signed archives:
an evaluation of internet trust

Jose Nazario

jose@crimelabs.net December, 2002

Overview

- Overview of Research
- Motivation
- Methods
- Results
- Discussion
- Related Work
- Future Directions

Research Overview

- "Download, verify"

- Identify signed archives on the net
- Download, grab the corresponding key
- Verify, investigate errors
- Meta-analysis on results and data set

Motivation

- 2002: Series of high profile trojans
  - irssi, dsniff, fragroute, fragrouter, openssh, sendmail, tcpdump
- How many others are compromised?

- About 10 years of PGP
- How well is the web of trust doing?

Methods

- Identify signed archives
  - used Google, generic search terms
  - weeded list down to 166 unique servers, 2804 archives
- Build a tool to autocheck
  - based on ’extract-0.1’
- Download archives
- Check
- Post-process the data

About ’extract-0.1’

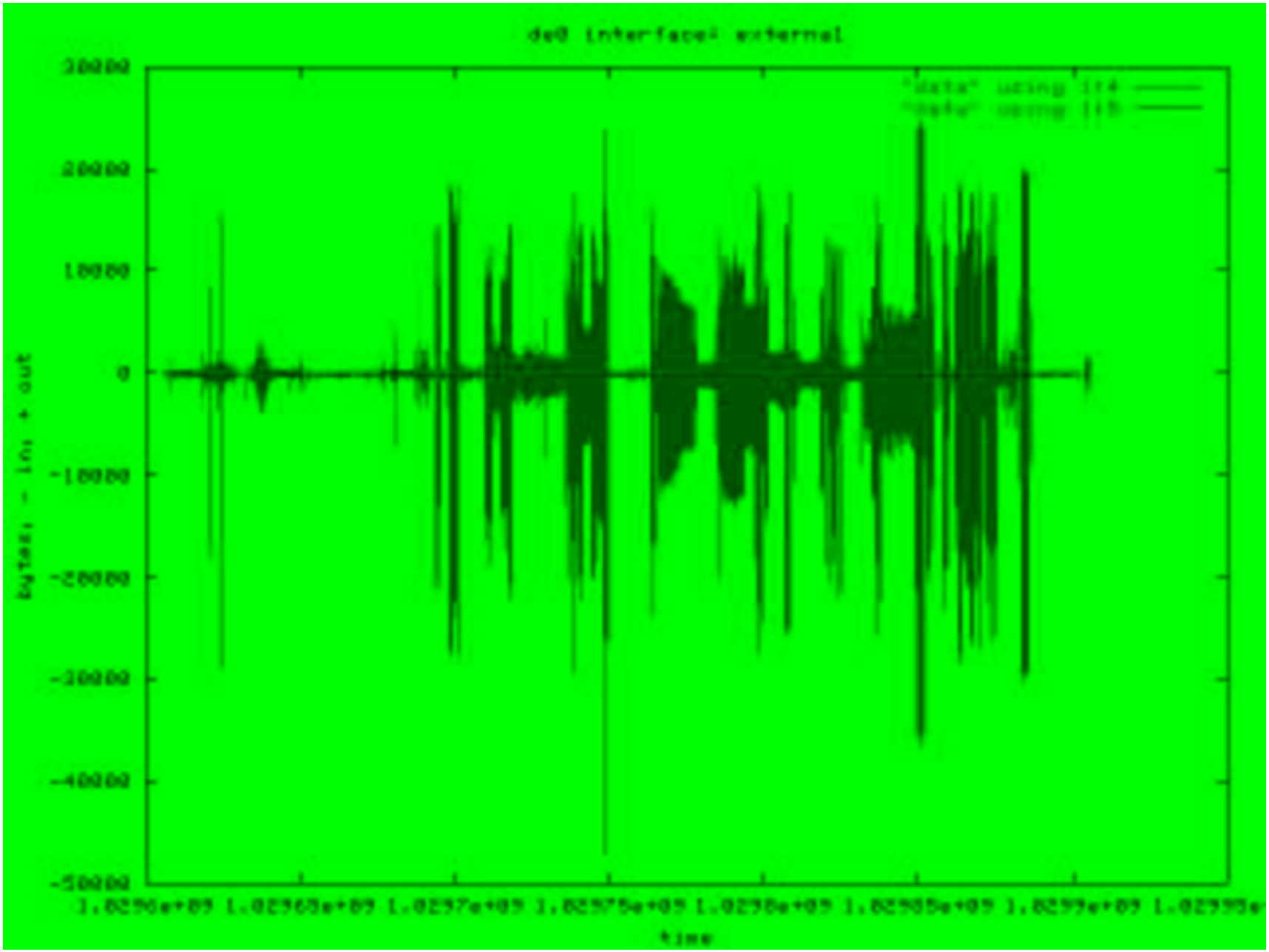
- Small shell script wrapper to tar and gpg
- Looks for detached signature for archive
- Checks the key
  - If the key exists, continue
  - If we dont have the key, fetch it, restart
- Check the signature with the key
- If it checks out ok, extract

- Small, efficient, easy to use

Downloading the Archives

- Took about 3 days (operated only serially)
- Nearly maxed out my cable modem
- Took up about 9 GB of storage

Traffic Impact



Bulk Analyzing

- Empty GPG keyring
- Used a modified ‘extract-0.1’ tool
- Small shell script:
  - find all archives in current directory
  - run modified ’extract’ on them
- Took a few hours on data machine (K6-2/300, 64MB)
- Log all actions, post process

GOOD SIG for 12.32.54.90/pub/gnuplot/gnuplot3.7cyg.zip in /home/jose/crimelabs/projects/signed-archives/sampling

signature FAILED for ftp.tlk-l.net/pub/mirrors/ssh.com/ssh-1.2.33.tar.gz

Results

- Number of archives checked: 2804
- Number of unique archives: 1426
- Number of sites downloaded from: 166
- Number of keys fetched: 93

- Success: 2799 archives
- Failures: 5

About the Failures

- Failure 1: Truncated download (OpenSSH, mirrored elsewhere)
- Failures 2 and 3: False negatives
  - unknown why, but repeating the analysis came up positive
  - bulk repeated showed random failures, but no random positives
- Failures 4 and 5: Legitimate
  - updated archive, no updated sig
  - author confirmed (V. A. Brennan)

Complete Failures

- Unable to verify ‘cmu-snmp’: old key, incompatable

Never contacted authors, but shows a breakdown of the system. No key ever found.

Uncovered Weaknesses in the System

- Inline key distribution
- Key compromise risk
- Few signatures
- Trust of the signatures and keys

Inline Key Distribution

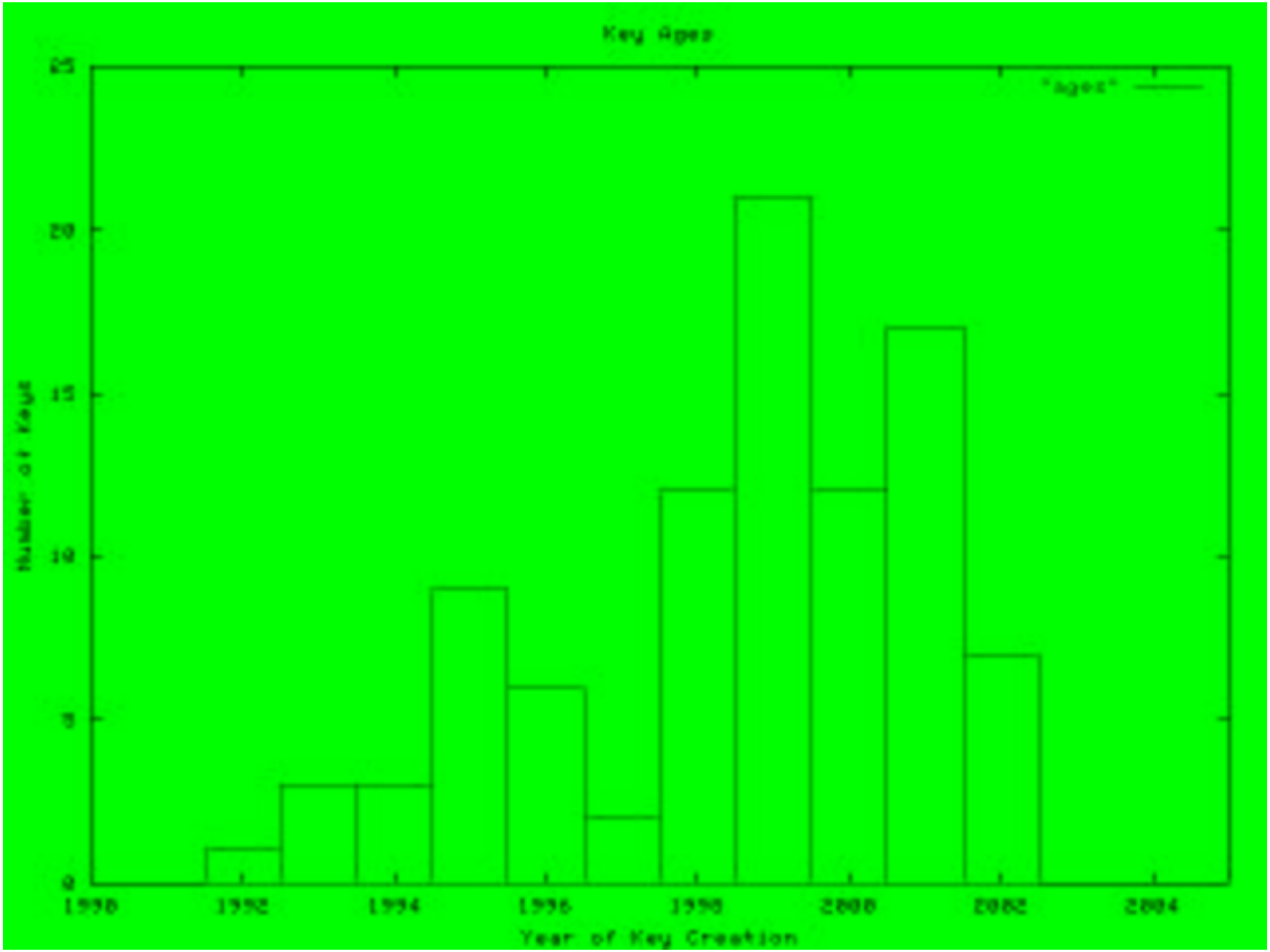
- Means placing the key along side the archive (on the server)
- Temptation for client to grab key there
- Attack: compromise binary, upload fake key which verifies
- Notable abusers: OpenSSH portable, SSH Communications, Cyrus, Gnuplot

Key Compromise and Risk

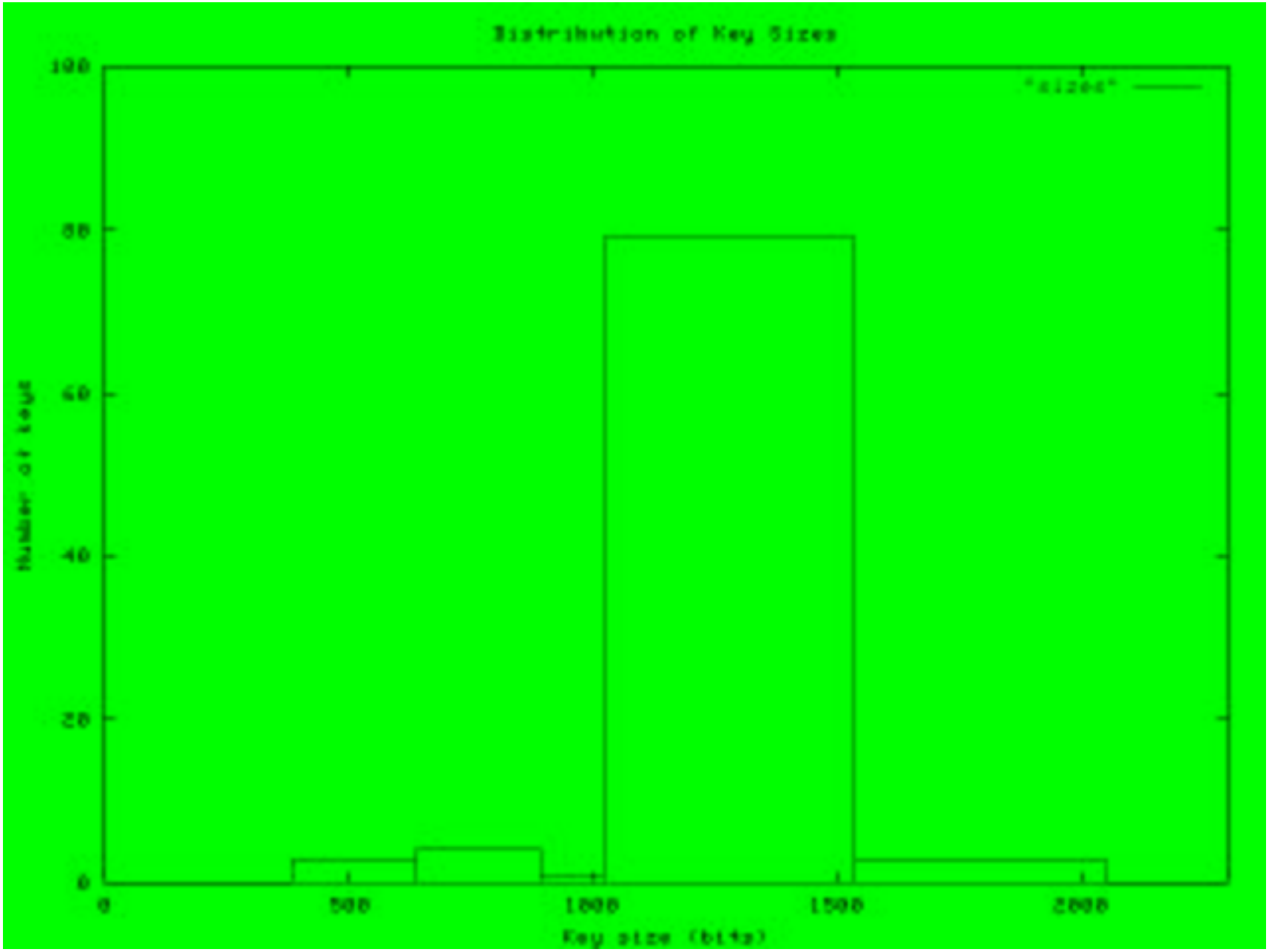
- Related to key age, size, interest level of an attacker

- 93 keys analyzed
- Most 3 or fewer years old
- Some 10 years old
- Most 1024 bit
- Some 512 bit
- 1024 bit keys persist through 10 year period

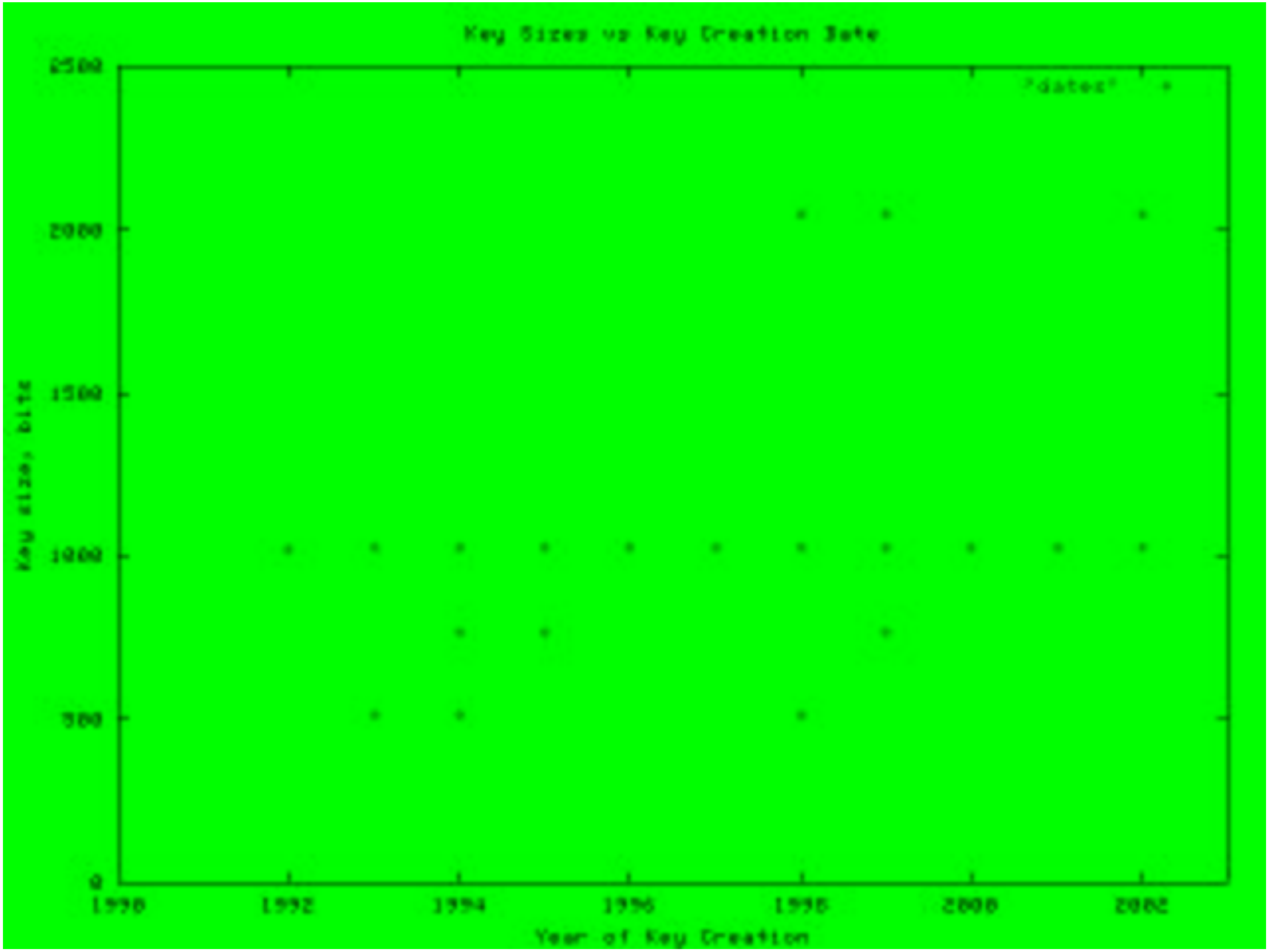
Key Ages


Key Ages
"Ages"
Number of Keys
0
5
10
15
20
25
1990
1992
1994
1996
1998
2000
2002
2004
Year of Key Creation

Key Sizes in Bits



Correlating Key Age and Size

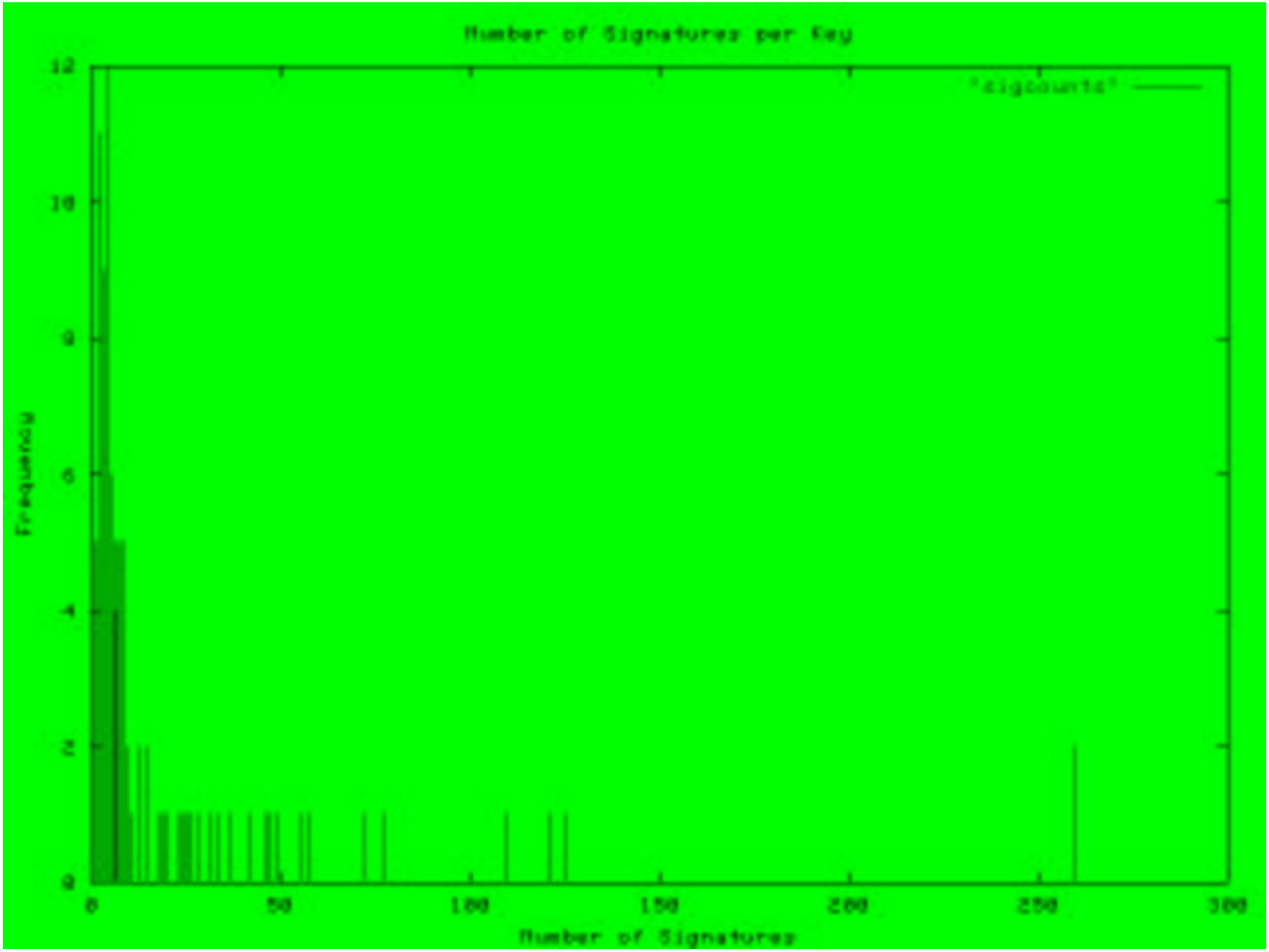


Few Signatures in Use

- Signatures help establish trust
- Web of trust model
- Fewer signatures means its harder to verify

- Average of 21 signatures per key
  - some had none
  - two have 261 signatures

Signatures per Key



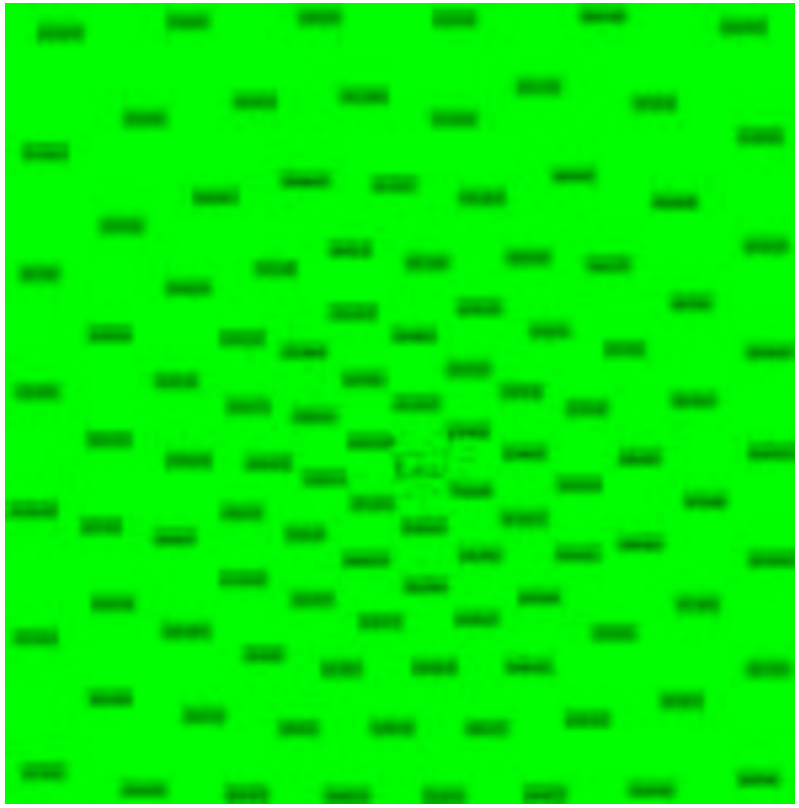
Verifying the Signatures

- Examine a key, trace signatures
- All sigs should go back to the "strong set"
- Strong set is 100,000+ keys which are self contained

- 36 could not be mapped back
- 57 mapped to strong set
  - average of 6 hops to center

- 2/3 keys check out OK

Tying Signatures to the Strong Set



http://monkey.org/~jose/graphing/csw03/csw03.png

Related Work

- Several studies on web of trust model
- Dugsongs ’gzsig’ tool
- Marius’ PGPwrap library
- GPGme library
  - both can be used to build ’extract’ like tools

- BSD ports tree
  - distributed set of MD5, SHA-1, and RMD160 hashes
  - what actually caught most of the trojans in 2002

Future Directions

- Currently seeking research partner for ongoing study
  - need bandwidth, storage
- Bring everyone into the strong set
- Get more signed archives out there
- Design a better system

Acknowledgements

- Beth (letting me throttle our cable modem)
- Marius, Dug, Niels, Alex, Seth for conversation
- Dtype.org people for strong set analysis

- UMeet organizers

- You