

# IT 4823

## Information Security Administration

### Traffic Analysis



Notice: This session is  
being recorded.

# Traffic Analysis

*“Traffic analysis, not cryptanalysis, is the backbone of communications intelligence.”*

— Susan Landau and Whitfield Diffie

# Six Degrees of Kevin Bacon

- Idea: any Hollywood actor can be linked to Kevin Bacon through mutual film appearances in six or fewer steps.
- It's a parlor game, but it illustrates an important concept.

# Bacon Number

- An application of the Erdős number concept
- Kevin Bacon has a Bacon number of 0.
- Any actor who appeared in a movie with Bacon has a Bacon number of 1.
- Example:
  - Edward Asner has appeared with Bacon; Bacon number 1
  - Elvis Presley has appeared with Asner, but not with Bacon: Presley's Bacon number is 2.

[http://en.wikipedia.org/wiki/Six\\_Degrees\\_of\\_Kevin\\_Bacon](http://en.wikipedia.org/wiki/Six_Degrees_of_Kevin_Bacon)

# The Small World Phenomenon

- Widely separated people can be linked together with amazingly few links.
- Original experiments (1967) by Stanley Milgram and others, mailing letters through a chain.
- People in the United States are connected by “about three” links.
- Computer simulations also give “about three” as the mean path of connections.

# Is This Really True?

Connect Bob Brown to Barack Obama:

1. Bob Brown knows Ralph
2. Ralph knows Terry Sanford
3. Terry Sanford knows Rajiv Shah
4. Rajiv Shah knows Barack Obama

# Another Example

This one is from a few years ago.



# What is "Traffic?"

- Originally, radio messages between military units.
- Now includes many other kinds of connections:
  - Telephone calls
  - Email messages
  - Social network friends
  - Twitter followers
- The world is small, and growing smaller every single day.

<http://die-augeuweide.de/byrds/songop/proshchai.htm>



# Telephone Traffic

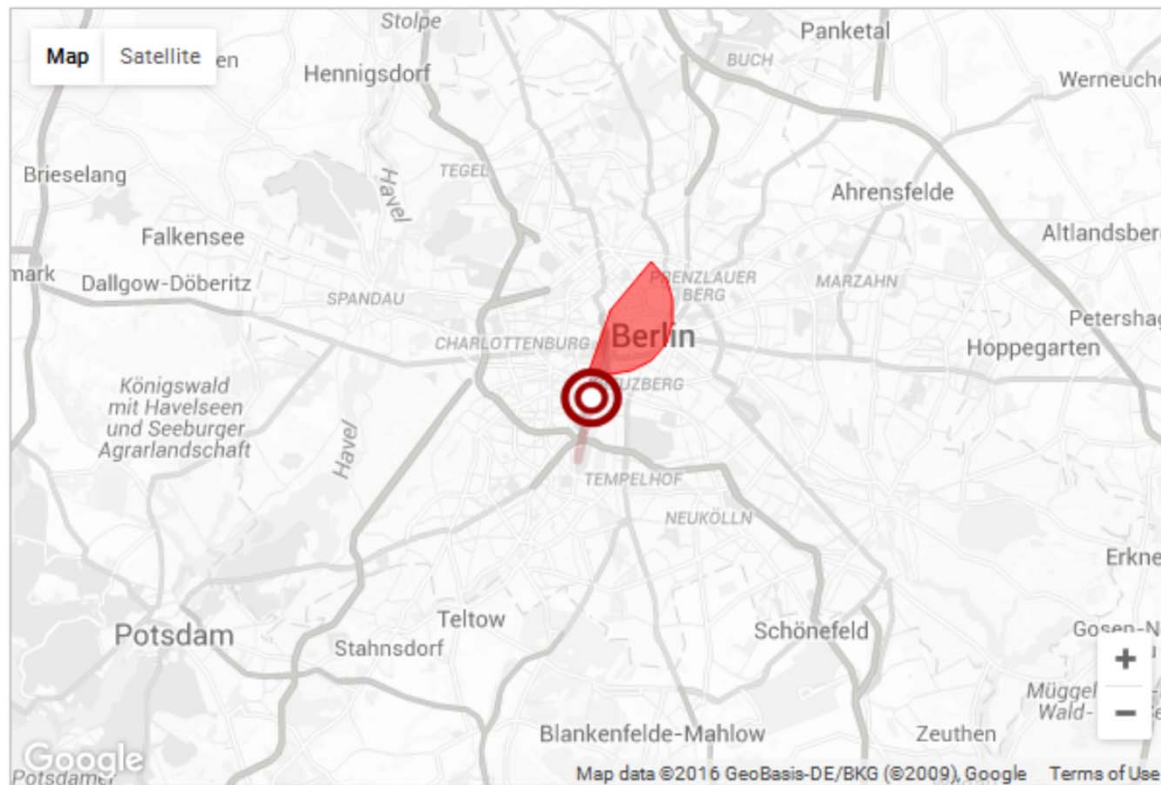
Call detail records (“metadata”):

- Calling number
- Called number
- Completion status
- Start data and time
- Duration
- Bill-to (account) number
- Cell tower identifier(s) when applicable

# Telephone Traffic: What Else?

- Phone numbers increasingly identify people, not places.
- One person, multiple numbers
  - Work
  - Home
  - Cellular
- “Reference” data:
  - Number to name, address, credit information
  - Cell tower identifier to approximate location

# Cell Phone Metadata Analysis



**Monday, 31 August 2009**



Malte Spitz gives a speech to the Greens in Erfurt against internet censorship.  
(source: [Parteiwebsite](#))



6 incoming calls  
21 outgoing calls  
total time: 1h 16min 8s



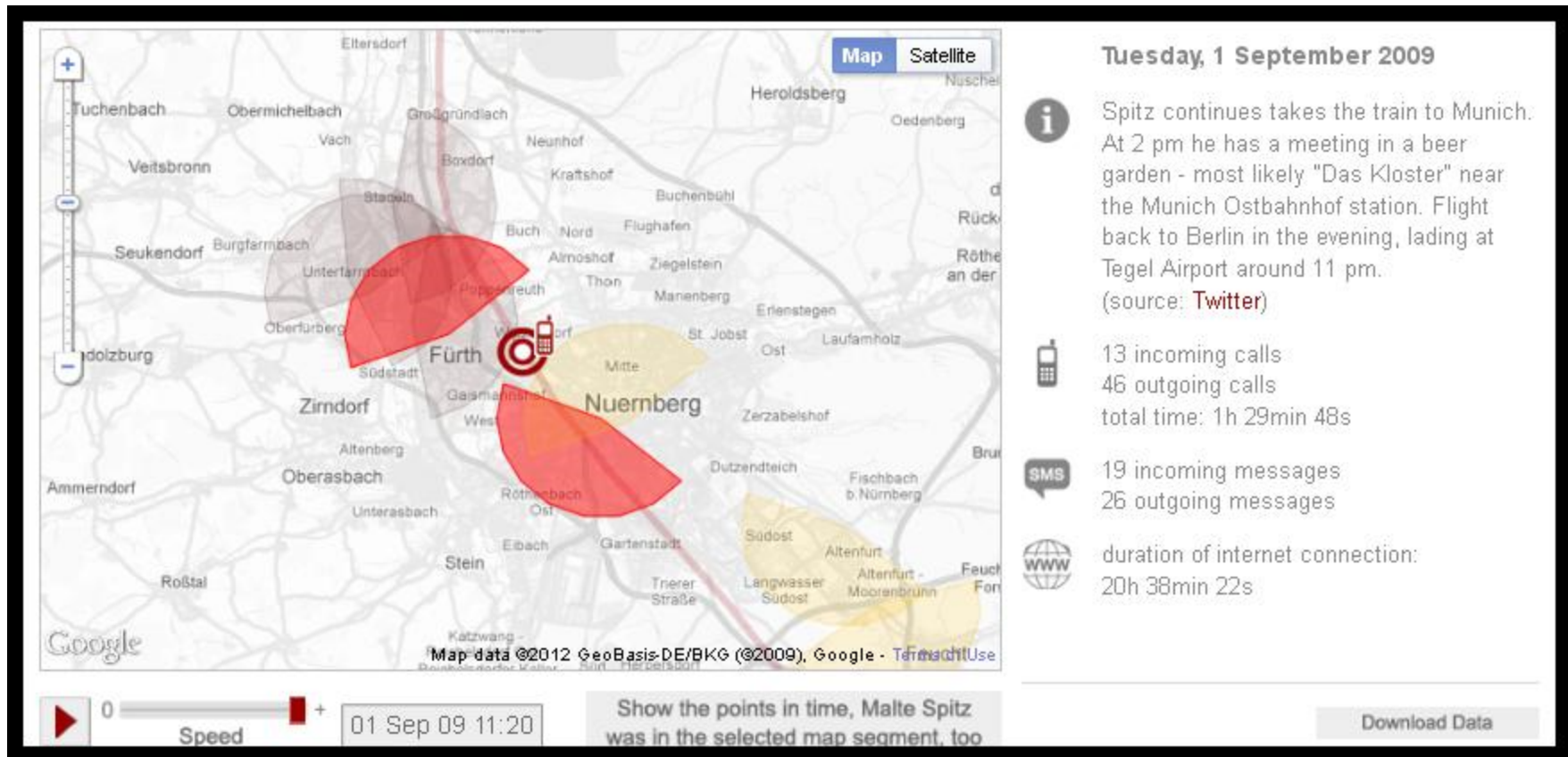
34 incoming messages  
29 outgoing messages



duration of internet connection:  
21h 17min 25s

<http://www.zeit.de/datenschutz/malte-spitz-data-retention>

# Cell Phone Metadata Analysis



# Email Traffic

An email header tells:

- Sender's email address
- Sender's IP number, often
- Sender's email provider
- Recipient's email address
- Recipient's email provider

# Email Traffic: What Else?

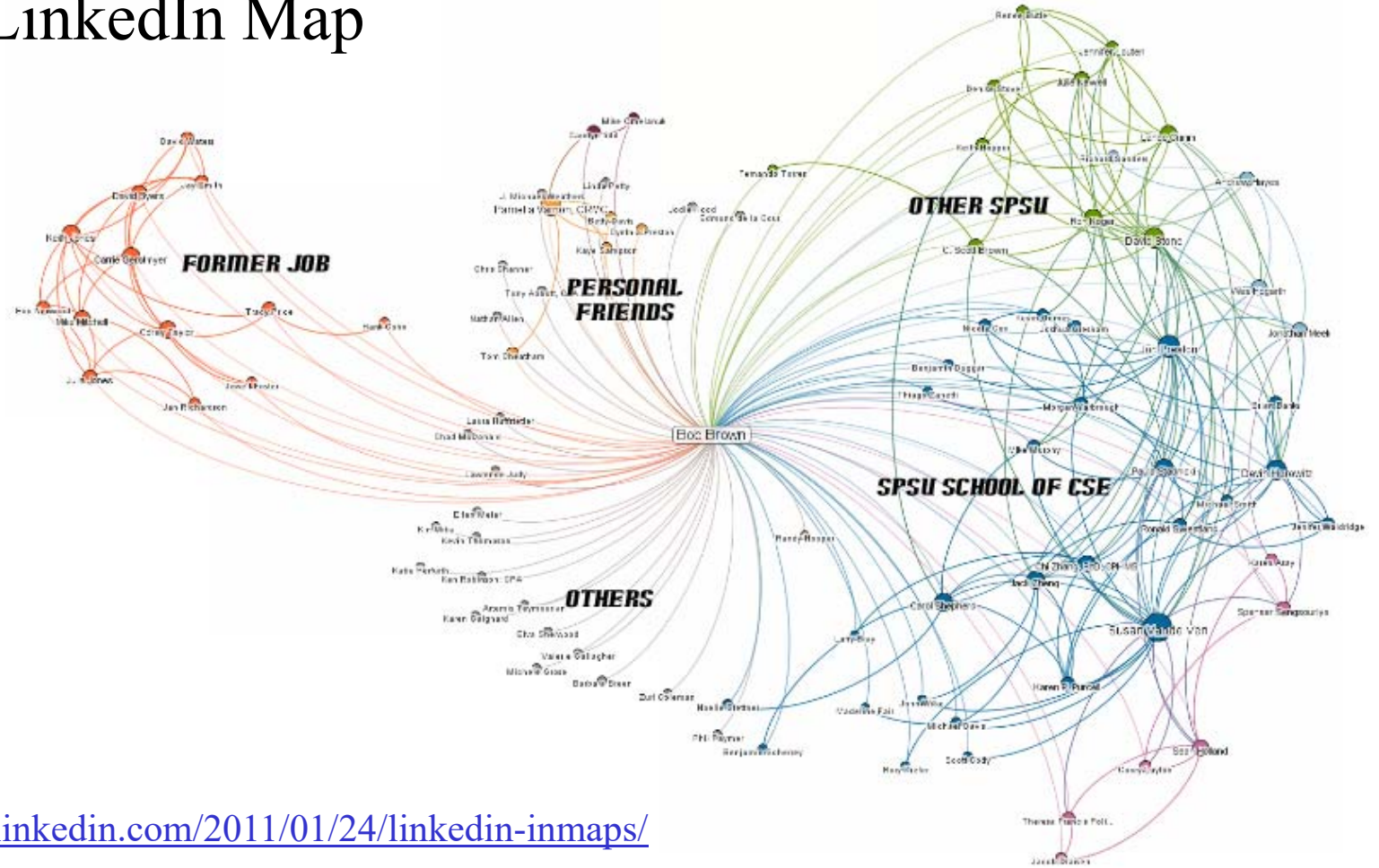
- Real names, possibly
- Where one works, possibly
- Approximate location (from origin IP address)
- Other contacts?
- “Reverse traffic” (*i.e.* replies) reveals recipient’s approximate location.

# Other Network Traffic

- Port numbers indicate protocols even if traffic is encrypted.
- Packet sizes characterize traffic
  - Many very small packets: interactive session; timing between keystrokes reveals content!
  - Large packets one way, small packets the other: file transfer.
- Specific clues
  - Skype will start with many 18 byte UDP packets.
  - Limewire starts with many 35 byte and 23 byte UDP packets

# Social Networks

## Brown's LinkedIn Map



<http://blog.linkedin.com/2011/01/24/linkedin-inmaps/>



# Traffic Analysis

- It's an inference attack.
- Can identify “centers of control”
- Can measure message volume (“chatter”)
  - Increase: a project is being planned
  - Decrease: planning complete, or fear of interception.

## Example: Operation Quicksilver

- Motivation: Convince the German high command that France would be invaded at Pas-de-Calais, not Normandy.
- Mechanism:
  - Fictitious “divisions” of a few soldiers each were equipped with real radios sending realistic messages.
  - German traffic analysts took these for actual troop units.

## Example: HMS Glorious

- British ship HMS Glorious was evacuating pilots and planes from Norway.
- Traffic analysis indicated two German battleships were moving into the North Sea.
- The British Admiralty discounted the report as unproven.
- HMS Glorious was surprised and sunk.

# Another Inference Example

- Google queries are encrypted with TLS.
- Can traffic analysis provide a basis for inference?

Protocol	Length	Info
DNS	83	Standard query 0xaba1 A configuration.apple.com
DNS	181	Standard query response 0xaba1 CNAME configuration.apple.com.edgekey.net CNAME
DNS	70	Standard query 0xb847 A <b>google.com</b>
DNS	246	Standard query response 0xb847 A 173.194.32.9 A 173.194.32.8 A 173.194.32.14 A
LLMNR	84	Standard query 0xb890 A wpad
LLMNR	64	Standard query 0xb890 A wpad
LLMNR	84	Standard query 0x42cc A wpad
LLMNR	64	Standard query 0x42cc A wpad
DNS	70	Standard query 0x3c62 A <b>www.aa.org</b>
DNS	148	Standard query response 0x3c62 CNAME aa.org A 166.78.39.56
DNS	79	Standard query 0xe4bb A ajax.googleapis.com
DNS	129	Standard query response 0xe4bb CNAME googleapis.l.google.com A 74.125.143.95
LLMNR	84	Standard query 0x87a6 A wpad
LLMNR	64	Standard query 0x87a6 A wpad
LLMNR	84	Standard query 0x3f50 A wpad

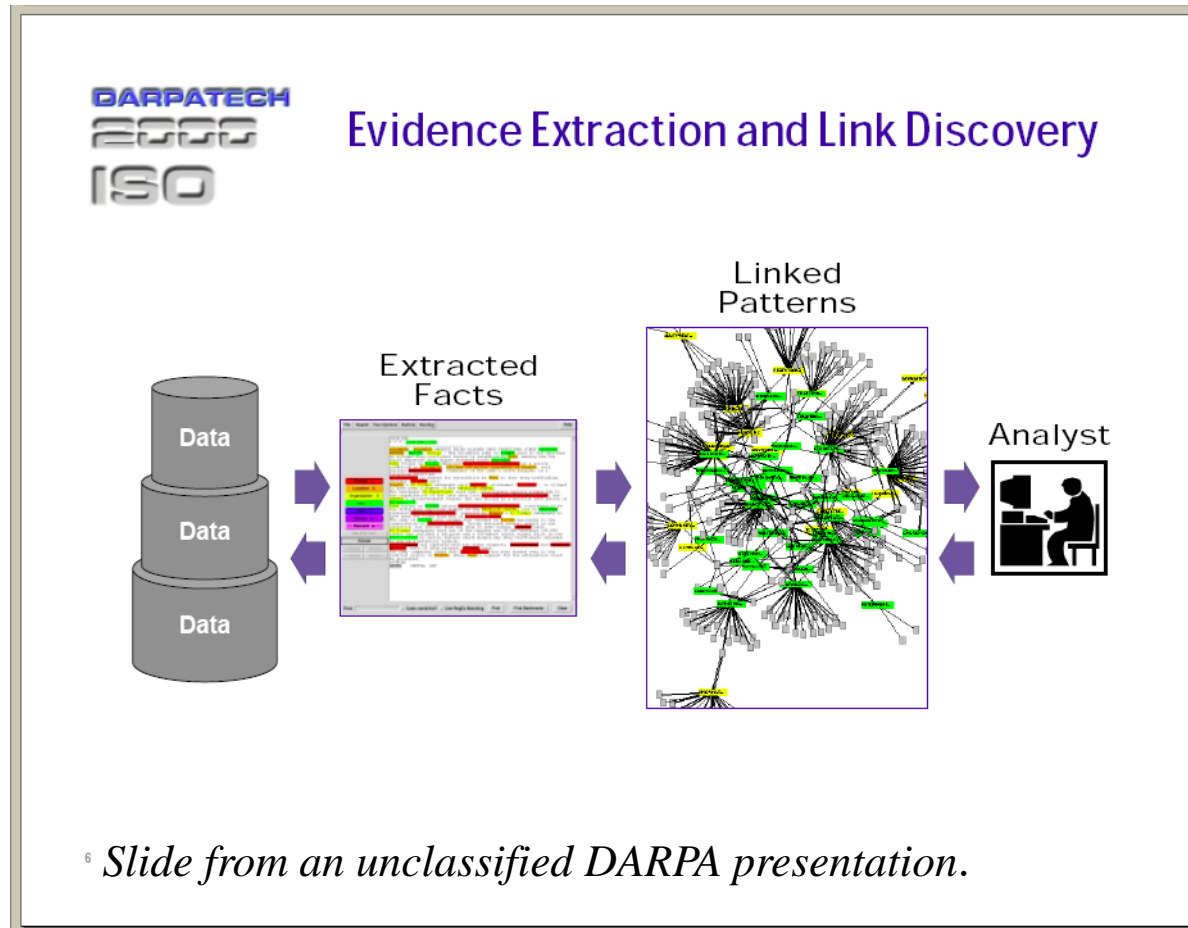
- That's why “metadata” matters!

<http://www.f-secure.com/weblog/archives/00002736.html>

# Distance from a Terrorist

- I have an acquaintance from Faroffistan; she works for HAL Computers.
- I call her with a business question; she calls me back.
- Later she calls home to check on her family.
- Still later, from home, she calls her parents in Faroffistan.
- Her father calls a shopkeeper who is suspected of supporting terrorists.

# Think “They” Can’t Associate it?



# The NSA and Traffic Analysis

- The NSA is known to be collecting call detail data about calls within the U.S.
- The NSA claims the right to “track” such data up to three “hops” from persons suspected of foreign involvement.
- Three hops encompasses almost everyone in the United States! (Remember Milgram?)
- The NSA also collects email header data and social media data.

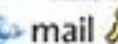


facebook



Hotmail®

YAHOO!

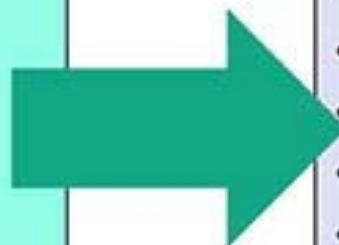


(TS//SI//NF) PRISM Collection Details



Current Providers

- Microsoft (Hotmail, etc.)
- Google
- Yahoo!
- Facebook
- PalTalk
- YouTube
- Skype
- AOL
- Apple



What Will You Receive in Collection  
(Surveillance and Stored Comms)?

It varies by provider. In general:

- E-mail
- Chat – video, voice
- Videos
- Photos
- Stored data
- VoIP
- File transfers
- Video Conferencing
- Notifications of target activity – logins, etc.
- Online Social Networking details
- **Special Requests**

Complete list and details on PRISM web page:

Go PRISMFAA



# Criminals and Traffic Analysis

- The criminal who can monitor your network traffic can determine:
  - Your email provider
  - Your bank
  - The web sites you visit
  - Everything else you do with the Internet
  - Even when you are at home or work.
- Even if the connections themselves are encrypted.

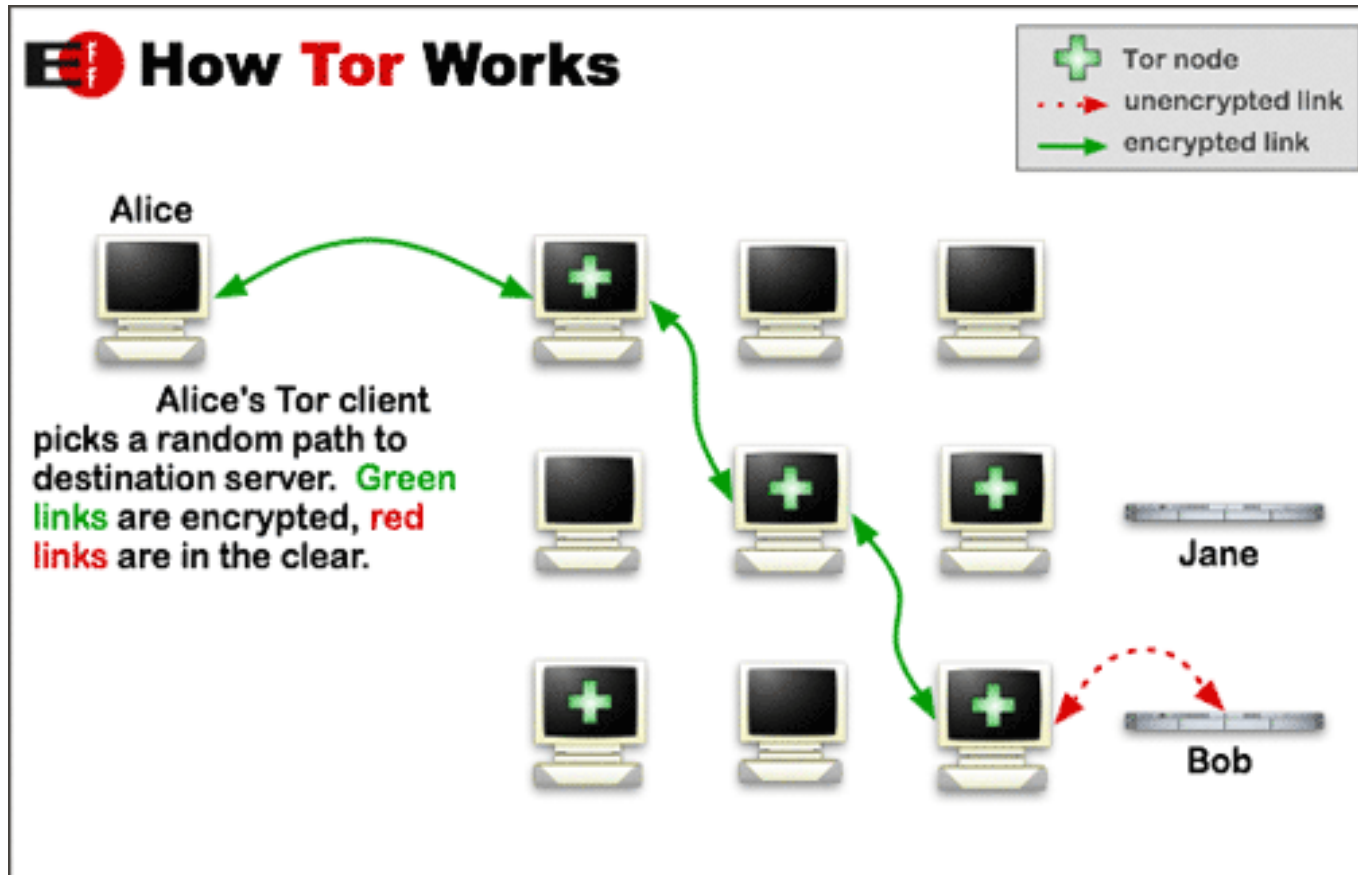
# Defeating Traffic Analysis

- Fictitious traffic, as with Operation Quicksilver.
- Hand-carried messages
- “Burner” cell phones, maybe. (Traffic is captured, but if phones used for short duration, may not be able to be analyzed.)
- Tor networks and remailers are only partially effective, and may be compromised.
- Encrypting links to email servers is effective if the email provider is not compromised.

# Virtual Private Networks

- A VPN encrypts all traffic, including headers.
- The fact of connecting to a VPN is known if the local connection is monitored.
- The VPN exit node may be vulnerable.
- Tor is a kind of multi-hop VPN.

# How Tor Works



## More About Tor

- Originated with DARPA. It's “The onion router.” The idea is that there are layers upon layers of “hops” and encryption.
- In theory, it prevents traffic analysis (and documents leaked by Snowden say it works.)
- If software on the originating computer can be compromised, Tor is useless. The FBI is said to have done this.
- Traffic from exit nodes can be tapped.

# Email: Link vs. Content Encryption

- Link encryption (TLS) protects traffic between end user and mail server, *e.g.* gmail.
  - If the encryption has not been compromised, either with a “back door” or by securing the private key.
  - If the operators of the mail server are not compromised.
  - Link encryption prevents content snooping in transit.
  - Link encryption conceals email headers.
- Content encryption (*e.g.* with GPG) protects the contents of email, but not the headers.

# Summary of Techniques

- TLS: Message content encrypted in transit; source and destination visible.
- Content encryption (GPG, etc.) Content is encrypted end-to-end. Source and destination are visible.
- VPN: All connections are to the VPN; both content and headers encrypted in transit. Exit is unencrypted.
- Tor: Like VPN, but with multiple hops.

# Traffic Analysis and the Law

- Fourth Amendment: are calls and messages “papers?”
- “Pen register” rulings.
- The PATRIOT Act, FISA, and secret courts.
- Encryption and the law.
  - Encryption is absolutely legal in the U.S.
  - The NSA (probably) cannot break strong encryption.
  - Content encryption cannot defeat traffic analysis (but link encryption can frustrate it.)



# Why Should We Care?

- As security professionals, we might use traffic analysis to identify problems, such as leaks from organizations where we work. (You would have your own mail server logs, for example.)
- We might want to defeat traffic analysis to protect our privacy for completely legal reasons, *e.g.* planning a merger.

## Postscript: Mailpile Beta

- Released fall, 2014, updated August, 2015:  
<https://www.mailpile.is/>
- A “front end” for mail services like gmail or any mail service supporting IMAP.
- Integrated support for GPG.
- TLS encryption between your computer and your email service.
- End-to-end encryption of content with GPG.

# Questions

