“Unwanted” Traffic: Roots of the Problem

Stefan Savage

Collaborative Center for Internet Epidemiology and Defenses

www.ccied.org

Department of Computer Science & Engineering
University of California, San Diego

Why do we have “bad” stuff?

- **Open communications architecture**
  - Anyone can send anything to anyone, federated management
- **Vulnerable computing platforms**
  - One bug ->millions of compromised hosts
- **Meaningful economics in exploiting computing and/or communications**
  - Affiliates progs, phishing/id theft, extortion, piracy support, etc.
- **Lack of meaningful deterrence**
  - Little forensic attribution capability, inefficient legal mechanisms

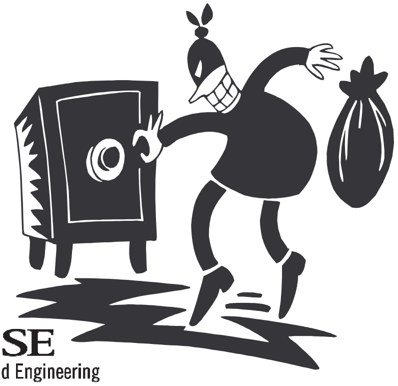
Open communications architecture

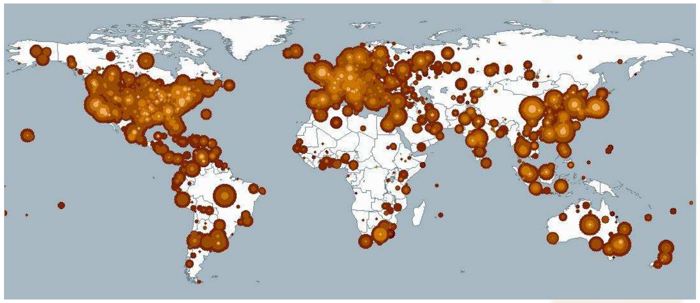
- Internet Service Model
  - Host-managed communications
  - Network poses (almost) no limits on what a user can send; addresses are virtual and easy to spoof
  - No widely deployed architecture for audit, authentication, etc
- No central administration or regulation (nor even settlement-based economic forces)
  - Feamster study shows 10% of SPAM from “tmp” prefixes
- Summary: we can depend on nothing, we’ve got little/no help from the infrastructure, and there are a huge number of independent stakeholders

Vulnerable Computing Platforms: a Threat transformation

- **Traditional threats**
  - Attacker manually targets high-value system/resource
  - Defender increases cost to compromise high-value systems
  - Biggest threat: insider attacker

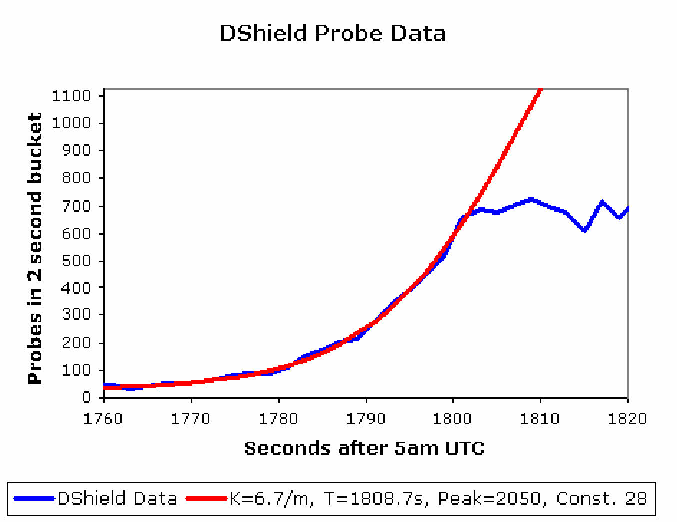
- **Modern threats**
  - Attacker uses automation to target **all** systems at once (can filter later)
  - Defender must defend **all** systems at once
  - Biggest threats: software vulnerabilities & naïve users

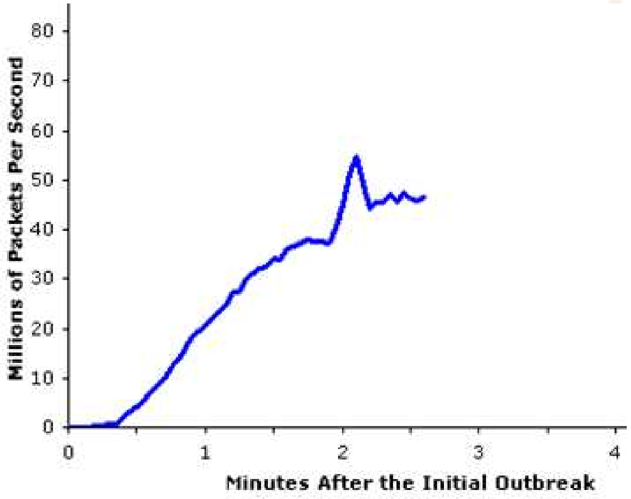




A pretty fast outbreak: Slammer (2003)

- First ~1min behaves like classic random scanning worm
  - Doubling time of ~8.5 seconds
  - CodeRed doubled every 40mins
- >1min worm starts to saturate access bandwidth
  - Some hosts issue >20,000 scans per second
  - Self-interfering (no congestion control)
- Peaks at ~3min
  - **>55million IP scans/sec**
- **90% of Internet scanned in <10mins**
  - Infected ~100k hosts (conservative)



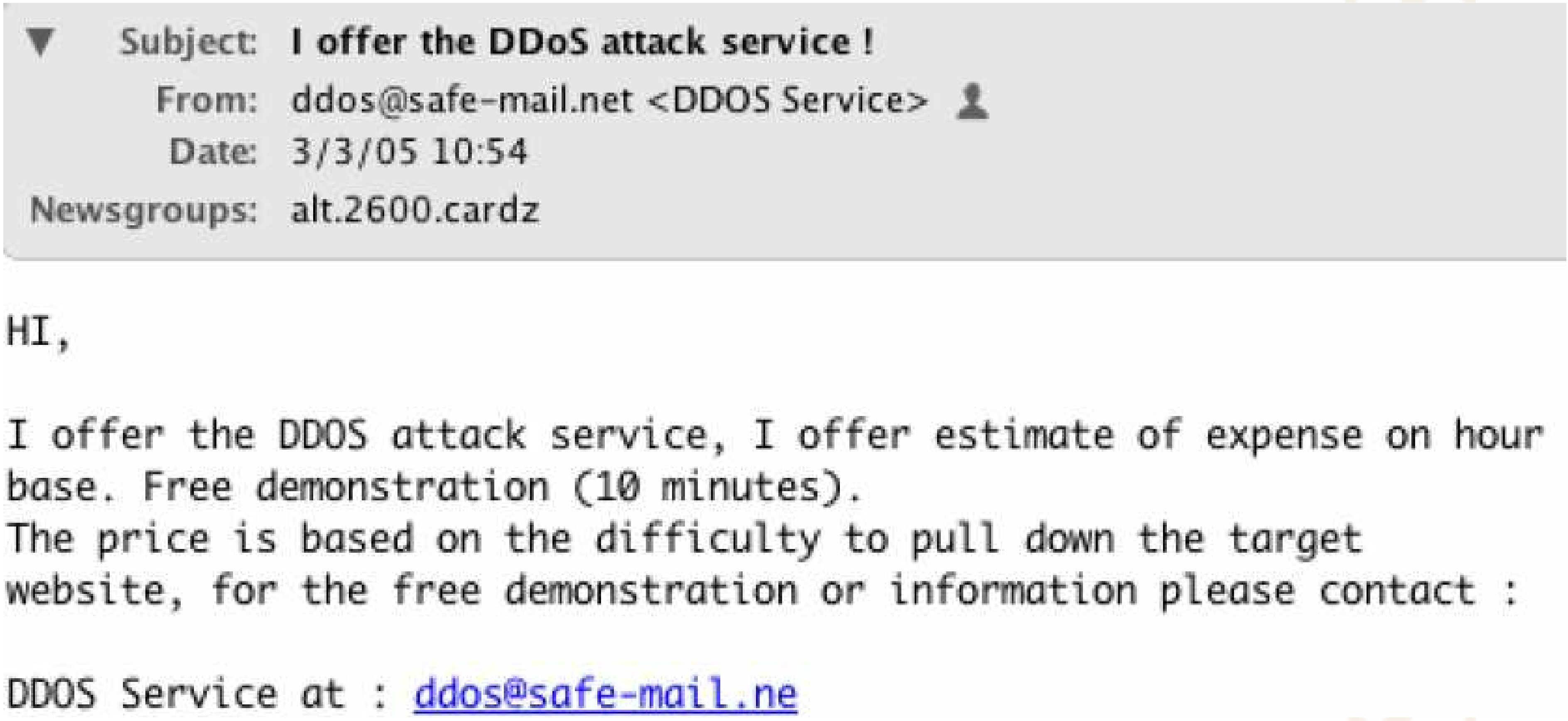


See: Moore et al, IEEE Security & Privacy, 1(4), 2003 for more details

In the last five years… an emerging criminal platform

- **Fundamental change**: profit-making criminal enterprise
  - "Virtuous" economic cycle **transforms** nature of threat
  - To wit: Botnet C&C evolution
- Commoditization of compromised hosts -> **platform**
  - Fluid third-party exchange market (**millions**)
    - SPAM proxying 3-10 cents per hostweek, Raw bot ~$1/host
  - Sophisticated command/control networks (botnets)
  - Market stratification/specialization; wholesale vs resale
  - With modern epidemic methods can take over 1M hosts in <<1min!
  - In recent measurements, one study tracked 3500 botnets (~700 active) with average size of 80,000 bots
- Innovation in criminal **applications w/viable economics**
  - DDoS extortion, SPAM, adware, piracy, phishing, identity theft

Example: what service-oriented computing really means…



Subject: **I offer the DDoS attack service !**
From: ddos@safe-mail.net <DDOS Service>
Date: 3/3/05 10:54
Newsgroups: alt.2600.cardz

```
HI,

I offer the DDOS attack service, I offer estimate of expense on hour
base. Free demonstration (10 minutes).
The price is based on the difficulty to pull down the target
website, for the free demonstration or information please contact :

DDOS Service at : ddos@safe-mail.ne
```

Botnet Spammer Rental Rates

> \>20-30k always online SOCKs4, url is de-duped and updated every
> \>10 minutes. 900/weekly, Samples will be sent on request.
> \>Monthly payments arranged at discount prices.

- 3.6 cents per bot week

> \>$350.00/weekly - $1,000/monthly (USD)
> \>Type of service: Exclusive (One slot only)
> \>Always Online: 5,000 - 6,000
> \>Updated every: 10 minutes

- 6 cents per bot week

> \>$220.00/weekly - $800.00/monthly (USD)
> \>Type of service: Shared (4 slots)
> \>Always Online: 9,000 - 10,000
> \>Updated every: 5 minutes

- 2.5 cents per bot week

September 2004 postings to SpecialHam.com, Spamforum.biz

How affiliates programs work…

- “Our first program pays you $0.50 for every validated free-trial registrant your website sends to [bleep]. Commissions are quick and easy because we pay you when people sign up for our three-day free-trial. Since [bleep] doesn't require a credit card number or outside verification service to use the free trial, generating revenue is a snap.

  The second program we offer is our pay per sign-up plan. This program allows you to earn a percentage on every converted (paying) member who joins [bleep]. You could make up to 60% of each membership fee from people you direct to join the site.

  Lastly, [bleep] offers a two tier program in addition to our other plans. If you successfully refer another webmaster to our site and they open an affiliate account, you begin earning money from their traffic as well! The second tier pays $0.02 per free-trial registrant or up to 3% of their sign-ups.”

What's next:
Large-scale Information Piracy

- The true value in a compromised host is the ***information*** it holds
  - Spreadsheets, e-mail, presentations, etc

- It is **easy** to implement distributed queries:
  - Find all spreadsheets containing "10-Q" & "2007"
  - Find all e-mail containing "From:" and "microsoft.com"

- *Cast a wide net and reel in…*
  - If ex-CIA director John Deutch connected his laptop to AOL, why do you think your organization is better?

The final piece of the economic puzzle

- How to turn virtual goods (e.g. CC#, paypal) into real assets?
- Back account info:
  - Wire $$$ to bank account under assumed name in Estonia, withdraw $$$ and close account
- Credit card/pay pal info
  - Use MC/paypal to order items and send to remailer
  - Use paypal $$$ to pay third party physical remailer
    - Accepts packages, rewraps and sends abroad
  - Foreign address is temporary PO box under assumed name (or generic c/o building address)
- CDUniverse story
  - Resell items on ebay at discounted prices
  - Cashiers check to foreign bank…
  - “Safe” because they only ask you to send **after** delivery
  - Not so safe… uncancellable

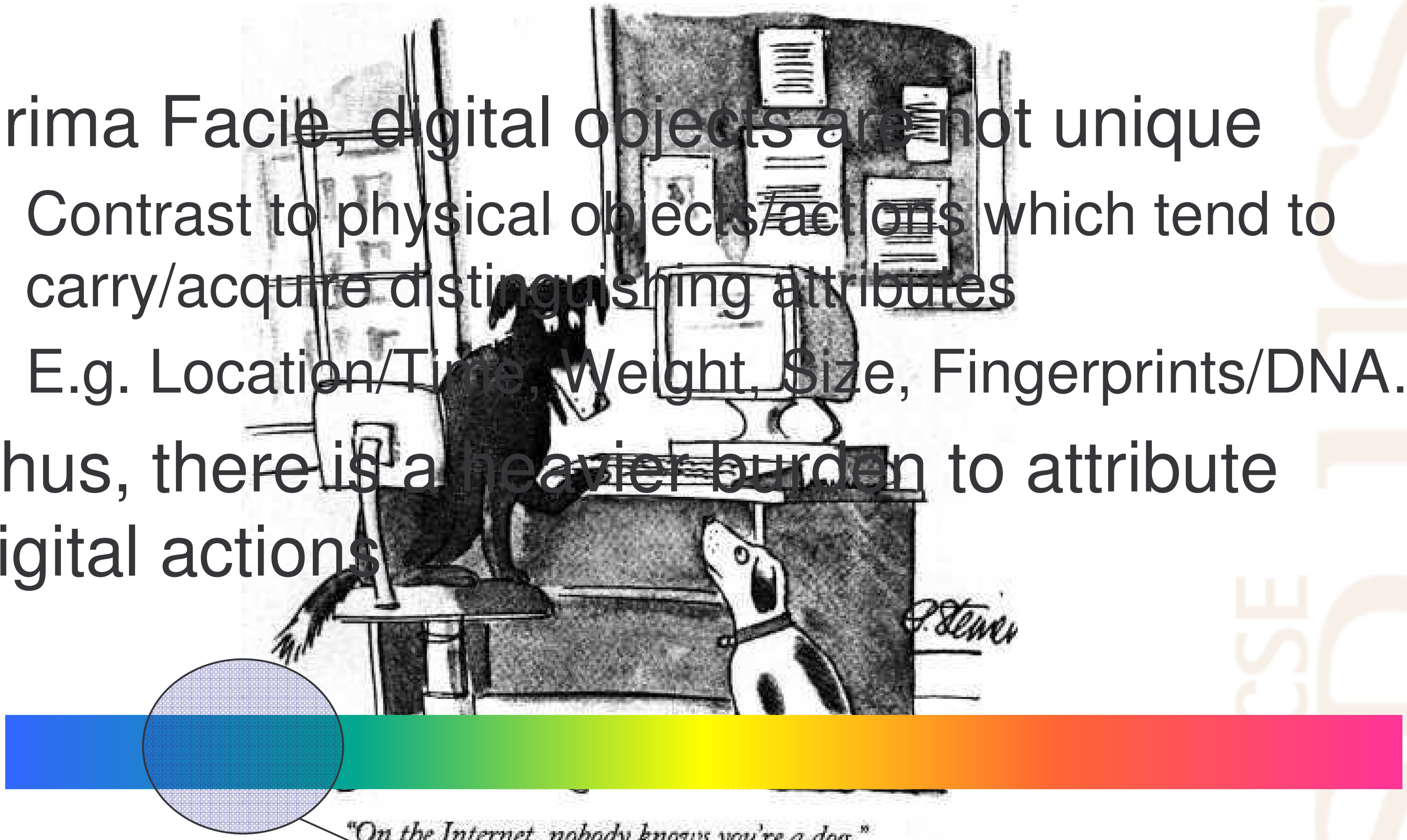
Deterrence

- Its very hard to attribute an attack to an individual
- Its very expensive
- It rarely gets done

- Bottom line: lack of cheap, effective, precise attribution capability

Attribution: Essence of the problem

- Prima Facie, digital objects are not unique
  - Contrast to physical objects/actions which tend to carry/acquire distinguishing attributes
  - E.g. Location/Time, Weight, Size, Fingerprints/DNA…
- Thus, there is a heavier burden to attribute digital actions



"On the Internet, nobody knows you're a dog."

Complete Anonymity

We are here

Or a thief, or a terrorist, or a predator, or a extortionist, etc…

Perfect Attribution

Kinds of attribution

- *Authentication*: who wants to do that?
  - Access control, non-repudiation
  - Technology exists, but too painful
- *Situational awareness*: who is doing that now?
  - Operational response (e.g. against DDoS, BotNet C&C)
  - Technology exists at coarse granularity
- ***Forensics*: who did that?**
  - Investigatory, evidentiary
  - Things are pretty terrible here

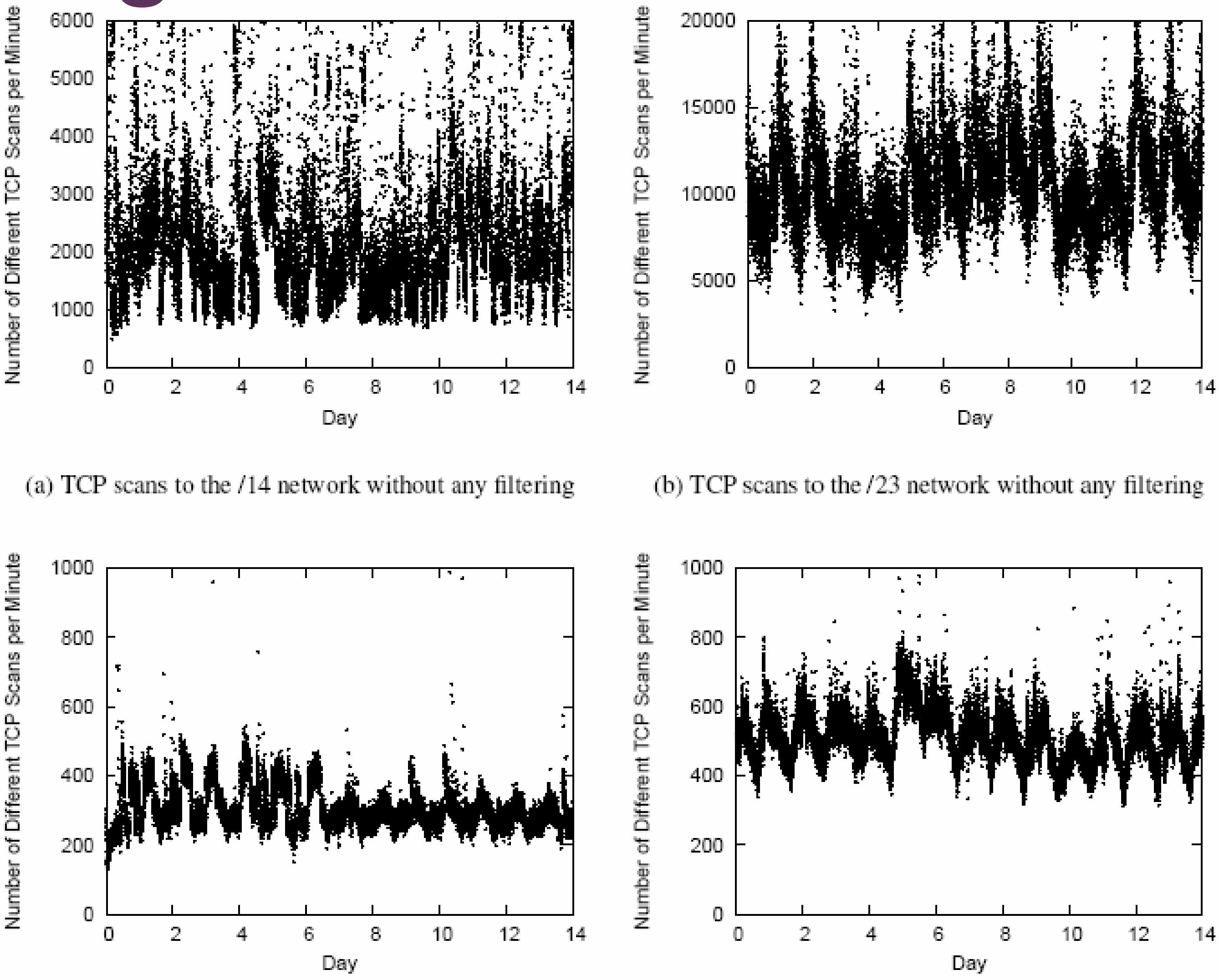
Why "traceback" doesn't help

- Best-case: find machine sending packets
  - Generally enough for operational purposes
- Stepping stones (identity laundering)
  - Attacker logs into A, from A logs into B, from B logs into C, mounts attack from C (esp bad for P2P C&C)
  - Can infer linkage via correlated timing, loss, etc but tricky
    - Must be done in real-time
    - Stepping stone machine can do *anything* to packets
    - Can lose trail if one party doesn't help
- Multi-user machines or broadcast networks
- First machine isn't necessarily attacker location (let alone identity)
  - Open wireless, unaudited dialup, physical proxies
  - Compromised machine (my machine was hacked defense)

Questions

?

Background Radiation



(a) TCP scans to the /14 network without any filtering

(b) TCP scans to the /23 network without any filtering

(c) TCP scans to the /14 network after scan filtering

(d) TCP scans to the /23 network after scan filtering