When One Has Lemons …

Martin Casado *(Stanford)*
Tal Garfinkel *(Stanford)*
Weidong Cui *(Berkeley)*
Vern Paxson *(ICSI)*
Stefan Savage *(UCSD)*

What this Talk is About ..

Spurious Traffic

(generally unwanted traffic on the Internet)

+

Opportunistic ***(parasitic)*** **Measurement**

(exploiting existing traffic for unrelated measurement)

=

Useful Internet Measurement(?)

Limits Traditional Measurement

- Much of Internet is “hidden” from view
  - NAT
  - Proxies
  - Firewalls
- Relatively Few Sources

  (planetlab has 620)
- Limited diversity

  (typically, academic networks, near the core)
- Limited to socially acceptable traffic patterns
- Privacy issues with passive measurement

However …
Lots of "Unwanted"Traffic on the Internet

- Malware,misconfigurations, …
- Many Sources
  - 359,000 for Crv2
  - 16,000 large automated scan
- Great Diversity
  - 159 countries from 38,000 spam sources
  - often from poorly administered machines
  - residential bias
- Extreme Traffic Patterns
  - Internet scale network events
  - antisocial

Opportunistic Measurement

- Use existing traffic for unrelated measurement purposes
- Collect using network telescopes
- Infer network properties
  - About the sender
  - About the Internet
  - About the collection site
  - …

Talk Outline

- Spurious Traffic Classes
- Example Opportunistic Measurement Studies
- Getting the most from your lemons
- Limitations
- Conclusions

Spurious Traffic ...

- Generally “unwanted”
- Worms
- Spam
- Automated scans
- Misconfigurations
- Static settings, defaults
- ...

Useful Traffic Generators

- Worms
  - Large number of sources
  - Code is readily available
  - Known scanning patterns
  - Initial outbreaks create “stress” conditions
- Automated Scans
  - large number of sources
  - can generate large, predictable bursts
  - often generate more traffic than worms (want to find more information)

Example Event

Eurasian Scan

- From China, Japan, Germany and Korea
- Have identified 16,000 unique hosts
- Visible each day at Stanford, CAIDA, LBNL
- Regular, within 5 minutes
- **LOUD** and intrusive
- SYN packets to 9898(Dabber), 1023(Sasser), 5554(Sasser)

Useful Traffic Generators

- Spam
  - Source of “significant” TCP flows
  - Many sources
  - Often sent from compromised end-hosts
    (doesn't require scanning worm to have infected)
  - Flows can be “directed” to different collection locations
- Network (Mis)Configurations
  - NetGear static NTP configuration
  - Preconfigured source address for DDoS tool

Example
NAT Study Using Code Red II

- How many sources behind NAT?

- Released in 2001
- Infects vulnerable IIS servers
- Still active source of traffic
- HTTP connection scans
- Preferential scanning
  - ½ of scans to local /8
  - 3/8 of scans to local /16
  - 1/8 of scans to full Internet

Code Red II

expect roughly 2^{10} more packets here

6 /24s in 192/8

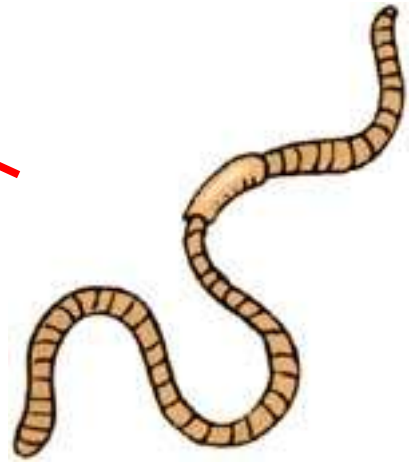
4 /16s not shared w/ private (used as baseline)

one /16 shared with 169 /8

192/8

192.168/16

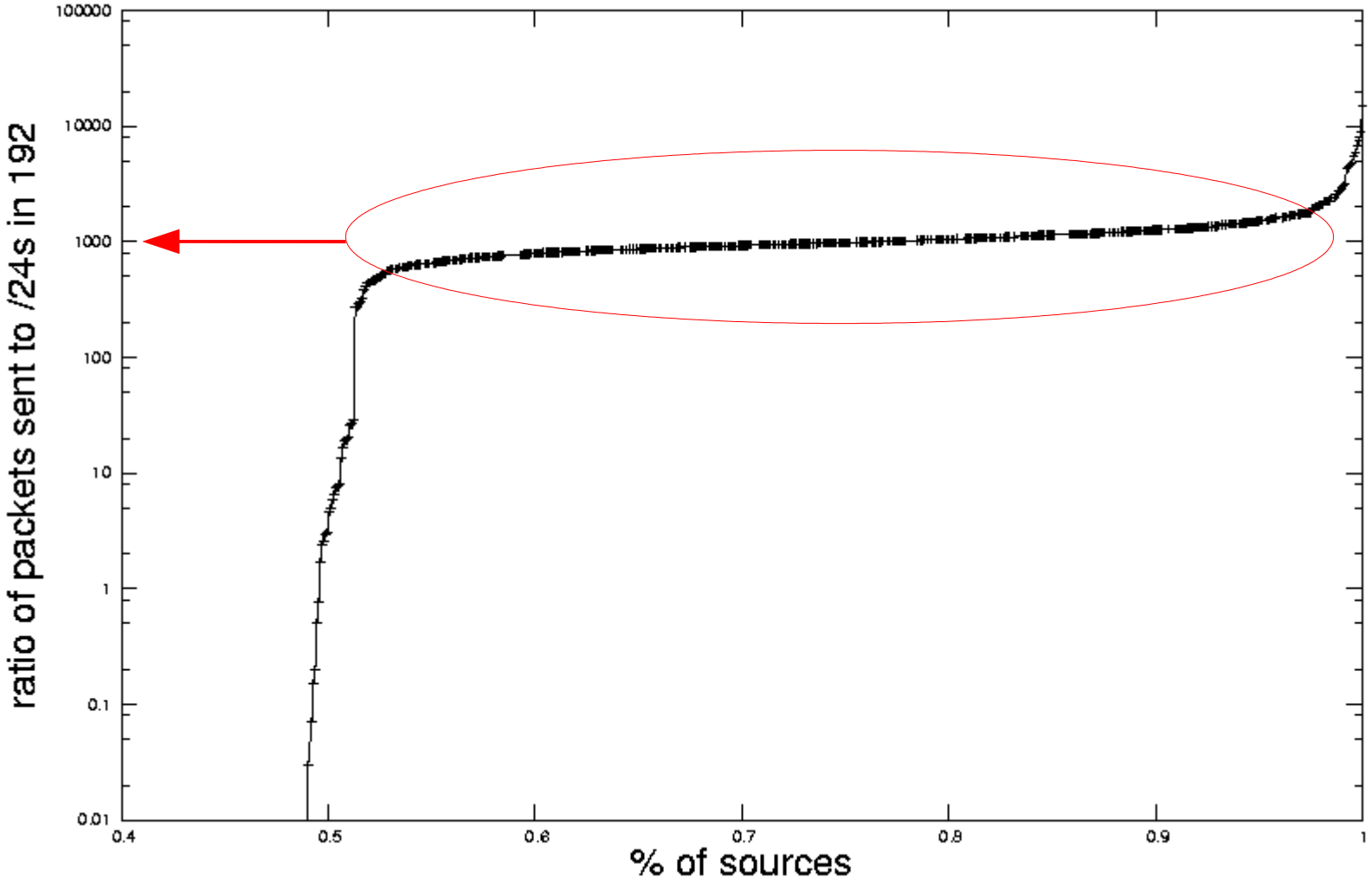
Internet



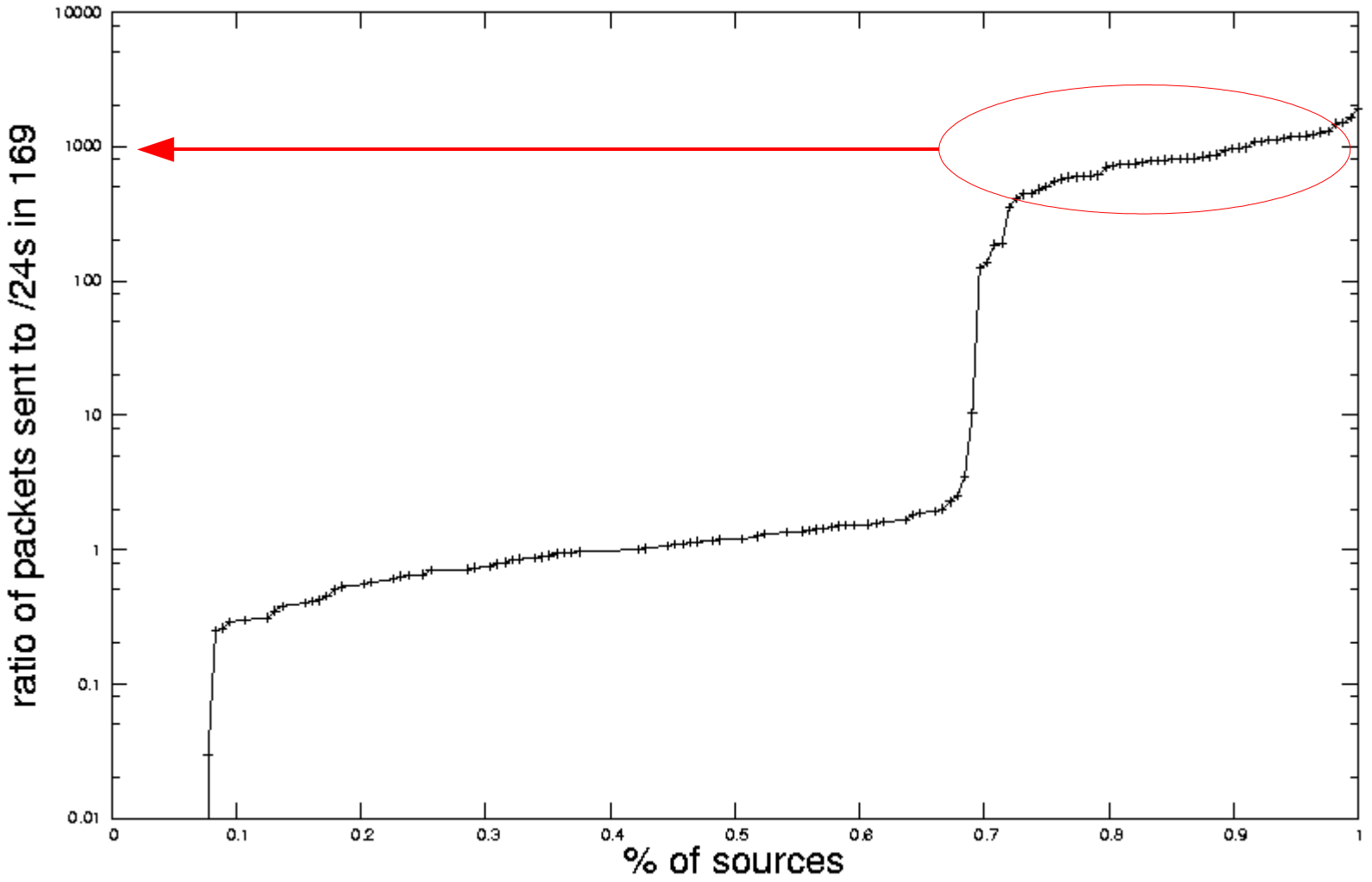
192.168.1.23

Used traces from 48-hour period
487,291 scans from 1,528 sources

Ratio of Packets Sent to 192



Ratio of Packets Sent to 169



ICSI

UCSD

Comments on Results

- 65% of machines appear to be behind NAT ..
  however
- NAT reduces chance of infection
- IIS less likely to be behind NAT
- Other private prefixes (e.g. 10/8 and 172.16)
- CRII is old … biased demographic

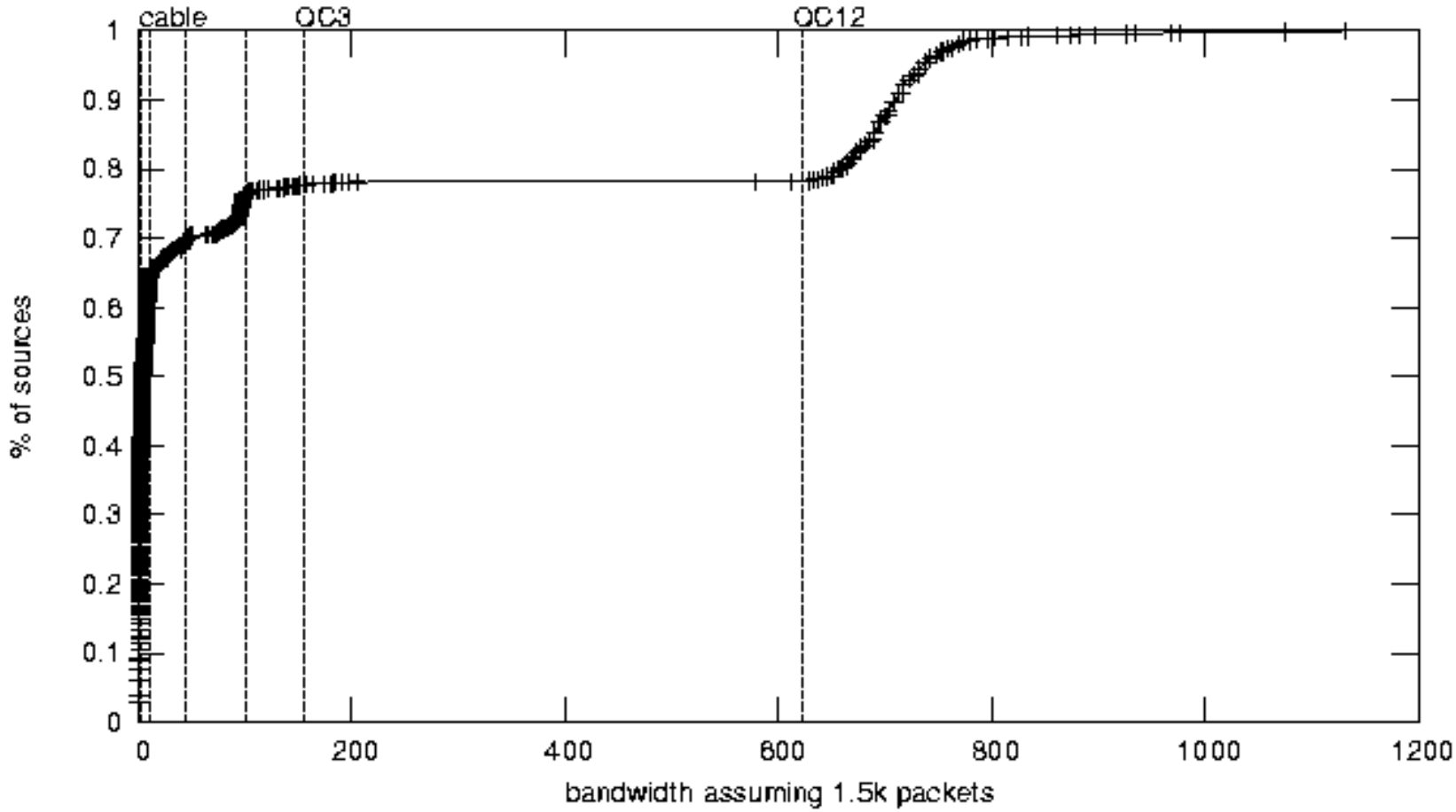
Example

A Heavily Spammed Domain

- Long online life
- Up to 1 million emails daily
- Over 40,000 source addresses
- Can produce “significant” tcp flows
  - MultiQ tool from M&M tool suit *(Katti, et. al 2004)*
  - 24,698 “significant” flows
  - 2,269 sources
  - 70% at cable speeds or below
  - 15-20% at OC12 speeds

UCSD

Bottleneck Link Bandwidth From Spam Sources



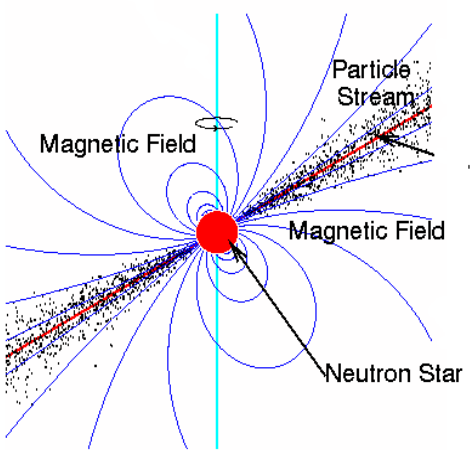
ICSI

UCSD

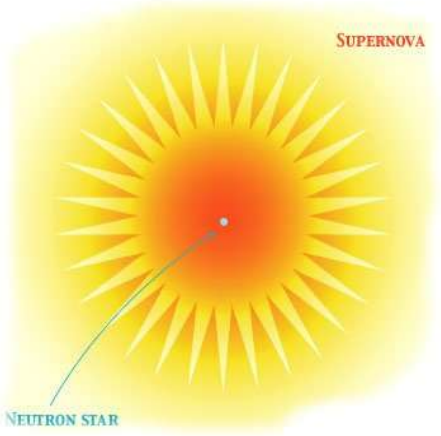
Getting the Most Juice out of your Lemons

General Classification

- Pulsars
  - Reliably send traffic over period of time
  - Endemic worms
  - Identified scans



- Super Novas
  - Flash events
  - Worm outbreaks
  - DDoS attacks
  - Large scans



Other Properties

- Queuing delay via packet-pair variance
- Bottleneck link bandwidth
- Drop/filtering rates
- TTL studies ("depth of the Internet")
- *Super Nova* can reveal capacity limits
- End host characteristics (e.g Witty Worm *Kumar et al, 2005)*
  - Number of disks
  - Link speeds
  - Uptimes
  - Infection tree
  - Patient zero
  - Target was military base

Calibrating Collection Sites

- Use *pulsars* with known distributions
- … or use movable target
- Determine telescope biases
  - filtering rules
  - drop rates
  - rate limits

Growing Lemons:
Attractors and Agitators

- Natural Attractors (hot spots)
  - Attract some or all of the traffic from a source
  - Examples
    - default configurations
    - scanning biases based on local configuration
- Agitators
  - Cause a source to send more traffic (e.g. honeynet responder)
- Chumming
  - Steps to draw attackers (e.g. Irc server)

Summary

- Traditional measurement is limited by
  - Sources
  - “Opacity” of the edge
  - Ability to send disruptive traffic
- Spurious traffic has complimentary properties
  - Many sources
  - Great diversity
  - Extreme traffic conditions

However …

- Noisy
- Lack of knowledge/control of sending environment
- Generally biased source sets
- Limited by available traffic

Yet …

- Preliminary studies provide promising results
- Can aid traditional measurement studies
- Much work remains to be done to develop and refine techniques

Questions?