



Designing a Framework for Active Worm Detection on Global Networks

Vincent Berk
George Bakos
Robert Morris



Outline

- Worms and Internet Epidemics
 - Internet Reachability
 - Code Red v2
- ICMP Destination Unreachables
- DIB:S
- Simulation Results
- Future work and Concerns



Worms

- What is a “Worm”
 - Email Virus?
 - Code Red, Sapphire?
- Definition:

“An autonomous, self-propagating piece of code that enters networked computer systems uninvited.”
- Payload



Internet Epidemics: Population

- Total population: N
- Susceptibles: $s(t)$
- Infectives: $i(t)$
- Recovered/Removed: $r(t)$

$$N = s(t) + i(t) + r(t)$$



Internet Epidemics: Equations

- Classic epidemiology
(Kermack&McKendrick, 1927):

$$\frac{ds}{dt} = -\beta si$$

$$\frac{di}{dt} = \beta si - \gamma i$$

$$\frac{dr}{dt} = \gamma i$$



Internet Epidemics: Parameters

- Beta: Infection parameter

$$\beta = \frac{C}{N} \times \frac{\alpha}{\tau}$$

- Where:
 - C target selection algorithm, 1 for random
 - N size of the address space
 - α number of unique concurrent scans
 - τ average time to determine reachability



Internet Epidemics: Parameters

- Tau: average time to determine reachability

$$\tau = r \times t_{latency} + (1 - r) \times t_{timeout}$$

- Where:
 - r the average reachability
 - $latency$ average connection latency
 - $timeout$ average time-out, usually depends on OS



Internet Epidemics: Parameters

- Gamma:
 - Removal Parameter
 - Mainly based on Human Response
 - Negligible if worm propagates very fast
 - Not a constant



Internet Reachability

	January 2002		March 2003	
Requests sent	2111469	100%	9288348	100%
No response	1937388	91.7%	6964182	75.0%
Acknowledgements	12038	0.6%	918287	9.9%
Destination Unreachable	145155	6.9%	571857	6.2%
Other, RST, TTL-exceeded	16888	0.8%	827655	8.9%



Code Red v2

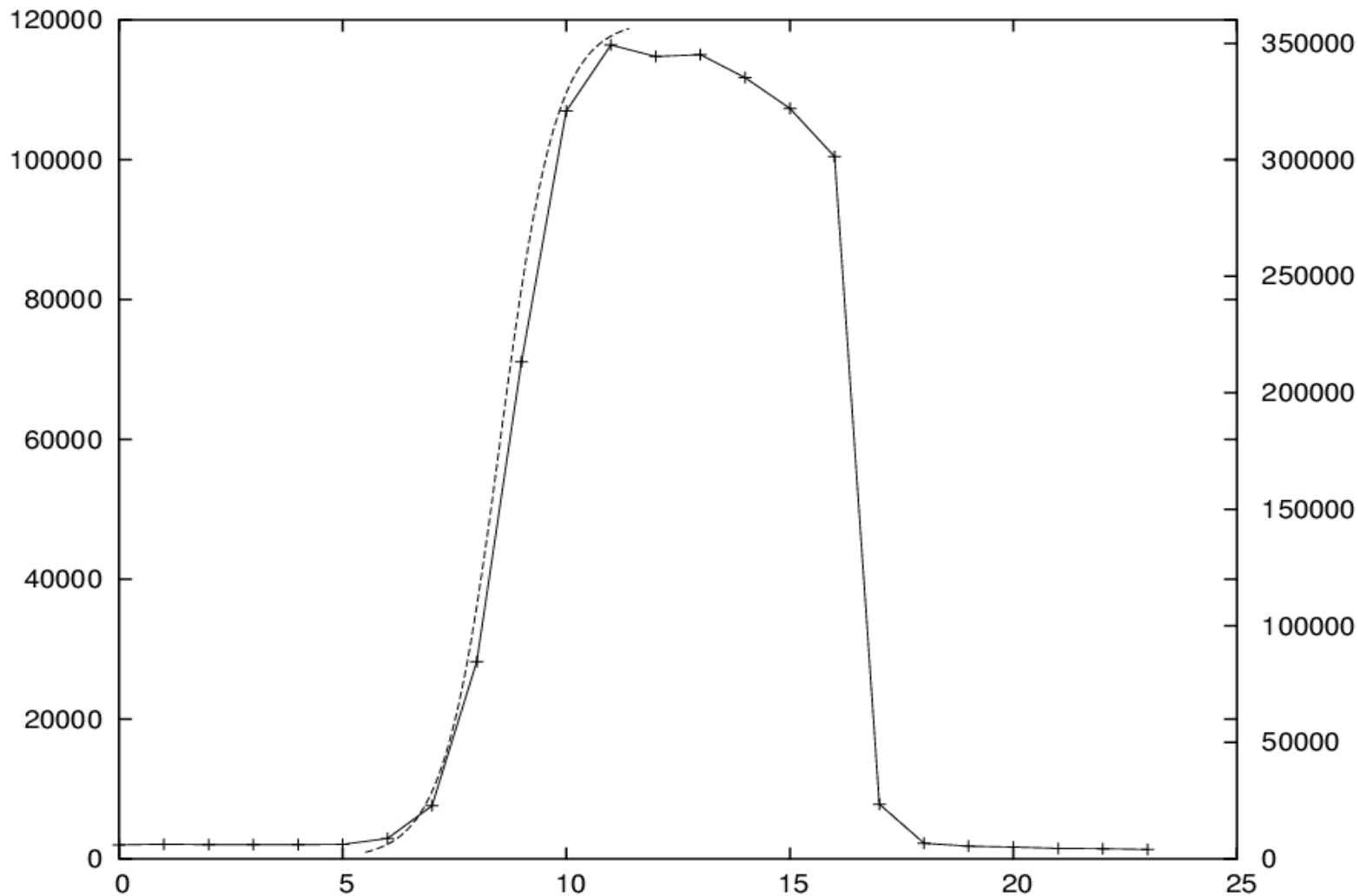
- Spread on July 19th 2001
- Infected nearly 360.000 hosts
- in 14 hours

$$\tau = \frac{1}{10} \times 1 + \left(1 - \frac{1}{10}\right) \times 21 = 19$$

$$\beta = \frac{1}{2^{32}} \times \frac{100}{19} = 1.23 \times 10^{-9}$$



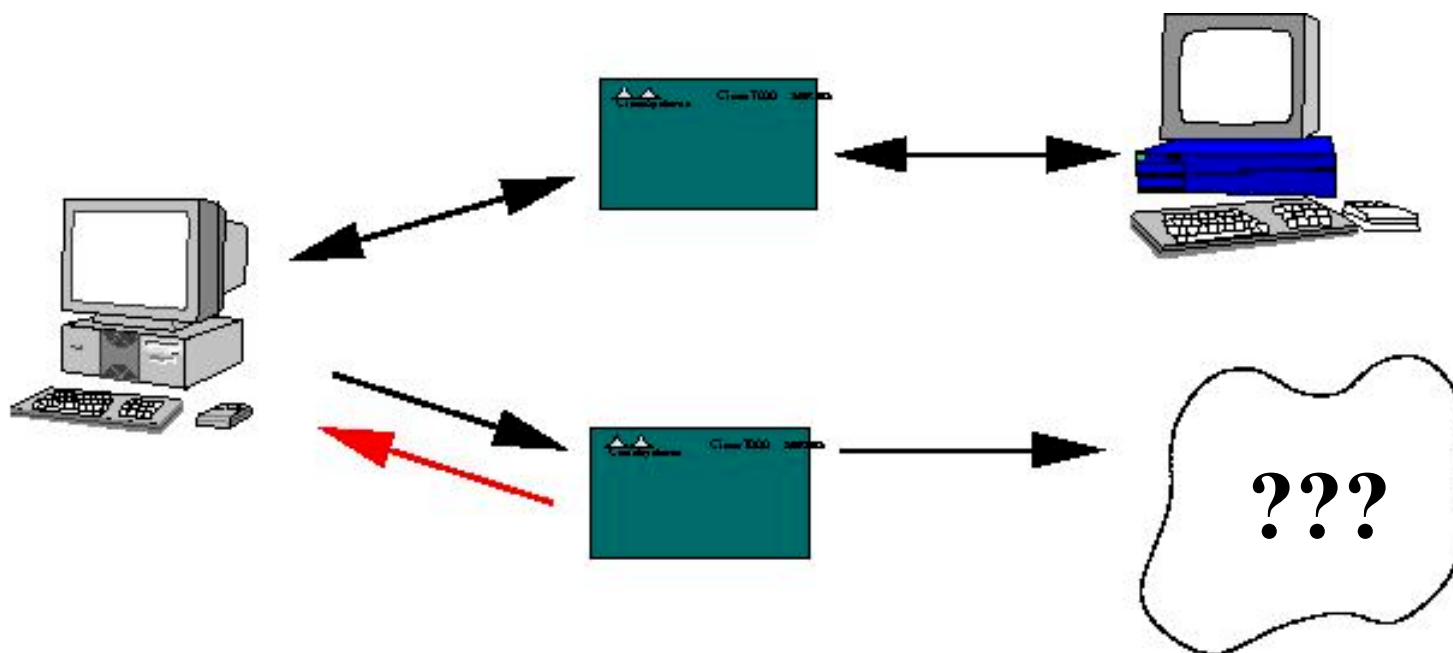
Code Red v2





ISTS

ICMP Destination Unreachable



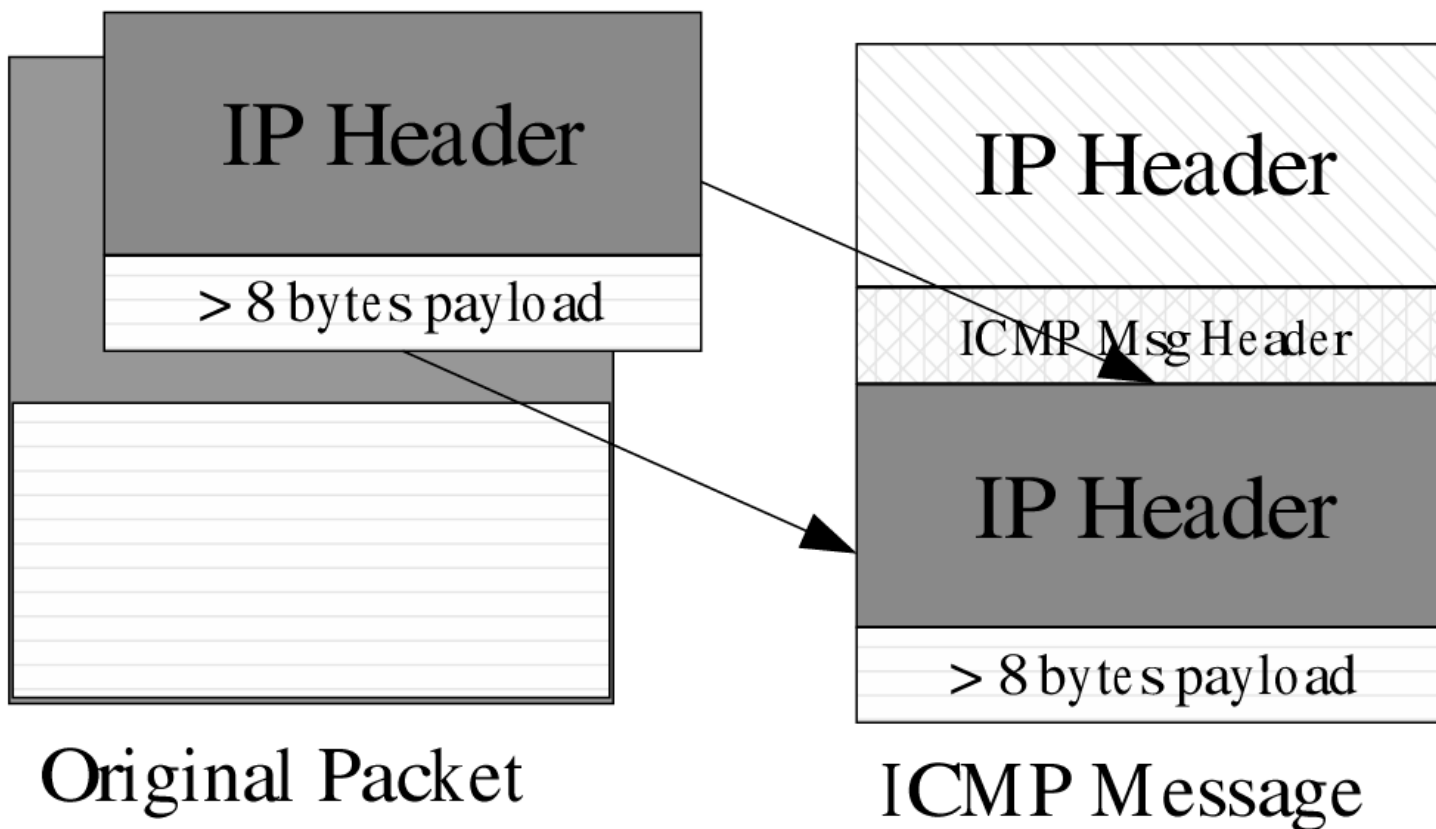


ICMP Destination Unreachable

- Can get generated when a host or network is not reachable. (About 6%-7%)
- An ICMP Destination Unreachable includes:
 - The IP header of the original packet
 - At least 8 bytes of its original payload
- This holds:
 - Source and Destination IP and protocol
 - For TCP and UDP: Source and Destination port



ICMP Destination Unreachable





Worms & ICMP Unreachables

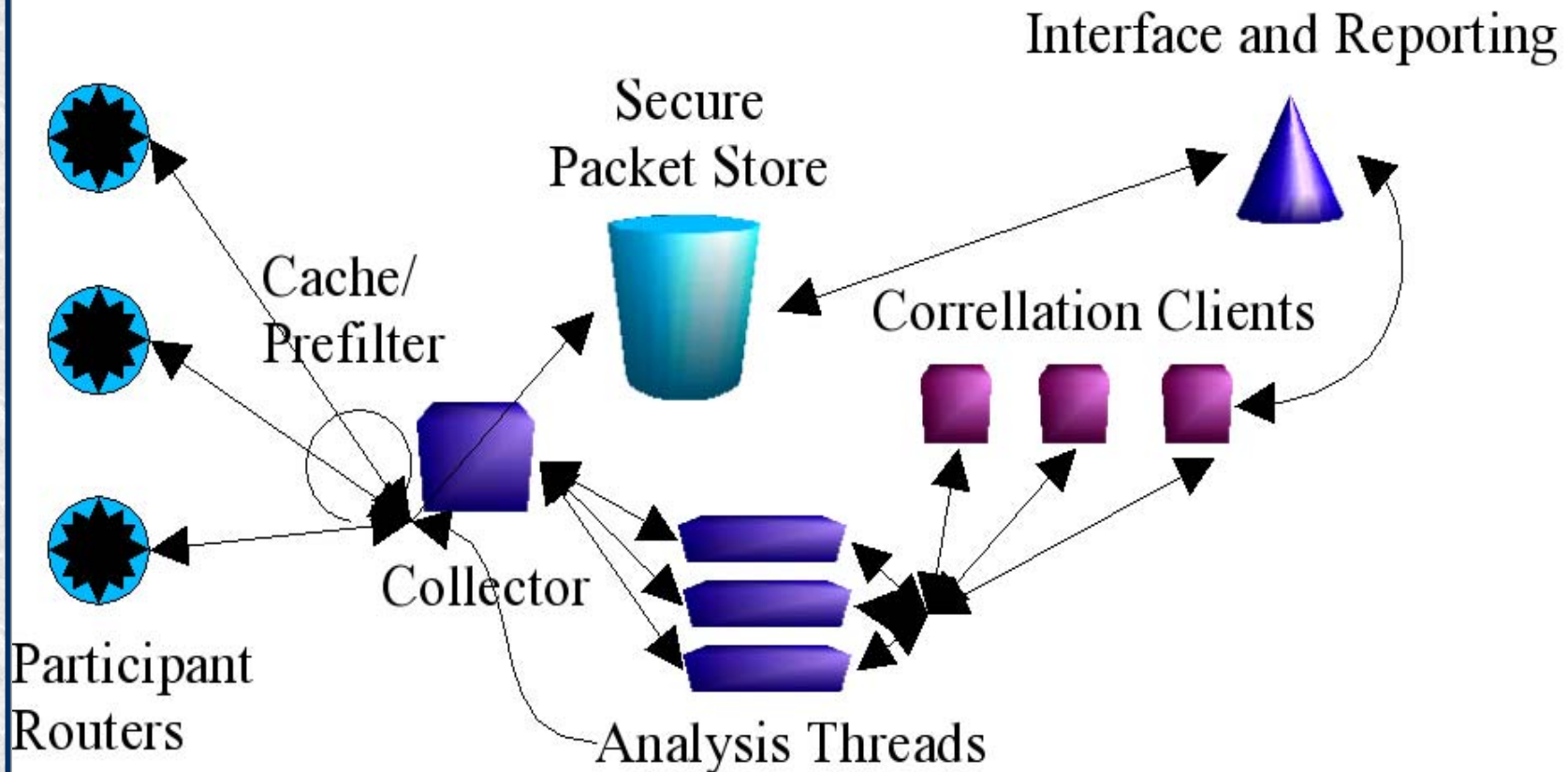
- Worm Paradigm:
 - Select Target
 - Probe for Vulnerability and (if possible) Infect
 - Repeat
- (Random) Target Selection Scanning will:
 - hit many unused IP addresses
 - in many different networks
- ICMP-T3s will come from many routers



DIB:S Collecting ICMP-T3

- Participating routers forward the ICMP Destination Unreachable messages that *they themselves* generate to an analysis point:
 - No privacy concerns regarding “sniffing”
 - Only “error messages” are analyzed
 - A Blind Carbon Copy is sent, optionally the original can still be dropped.
- Scanning shows as a “fan” or “bloom” of ICMP-T3s provoked by 1 IP address

DIB:S Framework





DIB:S Incoming Messages

- Incoming ICMP-T3 messages are:
 - Stripped of their IP/ICMP headers
 - Only Embedded data is used
 - Messages sorted on Embedded IP source and destination address
 - Analyzers are assigned an IP address range
 - Each Internet host will always be analyzed by the same Analyzer



DIB:S Reporting

- Parameters to the system:
 - N Threshold
 - t Timeout for received packets
- If one IP address contacts N other IP addresses on the same port with t seconds an Alert is issued.
- Other Alerts also possible
- A “track” of Alerts might indicate a Worm



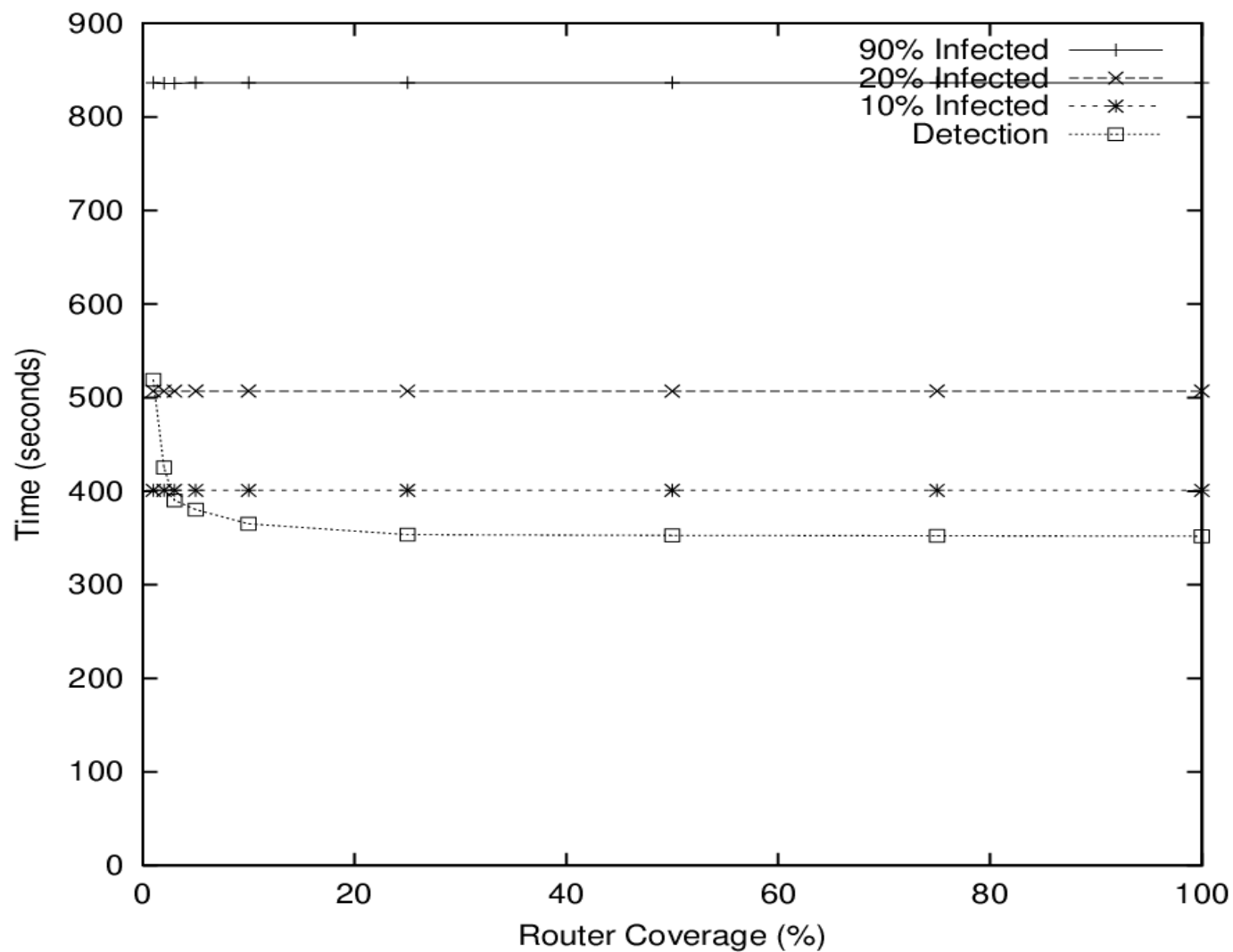
Simulation Results

- Simulation done on a network with:
 - 100.000 hosts
 - 25 percent is reachable
 - 0.1 percent is vulnerable
- DIB:S configuration:
 - $N = 10$
 - $t = 600$ seconds



Simulation Results

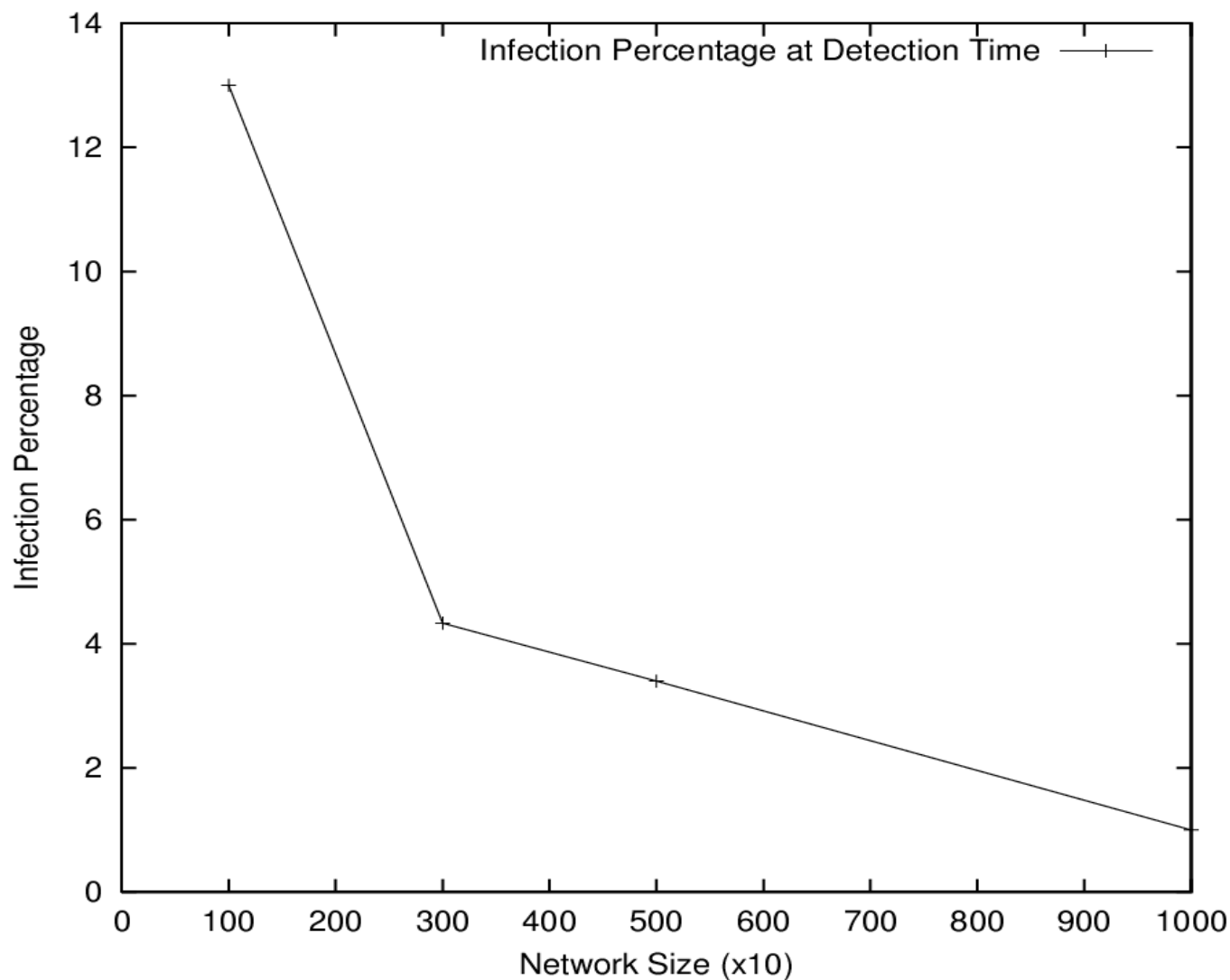
Worm Detection Time versus Router Coverage





Simulation Results

Infection Percentage at Detection Time versus Network Size





Simulation Results

- Random Noise:
 - Increases load on the system
 - Does not yield significant false-positives
 - BUT: don't set N too low ($N > 5$)
 - Detection based on initial exponential curve of worms
- Large scale Internet Simulation needed
 - Real Internet is less uniform



Future Work

- Deployment
 - LINUX based BCC routers
 - CISCO-IOS
 - For 1.5% coverage: 3.5 Class A networks needed
- Small network scale systems
 - Being a good Internet neighbour
- Use of other types of data
- Active Response



DIB:S

Dartmouth ICMP BCC: System

Vincent Berk

vberk@ists.dartmouth.edu