Scalable Internet Forensics

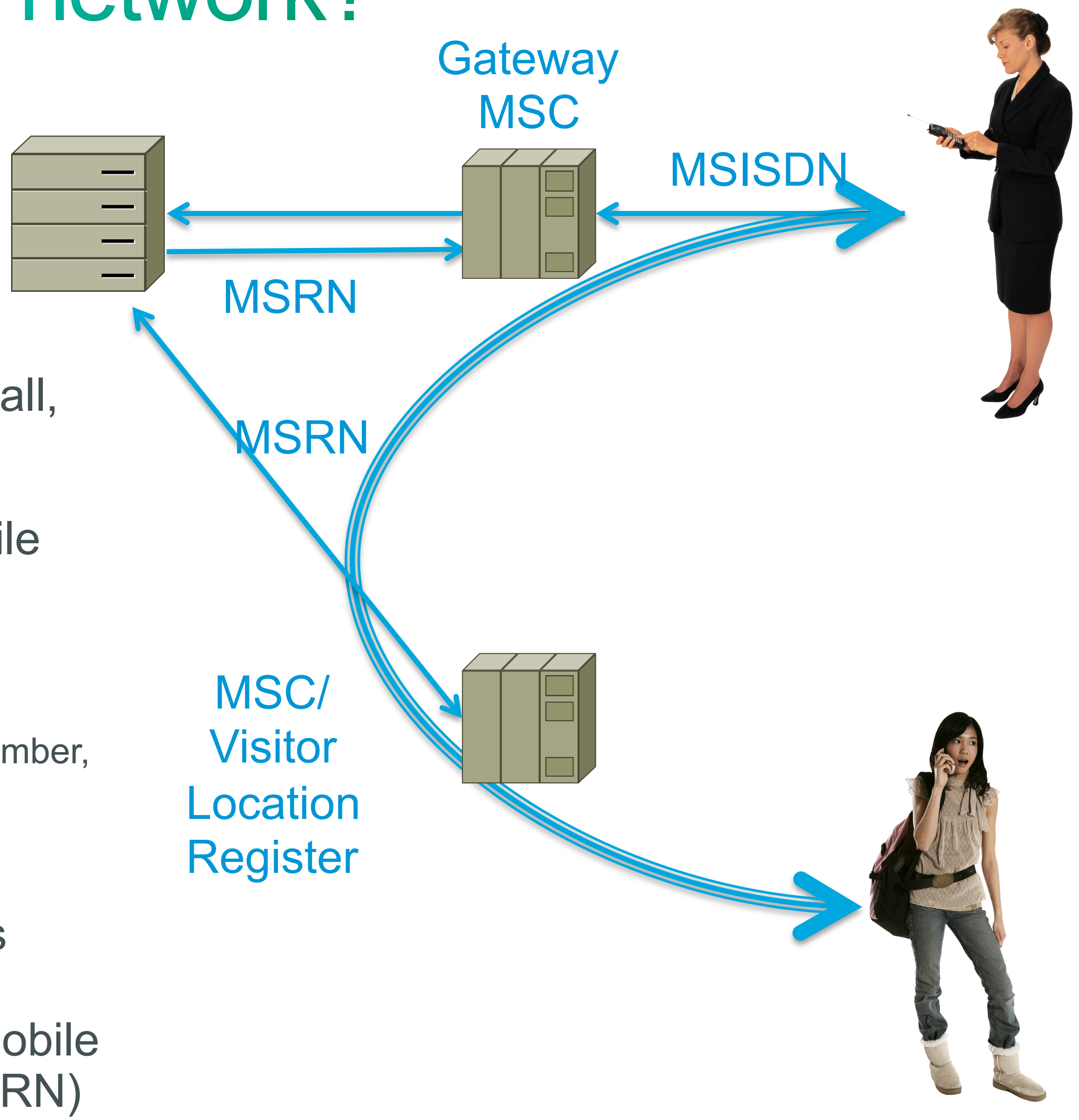
Fred Baker, Cisco Fellow and chair IETF IPv6 Operations WG

Issues I would like to consider today

- The Internet Architecture
  - How it is designed to work
  - What it does for us
- Conducting forensic studies in the Internet
  - How does one use the architecture to conduct forensic studies?
- Regulatory structure that can effectively use that
  - What works
  - What doesn’t
  - Some observations

But first a digression: The mobile telephone system, at a very high level

How does one locate a subscriber in the mobile telephone network?



This is the comic book version…

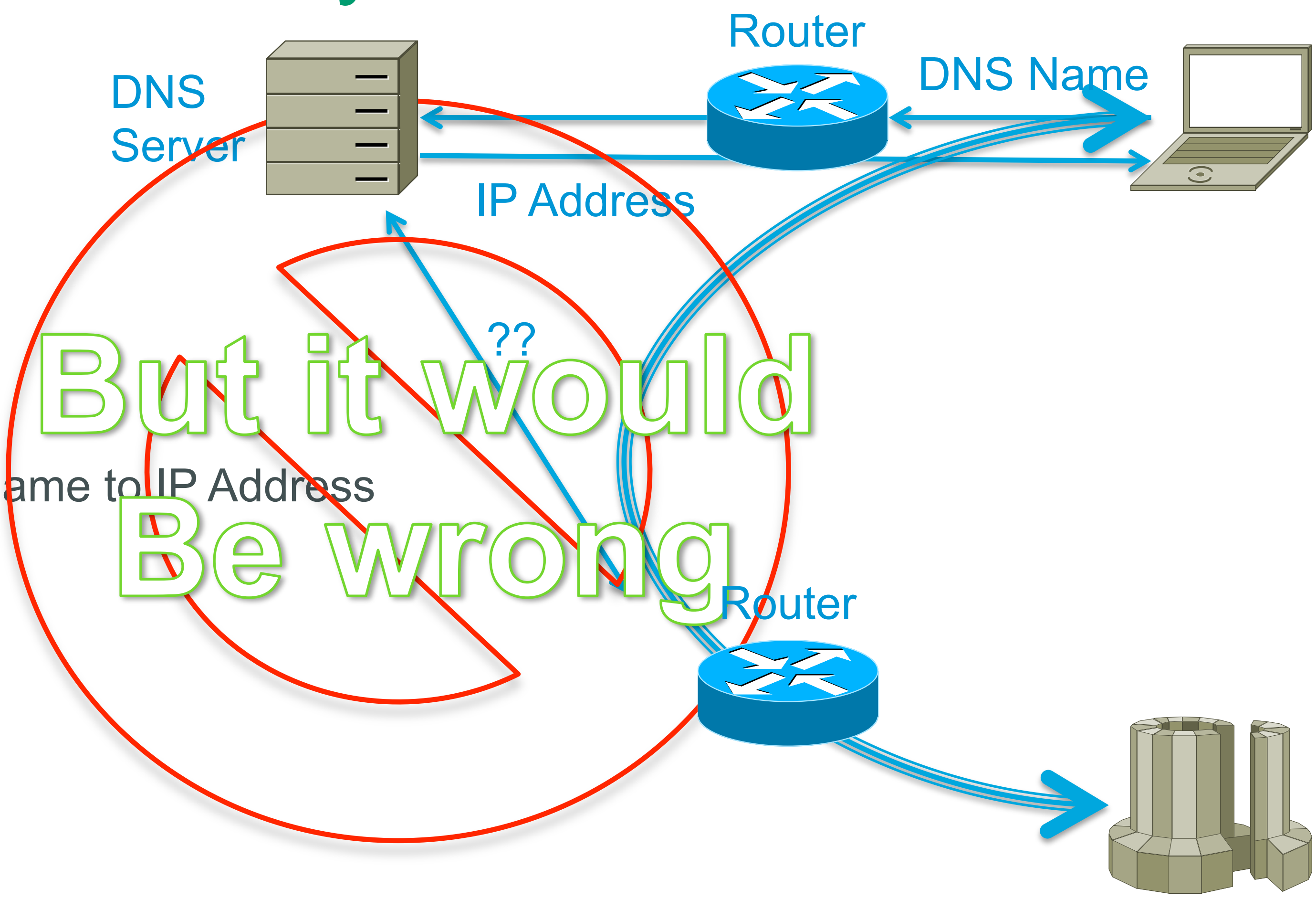
- E.164 Number (MSISDN): a telephone number to route a call, assigned to a subscriber
- E.212 IMSI: International Mobile Subscriber Identity

  A subscriber identifier used for authentication, billing

  Contains Country Code, Network Number, and Subscriber identifier

  64 bit value
- Circuit: a path between parties maintained by the telephone system, uses the temporary Mobile Station Roaming Number (MSRN)

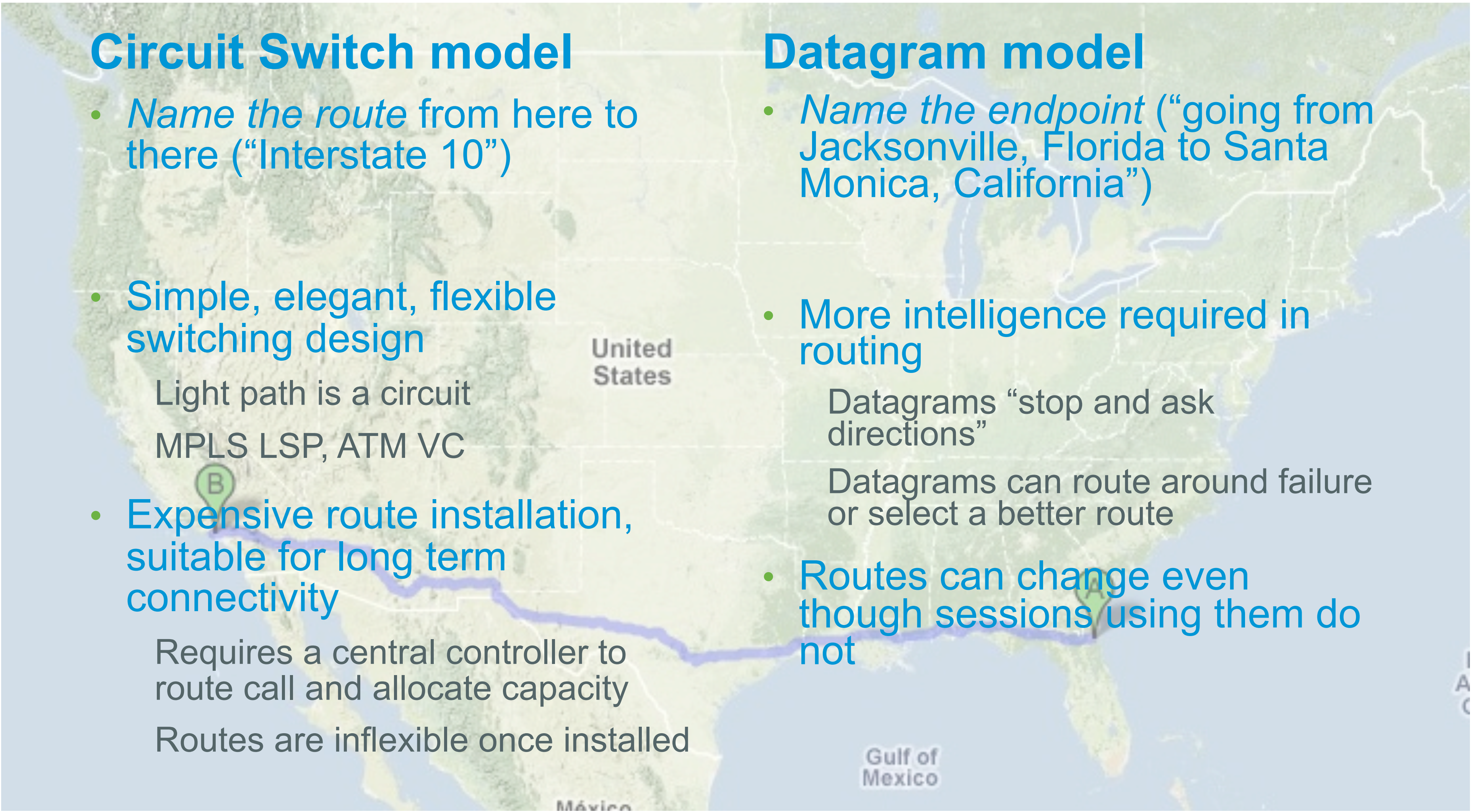
It would be tempting to assume the Internet works the same way…



But it would
Be wrong

- DNS: translates name to IP Address
- IP Address: …

How do we route calls and data?



Circuit Switch model

- *Name the route* from here to there (“Interstate 10”)
- Simple, elegant, flexible switching design
  - Light path is a circuit
  - MPLS LSP, ATM VC
- Expensive route installation, suitable for long term connectivity
  - Requires a central controller to route call and allocate capacity
  - Routes are inflexible once installed

Datagram model

- *Name the endpoint* (“going from Jacksonville, Florida to Santa Monica, California”)
- More intelligence required in routing
  - Datagrams “stop and ask directions”
  - Datagrams can route around failure or select a better route
- Routes can change even though sessions using them do not

The envelope: what does the postman use?

Sender's Address – a place
We usually write the name as well,
But it's not necessary

Street Address or PO Box
City, State, Country, ZIP code

Street Address or PO Box
City, State, Country, ZIP code

My friend
Rahuri
District Ahmednagar
Maharashtra 413705 India

Recipient's Address – a place
We usually write the name as well,
and may use a mail stop or other information
But it's not used by the postal service

How Internet Addressing really works

- **DNS:**

  Translates name to one **or more** IP Addresses

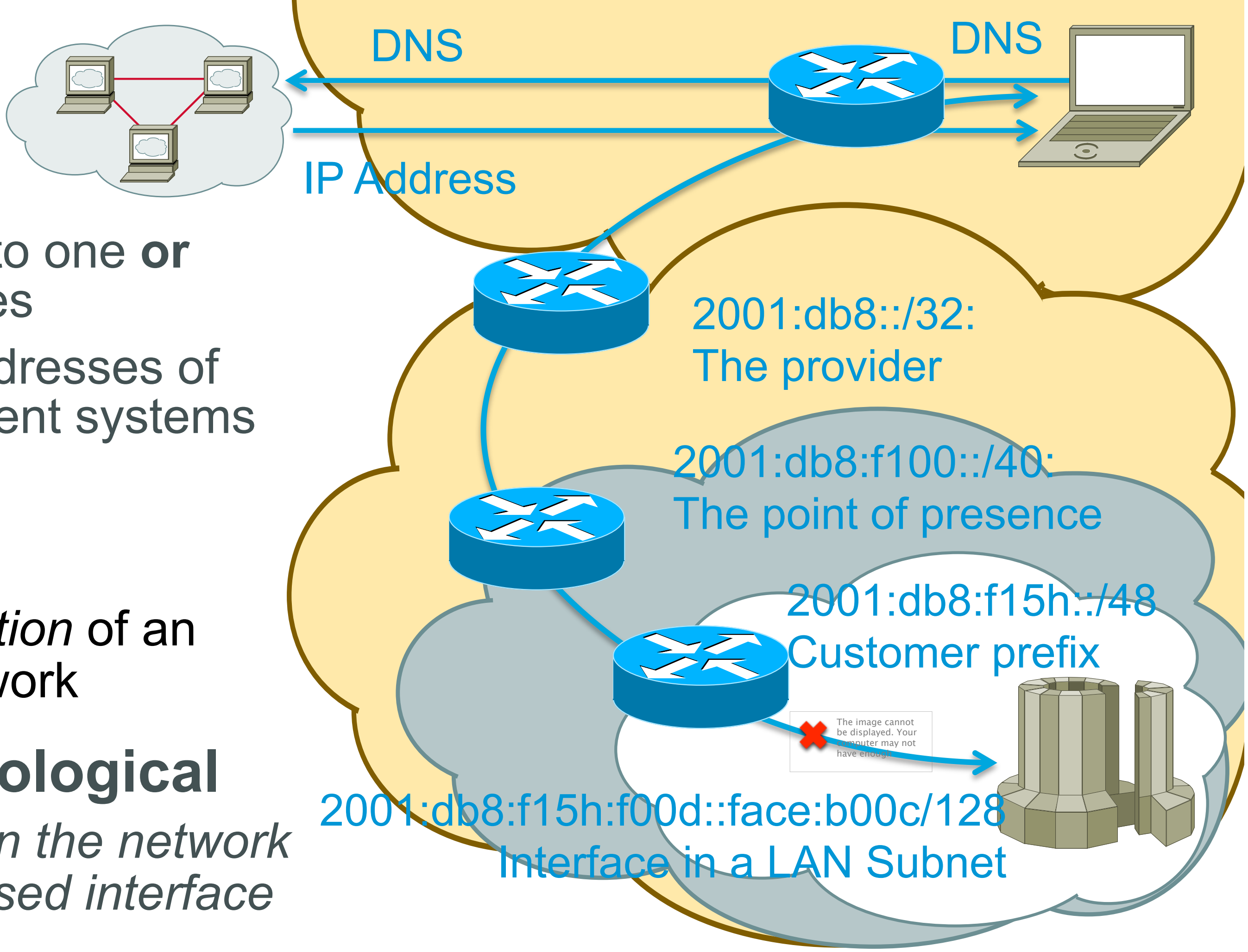
  They might be addresses of the same or different systems offering a service

- **IP Address:**

  Identifies the *location* of an *interface* in a network

- **Location is topological**

  It tells me where in the network to find the addressed interface



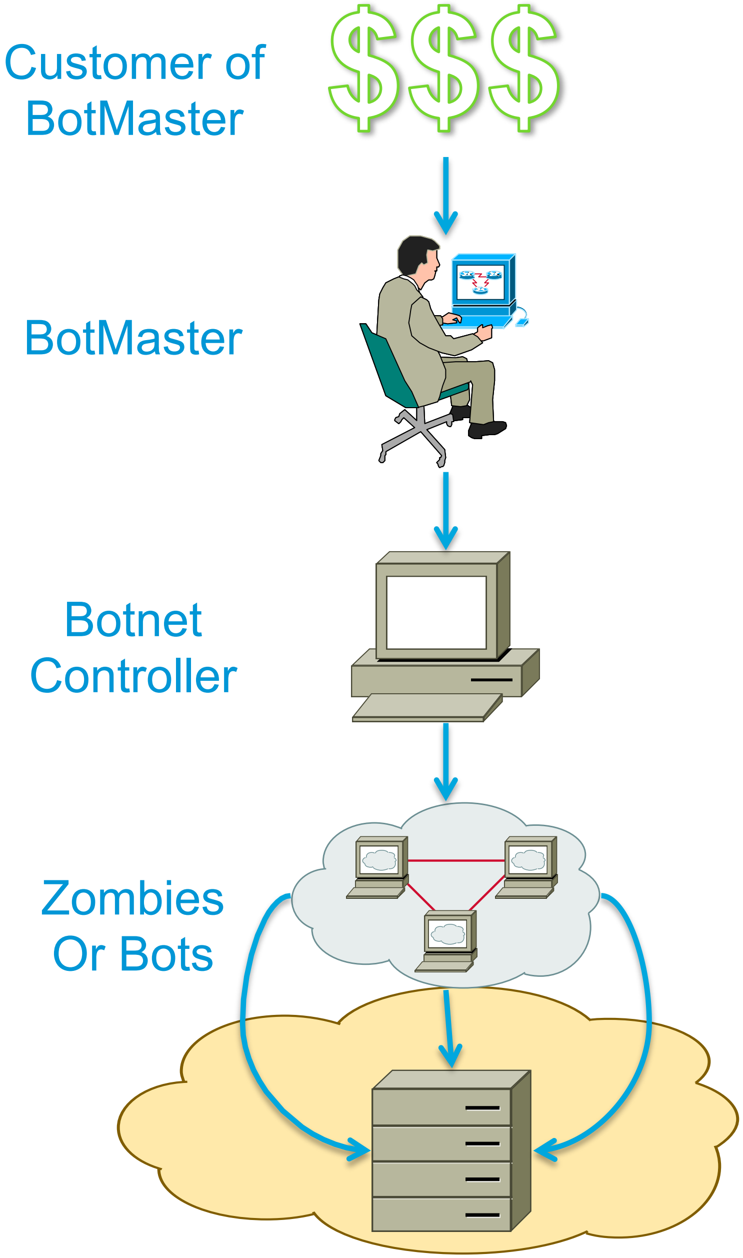
Conducting forensic studies in the Internet

What kinds of questions do forensic experts ask?

- Define “forensic”?
  - “the application of scientific methods and techniques to investigation”
- Where is this traffic coming from or going to?
  - Example: tracking down a denial of service attack
  - Example: identifying business relationships
    - “Alice often talks with Bob; Maybe I should have a contract with Bob’s provider”
  - Example: mapping a criminal network
    - “Hmm: Alice often talks with Bob…”
- Content of an exchange
  - Example: “what kinds of applications are using my network?”
  - Example: Lawful Intercept - “When Alice calls someone, what does she talk about?”
  - Example: Can I prevent communication, or collect evidence of a crime?

Attribution of attacks

1. What “zombie” system sent this message?
2. What command/control system controls the zombie?
3. What person (bot master) controls the program on that system?
4. What motivates (usually, who pays) him/her to do so?



Customer of BotMaster

$$$

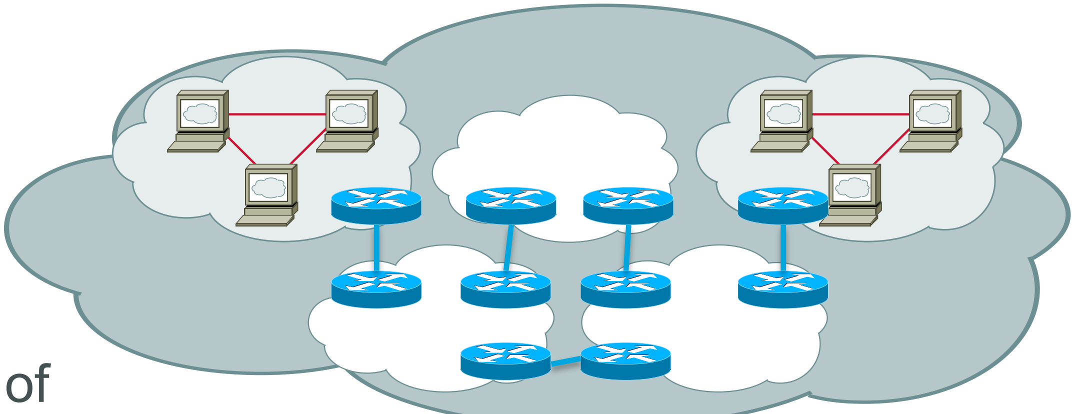
BotMaster

Botnet Controller

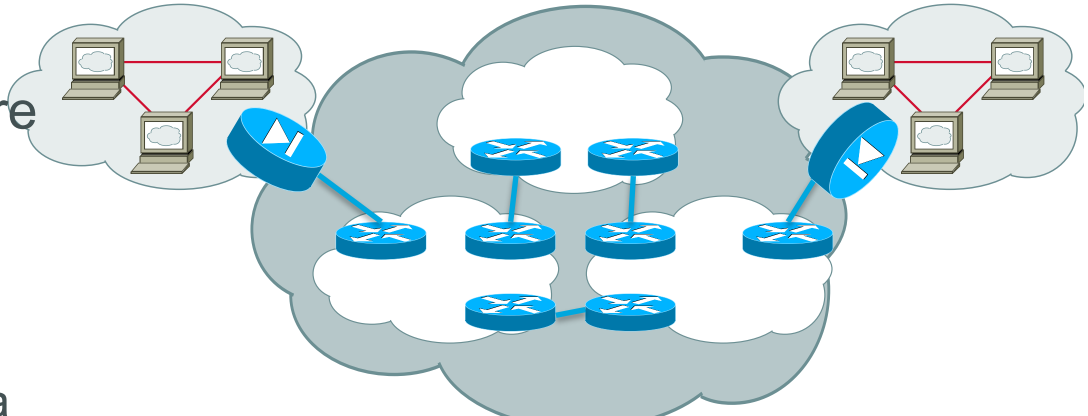
Zombies Or Bots

IPv4 complicates the problems of forensics (or, if you like, IPv6 simplifies them)

- Prefixes have less endpoint addresses in them
  - 32 bits vs. 128 bits
  - There are a lot more prefixes
- Stateful Network Address Translation forces the analysis of logs to determine what subscriber was using a given address at a given time
- Multiplexing of addresses (more than one subscriber using the same address)
  - Complicates description of a warrant, and
  - Prevents using the address for a predictable service



IPv6: edge networks and transit networks use a common address space



IPv4: edge networks and transit networks use different address space, and smaller prefixes

Data Retention

- Call Detail Records (CDR)
  - Common in telephone system; used to account minutes for billing
  - Not used in the Internet; statistics routinely kept are for packets and bytes crossing an interface in a direction
- Nearest Internet corollary: IPFIX/Netflow/Sflow
  - Record of a set of related traffic crossing an interface at a time
  - Used as a temporary diagnostic tool in troubleshooting networks
- Could be used in EU-style Data Retention (is used in Denmark)
  - Very expensive for the provider, in terms of rotating storage and electricity
- NAPT/SMTP/Web/etc. logs a little easier and less costly
  - Lower volume of data, and stored out of band
  - Already stored by service operators as diagnostic tool, but deleted quickly

The regulatory environment

Blocking content does not prevent crime

- Various ways to block content espoused in CleanFeed, HADOPI, SOPA, PIPA, OPEN, Great Shield Project, and so on
  - DNS Blocking
  - Null routing
  - Search editing
- The argument is that this is not so different from what network operators do in firewalls, and may use firewall technology
- If these tools in fact worked, there would be no
  - Cybercrime, pornography, attacks, viruses, and in some networks, peer-to-peer applications
  - **This is clearly not the case**
- In fact, in Wikileaks case, taking the content down from one place resulted in
  - It being mirrored in O(100) places and
  - Of far more interest to journalists and other evil beings.

Forcing networks to use common address space…

- Makes business harder and more expensive
- Why?
  - If done on a per-user basis, it drives up capital expenses of equipment due to larger route tables, heat, power, because routers cannot aggregate
  - It makes inter-network coordination more tedious
  - It doesn't actually fix the problems
- This is true whether it is a large block for a nation or individual addresses for citizens (and btw, tourists and business travellers need addresses too)

What does work in regulation?

- Use the same laws for cyberspace as you use in people space

  Theft is theft regardless of personal or intellectual property, etc.

- Make laws consistent among jurisdictions

  And base them on consistent, proven, legal theories

- Use digital investigation to guide and support traditional police methods

  MegaUpload, for example, was cracked primarily due to the testimony of an unindicted co-conspirator, not Lawfully Authorized Electronic Surveillance

- Cooperate with and draw on industry policy and experience

  Industry wants to be good citizens

  Industry needs a regulatory environment in which it can thrive.

“In short, business knows how to run the network and has similar problems to those of government - for which we have solutions.

Talk with us about your needs. We might be able to help.”