a) be authorized by statute, and (b) be compatible with the Fourth Amendment
- Collection of content is legal *only* with a warrant (there are minor exceptions)
- Collection of metadata is legal only if there is a warrant *or* (a) It is data "given" to a third party and (b) the statute explicitly permits its collection

What is Metadata on the Internet?

- IP addresses?
- Email addresses?
- What information is *voluntarily* "given"?
- We took a deep, technical look at how the Internet actually works and analyzed the protocols from a legal perspective
  - Note: we looked *only* at the criminal law context; foreign intelligence collection is covered by different legal standards

The Internet: A Layered Architecture



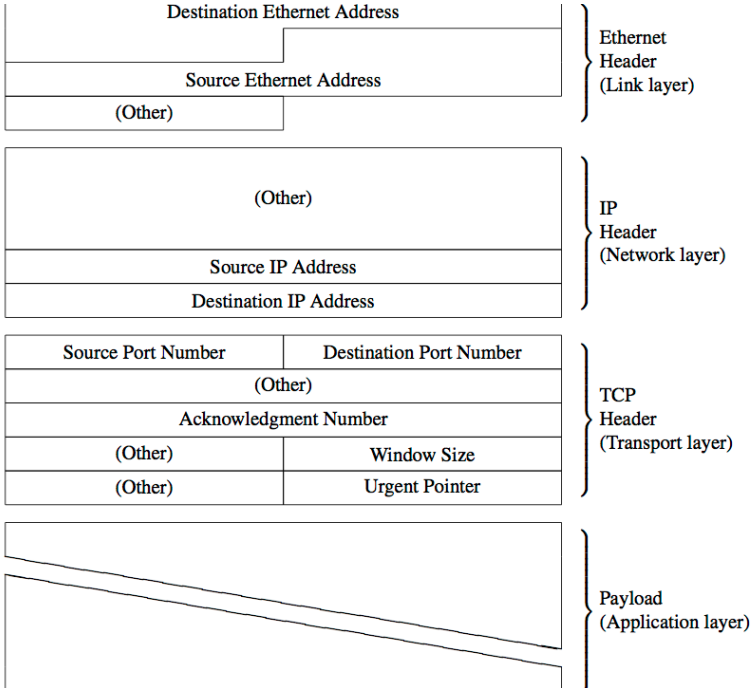
Third Parties?

- The transport and application layers are “end-to-end”; information in them is not given to a third party
- The network layer has lots of third parties
- The link layer is (often) local, so the other end-point istypically another device of yours
- It’s not so simple…

Statutory Definition & Issues

- Some of the transport layer isn't "the substance, purport, or meaning of [a] conversation"
  - Example: TCP port numbers
- Some of the network layer isn't "dialing, routing, addressing, and signaling"
  - Example: the IPid field
- Our definition: "Architectural content" is transported, unexamined, *by some layer in the stack*
- "Architectural metadata" is the complement

Architectural Content



- To the link layer, the IP layer and above are architectural content, even though the IP header contains DRAS.
- To the IP layer, the TCP header is architectural content—but the port numbers turn out to be more complicated

Link Layer Addresses

- Link layer addresses (e.g., WiFi MAC addresses) are often DRAS
- If the device is used at home these addresses are not given to third parties, and hence cannot be collected with a pen/trap order
- But: if used on a public WiFi network, the hotspot operator is a third party

IP Addresses

- Architectural and statutory metadata
- Clearly given to third parties: intermediate routers along the path
- Also: clearly "addressing" information
- But—do most individuals know their computers have IP addresses?
  - In *Smith*, the court noted how much consumers would know from phone bills, phone books, popular culture, etc.

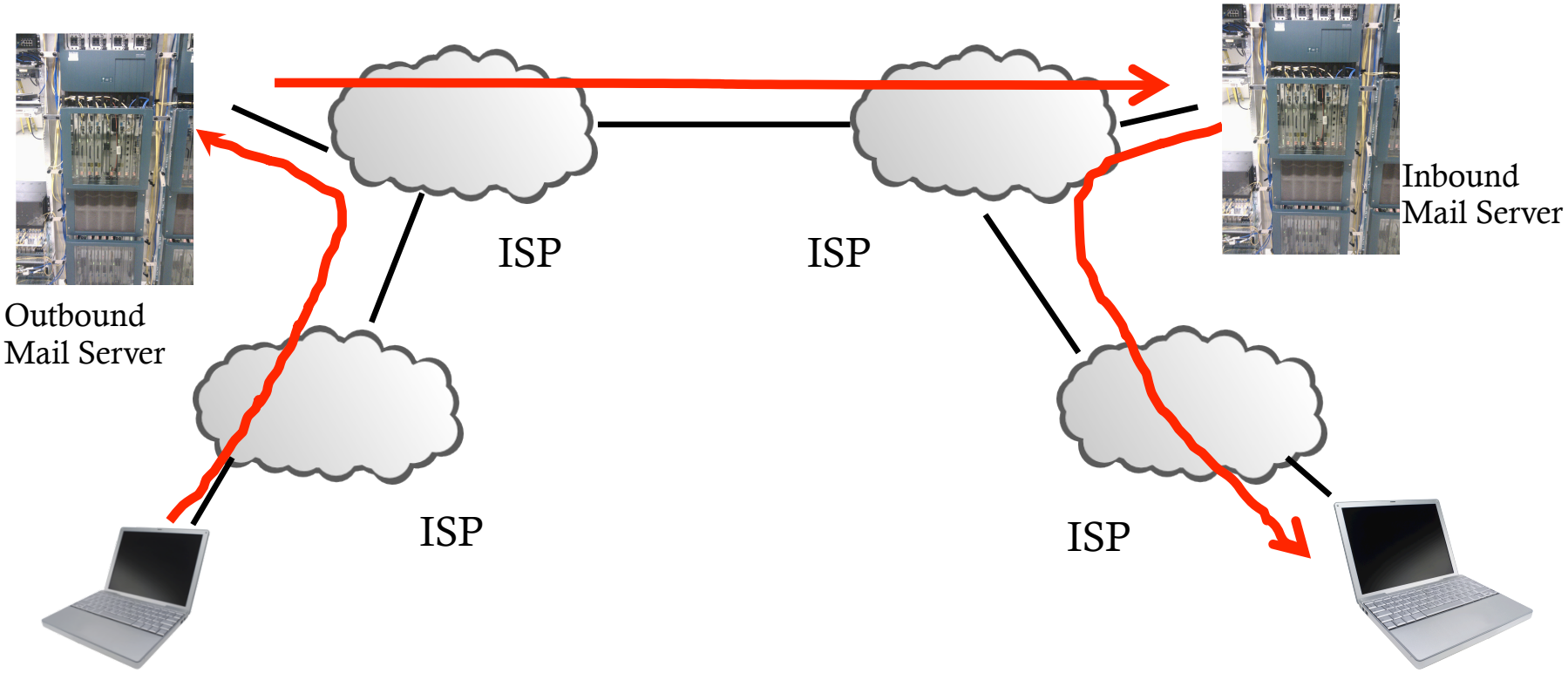
TCP Port Numbers

- Clearly DRAS—but part of the TCP layer, so there are no third parties involved
- Or are there?
  - Mobile phone users' connections go through "carrier-grade NAT" (Network Address Translation), which uses port numbers
  - ISPs monitor and sometimes block based on port numbers
- But—do ordinary users know any of that? Is the conveyance "voluntary"?

Signaling

- Signaling is the exchange of messages that set up a connection
- On the phone network, this is done by phone switches operated by the telephone company
- Internet signaling is done by TCP—and TCP is end-to-end, with no third parties involved
  - But: NATs used on mobile phones do look at the TCP signaling bits. Voluntary and knowing?

Sending Email



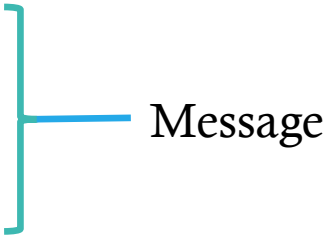
Email (Simplified)

- Mail goes from a sender's device to an "outbound mail server"
- From there, it is sent to the recipient's "inbound mail server"
- The recipient downloads it from that machine
- The mail servers are generally ISP- or enterprise-operated

Sending Myself Email

```
220 machshav.com ESMTP Exim 4.82 Tue, 11 Mar 2014 19:43:03 +0000
HELO eloi.cs.columbia.edu
250 machshav.com Hello eloi.cs.columbia.edu [2001:18d8:ffff:16:12dd:b1ff:feef:8868]
MAIL FROM:<smb@eloi.cs.columbia.edu>
250 OK
RCPT TO:<smb@machshav.com>
250 Accepted
DATA
354 Enter message, ending with "." on a line by itself
From: Barack Obama <president@whitehouse.gov>
To: <smb2132@columbia.edu>
Subject: Test

This is a test
.
250 OK id=1WNSaS-0001z5-1d
QUIT
221 machshav.com closing connection
```

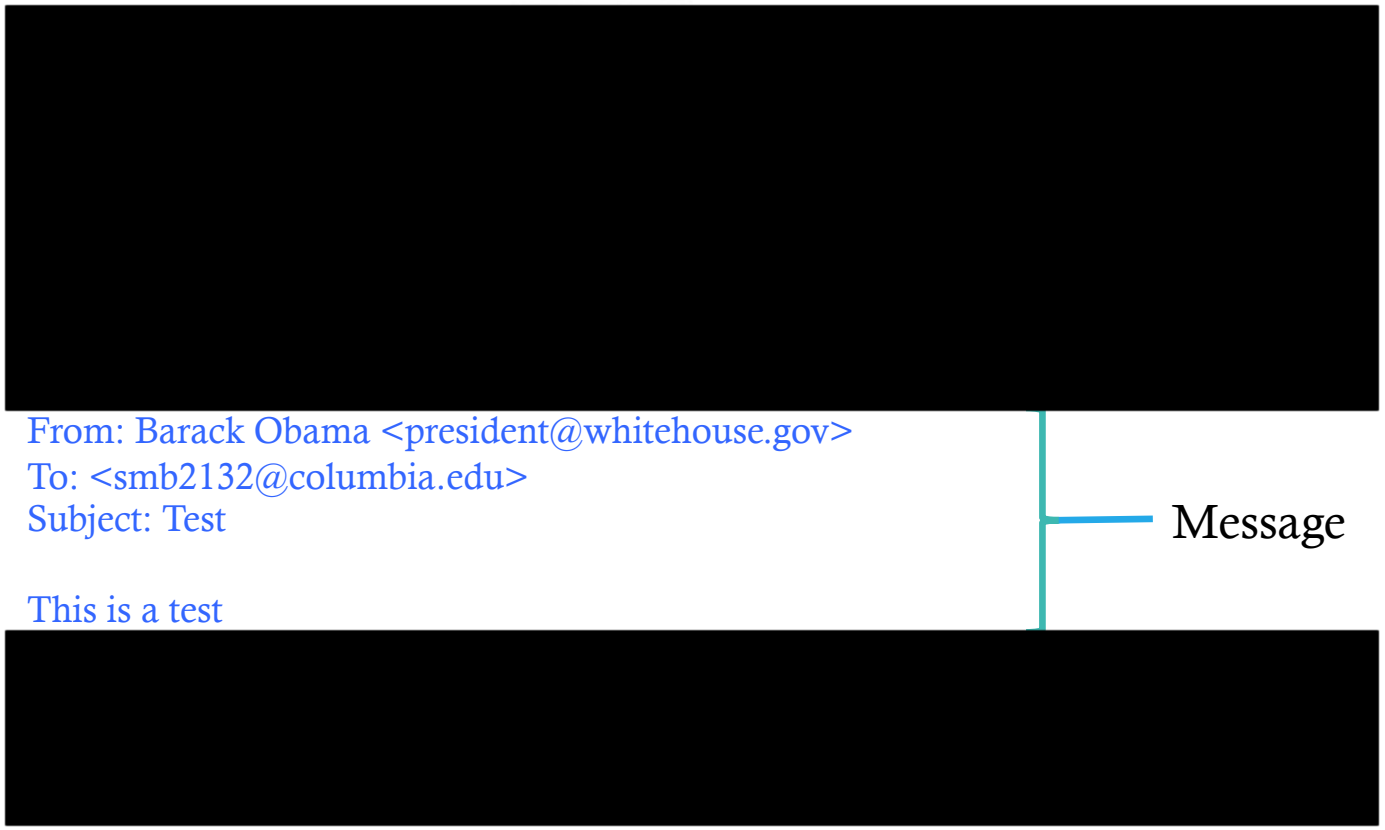


Conversation With A Third Party

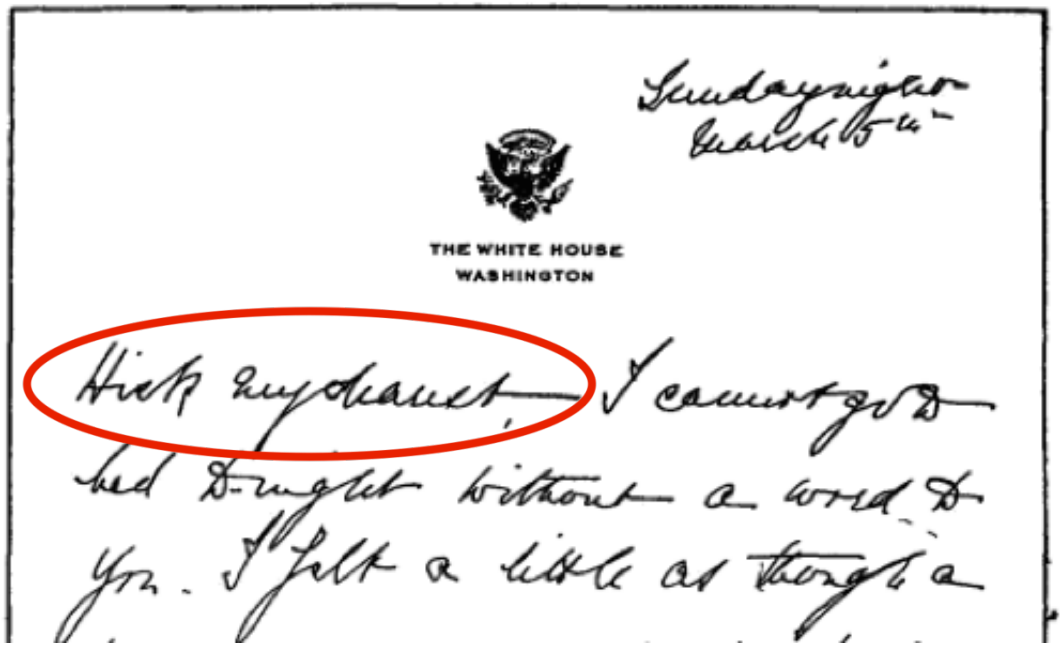
220 machshav.com ESMTP Exim 4.82 Tue, 11 Mar 2014 19:43:03 +0000
HELO eloi.cs.columbia.edu
250 machshav.com Hello eloi.cs.columbia.edu [2001:18d8:ffff:16:12dd:b1ff:feef:8868]
MAIL FROM:<smb@eloi.cs.columbia.edu>
250 OK
RCPT TO:<smb@machshav.com>
250 Accepted
DATA
354 Enter message, ending with "." on a line by itself

.
250 OK id=1WNSaS-0001z5-1d
QUIT
221 machshav.com closing connection

What the Recipient Sees



A Letter from Eleanor Roosevelt to Lorena Hickock (March 1933)



THE WHITE HOUSE
WASHINGTON

Hick my dearest— I cannot go to bed tonight without a word to you. I felt a little as though a

It begins "Hick my dearest".

(excerpt from Amazon.com)

Things to Note

- The SMTP *envelope*—that's the technical term!—can have different information than the message headers
- Unlike the phone network, anyone can run their own mail servers
  - I personally run two, one personal and one professional
  - This complicates third party doctrine analysis
- The reality of email is far more complex than I've outlined here
  - Example: many people read their email via a Web browser—and the NSA has stated that even for them, picking out just the From/To information from a Webmail session is very difficult

Two Sets of Addresses

- To mail servers, the *header* From/To lines are *architectural content*—they belong to a different sublayer
- But courts and the Justice Department have gotten it wrong:
  - DoJ: "Pen register and trap and trace devices may obtain any non-content information . . . Such information includes … the 'To' and 'From' information contained in an e-mail header"
  - Court: "That portion of the "header" which … reveals the e-mail addresses … would certainly be obtainable using a pen register and/or a trap and trace device."

The Web and URLs

- URLs seem simple:

  http://en.wikipedia.org/wiki/Metadata

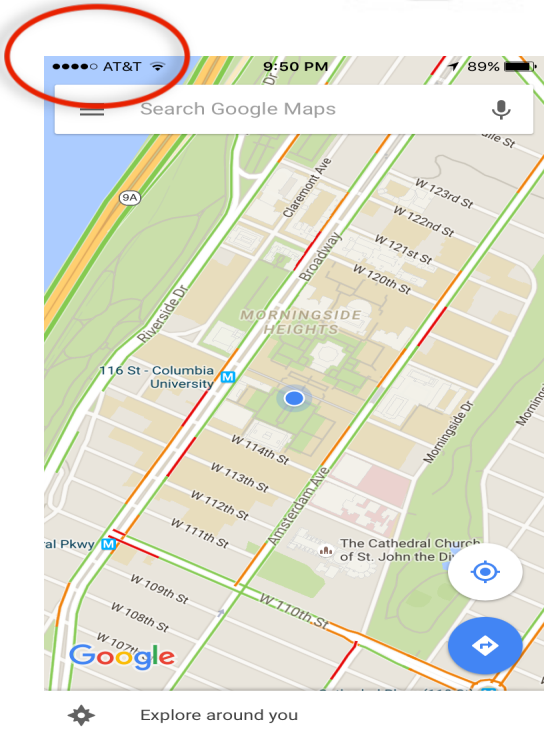
- There's an "authority" (en.wikipedia.org), which seems to be DRAS, and there's a path (/wiki/Metadata), which seems to be content

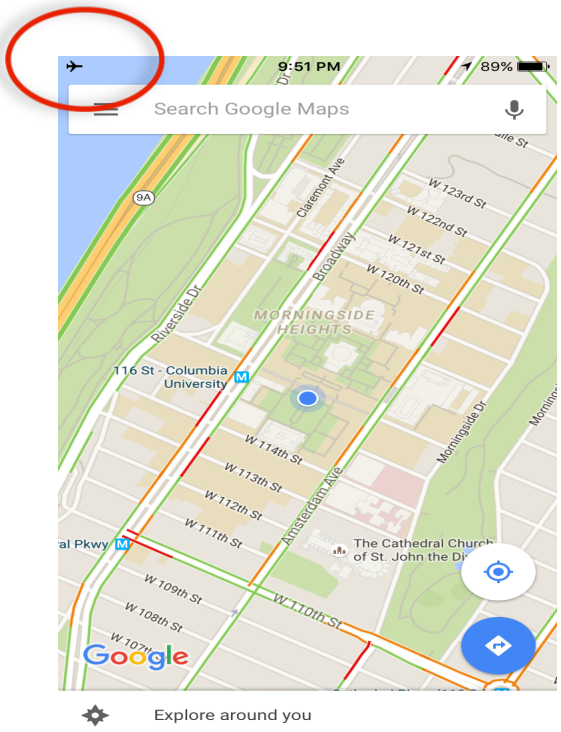
- It's not that simple…

URL Complexities

- If a server hosts multiple web sites, the IP address—which is DRAS—points to the server, but the actual web site goes to the web site operator, who may own all of the sites (i.e., there's no third party)
- When clicking on a Google search result, the real path may go to Google first—a third party!—but the user doesn't know this
- The user cannot tell whether or not there are third parties involved, so the data is not *voluntarily* given

When is Location Sent?





Location: It's Worse Than That

- Even when online, phones can use cached maps
- Even if not downloading maps, WiFi base station identifiers are sent to the server to aid in location determination
- Standalone GPS (satnav) units *never* transmit data
- Is the location conveyance "voluntary", per *Smith*?

Voice over IP

- *Ex Parte Jackson* (1878) said that the "outward form and weight" of a letter or package was not protected
- White *et al.* showed that they could use packet lengths of *encrypted VoIP* conversations to recover some phrases
- Metadata now reveals content

Conclusions

- We have other (and more complex) examples.
- Some information is clearly third party data per *Smith*—but other information is much harder to classify as content or metadata.
- It is very hard to use "voluntariness" as a touchstone
- The content/non-content distinction and the third party doctrine are no longer workable rules for an IP-based communications environment. We need new constitutional and statutory frameworks to govern law enforcement access to wire and electronic communications data.

Our Paper

It's Too Complicated: How the Internet Upends Katz, Smith, and Electronic Surveillance Law

Steven M. Bellovin, Matt Blaze, Susan Landau, and Stephanie K. Pell

Soon to appear in 30 Harvard J. of Law and Technology

Draft at
https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2791646