# 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…

# 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* (277 U. S. 438 (1928)): No—conversations are not "persons, houses, papers, [or] effects"

- *Katz v. United States* (389 U.S. 347 (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

# A "Reasonable Expectation of Privacy"

- "[F]irst that a person have exhibited an actual (subjective) expectation of privacy"
- "[S]econd, that the expectation be one that society is prepared to recognize as 'reasonable.'"

(Justice Harlan, concurring, in *Katz*)

# Are Pen Registers Searches?

- *Smith v. Maryland* (442 U.S. 735 (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."
- No "reasonable expectation of privacy"

# Pen Registers?

- A “pen register” is a device that records the numbers you dial
  - “These devices do not hear sound. They disclose only the telephone numbers that have been dialed.” (*Smith*)
- A “trap-and-trace” device records who calls you

# Pen Registers, Old and New

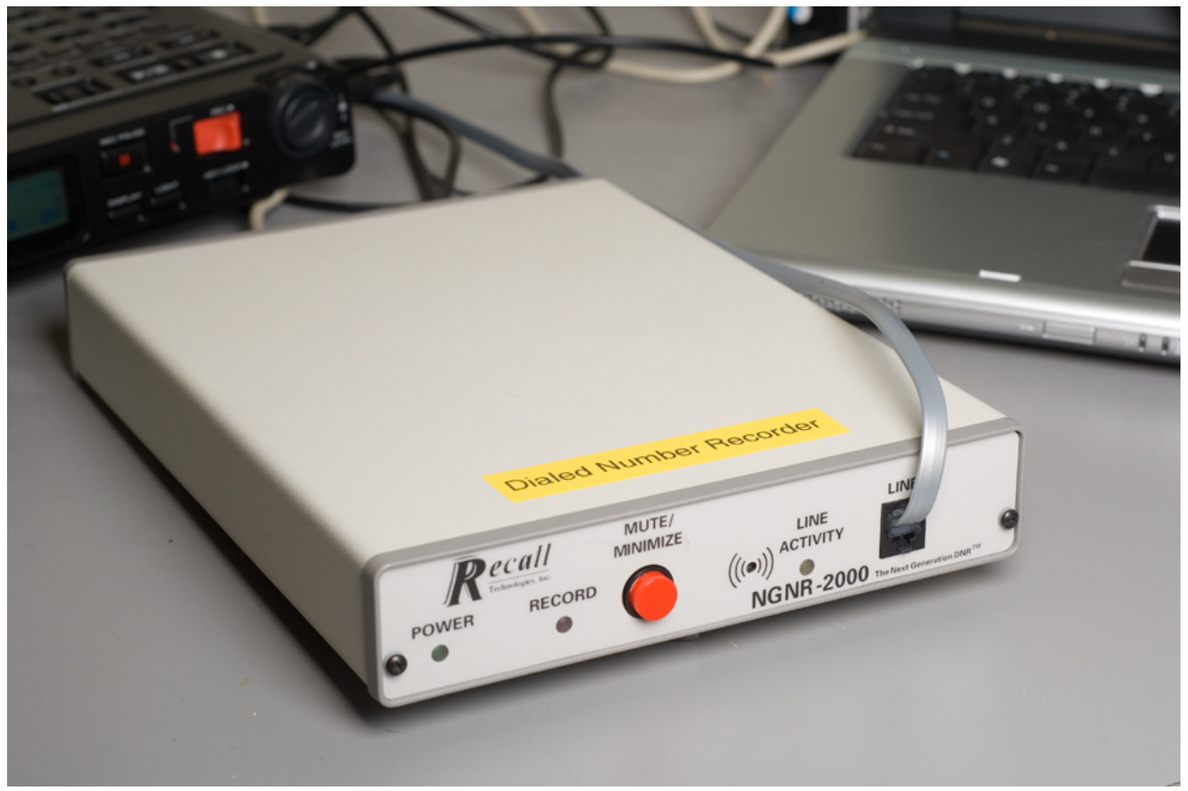




Photo courtesy Matt Blaze

# 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
  - Police need probable cause and other factors
- Pen registers/trap-and-trace orders are easy to get
  - 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. (There may be other (minor) exceptions.)

# What is "Content"? 18 U.S.C. §2511(7)

"'Contents', when used with respect to any wire, oral, or electronic communication, includes any *information concerning the substance, purport, or meaning* of that communication"

[Emphasis added]

# Pen Register
# 18 U.S.C. §3127(3)

"[T]he term "pen register" means a device or process which records or decodes *dialing, routing, addressing, or signaling information* [DRAS] transmitted by an instrument or facility from which a wire or electronic communication is transmitted, provided, however, that such information shall not include the contents of any communication"

[Emphasis and abbreviation added]

# 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

# Content versus DRAS

- To be “content”, it must convey the “substance, purport, or meaning”
- To be collectible via a pen register order, it must be “dialing, routing, addressing, or signaling” information

# Two Types of Content!

**Communicative Content**

- Defined by statute
- "Substance, purport, or meaning"

**Architectural Content**

- "Content" due to the *Internet Architecture*
- Information that is *end-to-end* and not given to a third party

# Why "Architectural Content"?

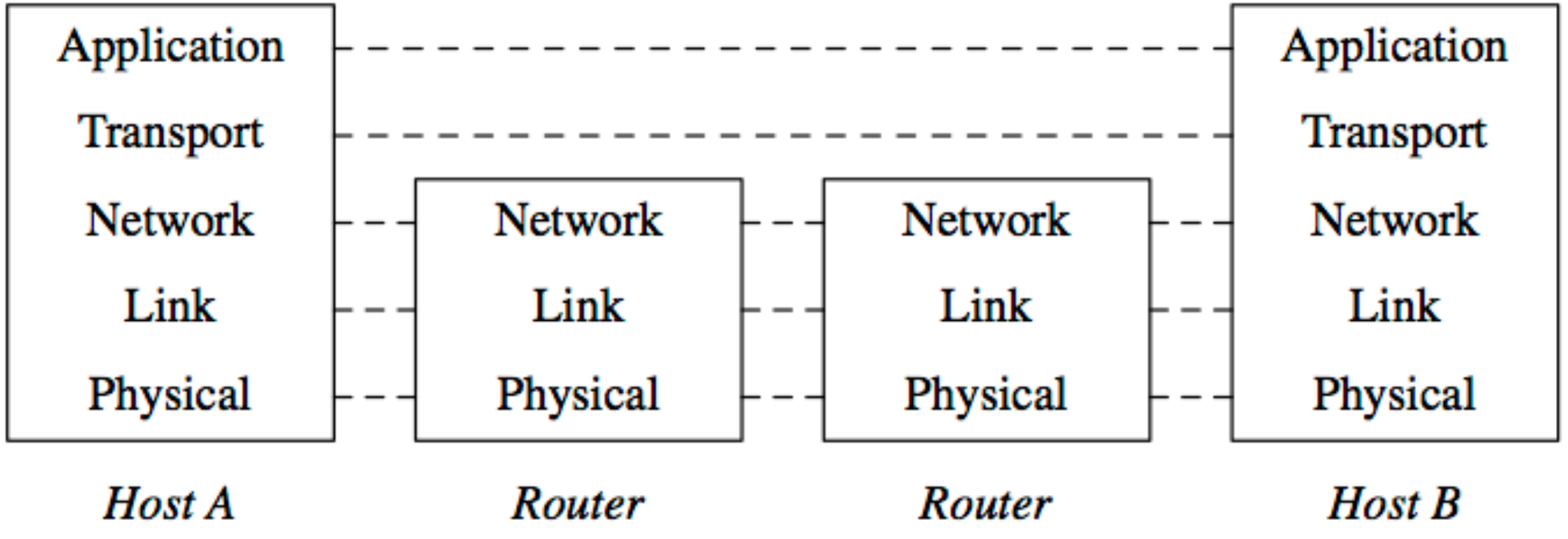
- The Court's reasoning in *Smith* relies on information being "conveyed" to a third party

- There is information that is *not* shared, but is not the "substance, purport, or meaning"

- Because it is not shared with a third party, a warrant may be needed (if, of course, there is no other reason for saying there is no "reasonable expectation of privacy")

- But the Wiretap Act doesn't permit its collection!

# 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 is typically another device of yours
- It’s not so simple…

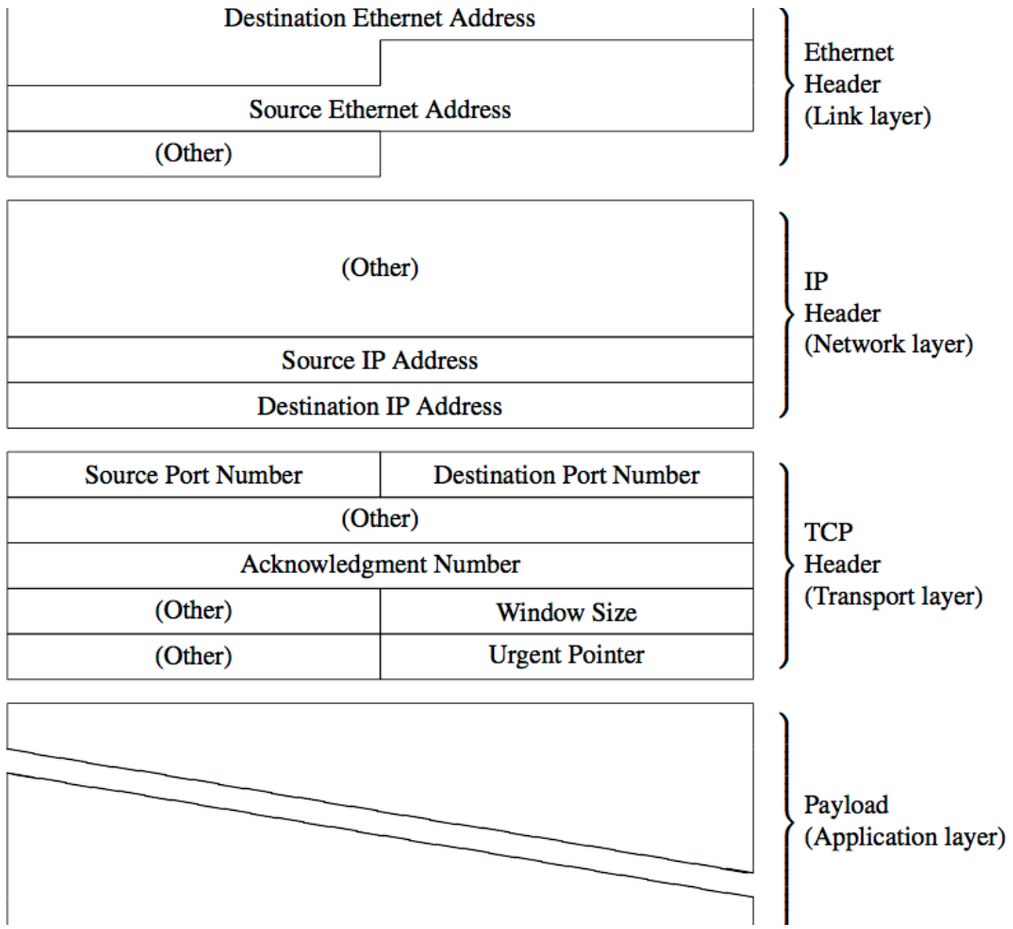
# Intuitive Understanding

- What can the user change without causing problems?
  - If you speak in a non-English language on a US phone, the conversation will still go through
  - If you encrypt the call, it will still go through
  - But: you send different tones to dial, the call will not be completed
- On the Internet, you *must* use a standard network layer, but what's above it is unimportant to the network
- This is oversimplified—things sometimes work differently on the Internet

# 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

# 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 generally not given to third parties, and hence cannot constitutionally 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.
  - Have you ever gotten a bill itemizing IP addresses you connected to?

# What's a Port Number?

- If an IP address is like a building's street address, the port number is the room within the building

- Different ports are used for different services: port 25 is for receiving email, port 80 is a web server, etc.

- In other words, a port number is a *service address*; seeing the port number (often) says why someone is contacting another computer

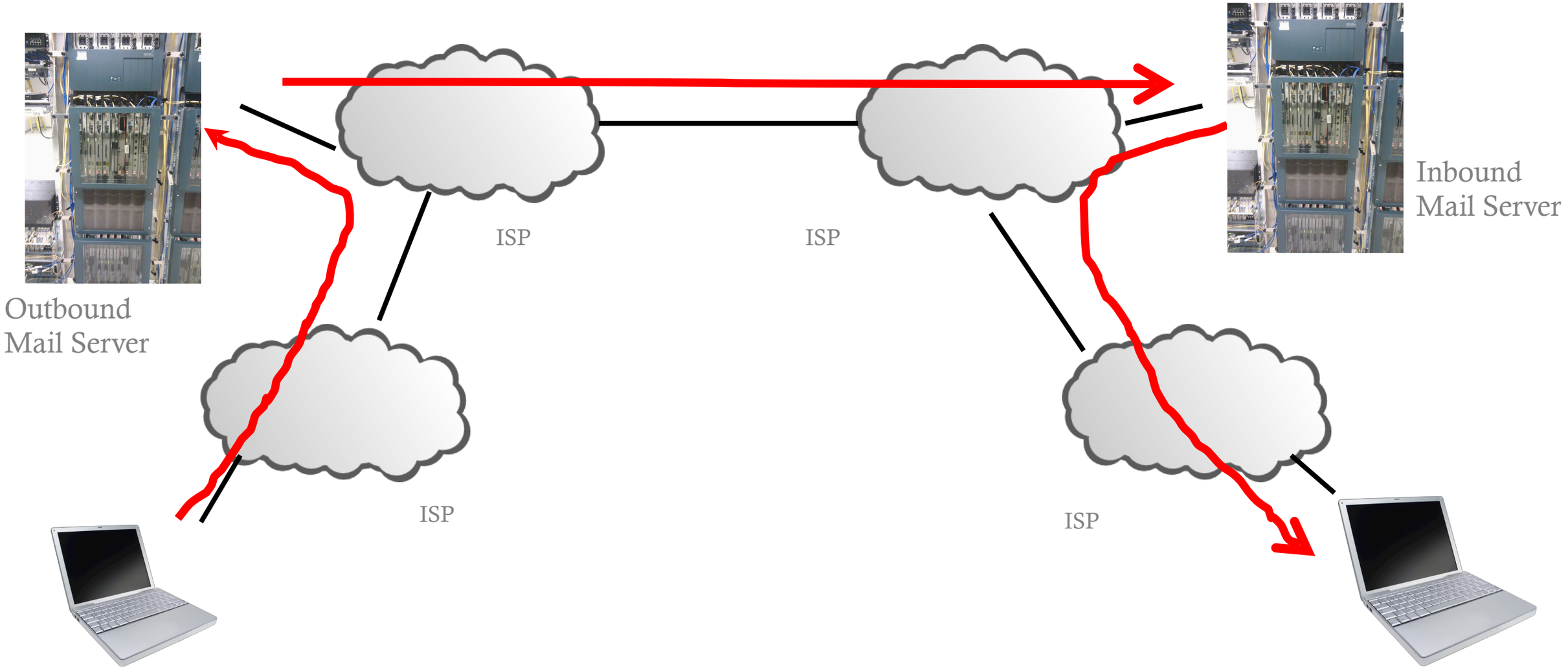
# TCP Port Numbers

- Clearly DRAS—but they're architectural content. They're 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"?
- Port numbers are *taken*, not given!

# 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
  - Signaling is architectural content
  - But: NATs used on mobile phones do look at the TCP signaling fields.  Voluntary and knowing?

# Sending Email



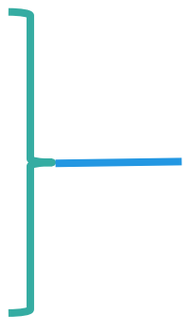
# Email (Simplified)

- 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

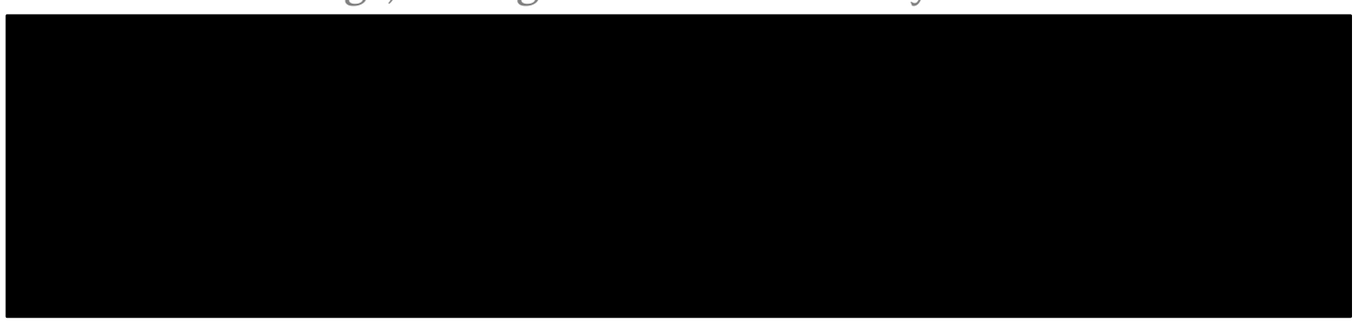
This is a test
.
250 OK id=1WNSaS-0001z5-1d
QUIT
221 machshav.com closing connection



Message

# Conversation With A Third Party

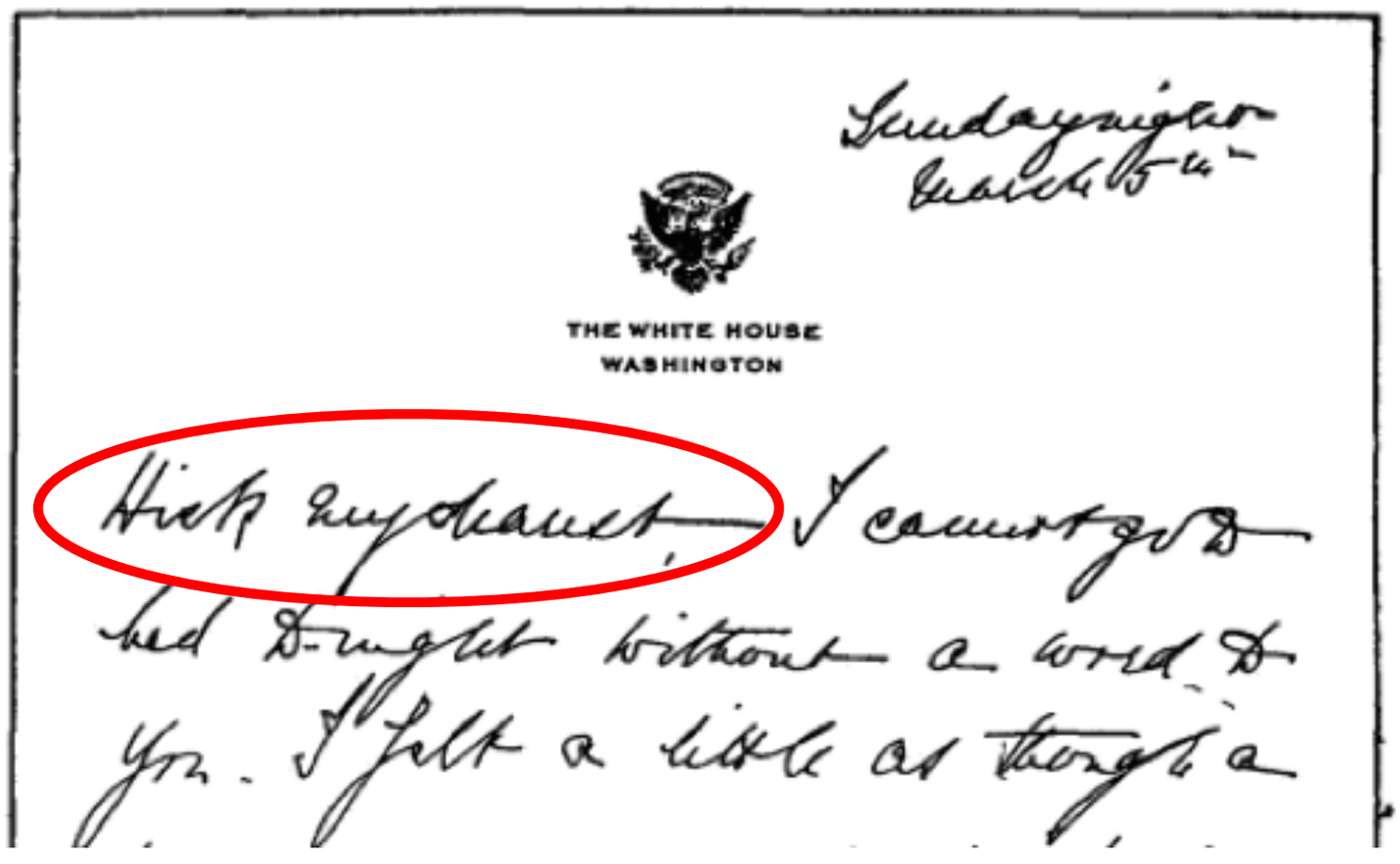
220 machshav.com ESMTP Exim 4.82 Tue, 11 Mar 2014 19:43:03 +0000
HELO eloi.cs.columbia.edu
250 machshav.com Hello eloi.cs.columbia.edu [2001:18d8:ffff:16:12dd:b1ff:feef:8868]
MAIL FROM:<smb@eloi.cs.columbia.edu>
250 OK
RCPT TO:<smb@machshav.com>
250 Accepted
DATA
354 Enter message, ending with "." on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



Sunday night
March 5th

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

(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.”

# Complexity

- The format of an email message, and in particular the format of header lines, is examined and (to some extent) modified by the mail servers
- Are the headers end-to-end or third-party data?
- What do users actually know?

# The Web and URLs

- URLs seem simple:

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

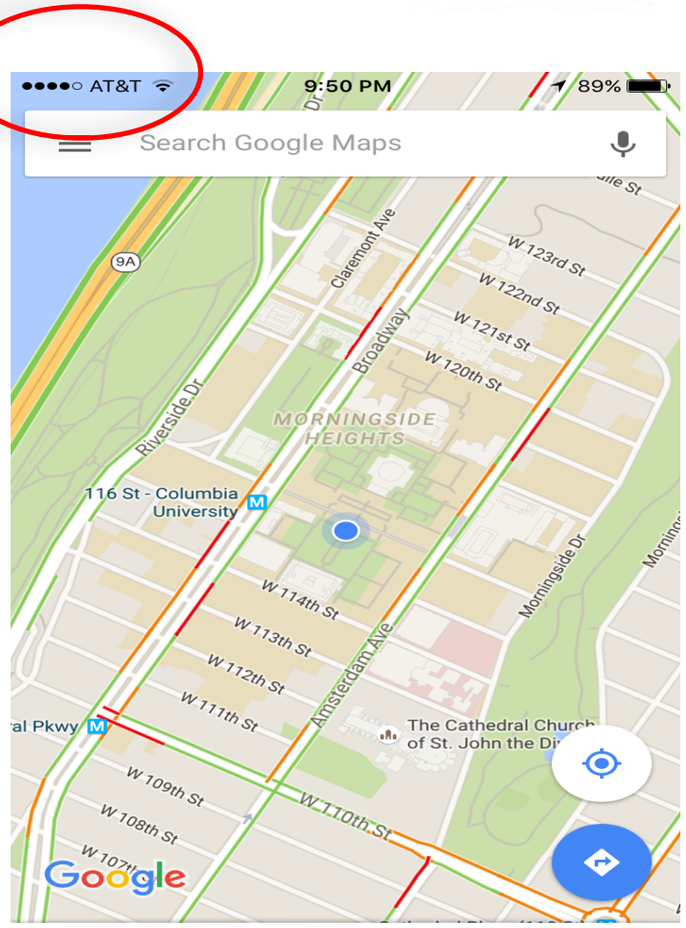
- There's an "authority" (en.wikipedia.org), which seems to be DRAS, and there's a path (/wiki/Metadata), which seems to be content

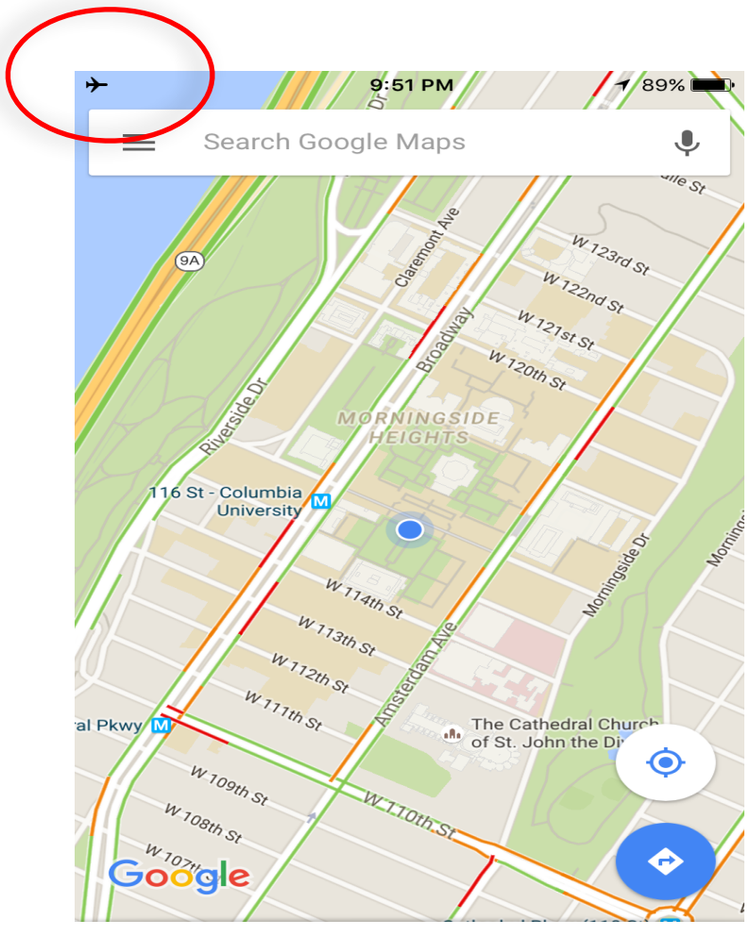
- It's not that simple…

# URL Complexities

- If a server hosts multiple web sites, the IP address—which is clearly 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)
- Is there a *legal* difference between `patents.google.com` and `www.google.com/patents` ?
- 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 units *never* transmit data
- Is the location conveyance "voluntary", per *Smith*? And what about *Carpenter*?

# Voice over IP

- *Ex Parte Jackson* (96 U.S. 727 (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!

# Statutory Problems

- There is information that is end-to-end but is not covered by the statutory definition of “content”
  - No one (including us) has done a “reasonable expectation of privacy” analysis of this information
- There is information that is shared with third parties that is not DRAS
- What are the appropriate legal definitions?

# 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.
  - The classification of some information depends on the situation and the details—you don't know until you look, but without a warrant, you can't look until you know
- It is very hard to use "voluntariness" as a touchstone—and *Smith* relied on voluntariness
- 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

30 Harvard J. of Law and Technology 1

http://jolt.law.harvard.edu/assets/articlePDFs/v30/30HarvJLTech1.pdf

# Questions?