DIMACS Working Group on Analogies between Computer and Biological Viruses And Immune Systems

Lora Billings
billingsl@
Mail.montclair.edu

Ira Schwartz
schwartz@
nlschaos.nrl.navy.mil

Alun Lloyd
alun@
Alunlloyd.co

Stephanie
Forrest@
Cs.unm.edu

Organize

What Our Objectives Are

Main Working Group Objective-

Identify possible new math analysis, modeling, and analysis techniques for computer virus spread

Approach-

Presentations (new ideas)

Discussion and debate

Three breakout sessions for final presentation (if possible) to identify follow on problems

New models to account for differences between computer and biological viruses

Measures and Identification of virus spread

Control and Constraints to preserve information flow

(Group leaders to be assigned)

Operational Issues

Military and Non-military networks are dependent on:

- commercial technology
- global Internet

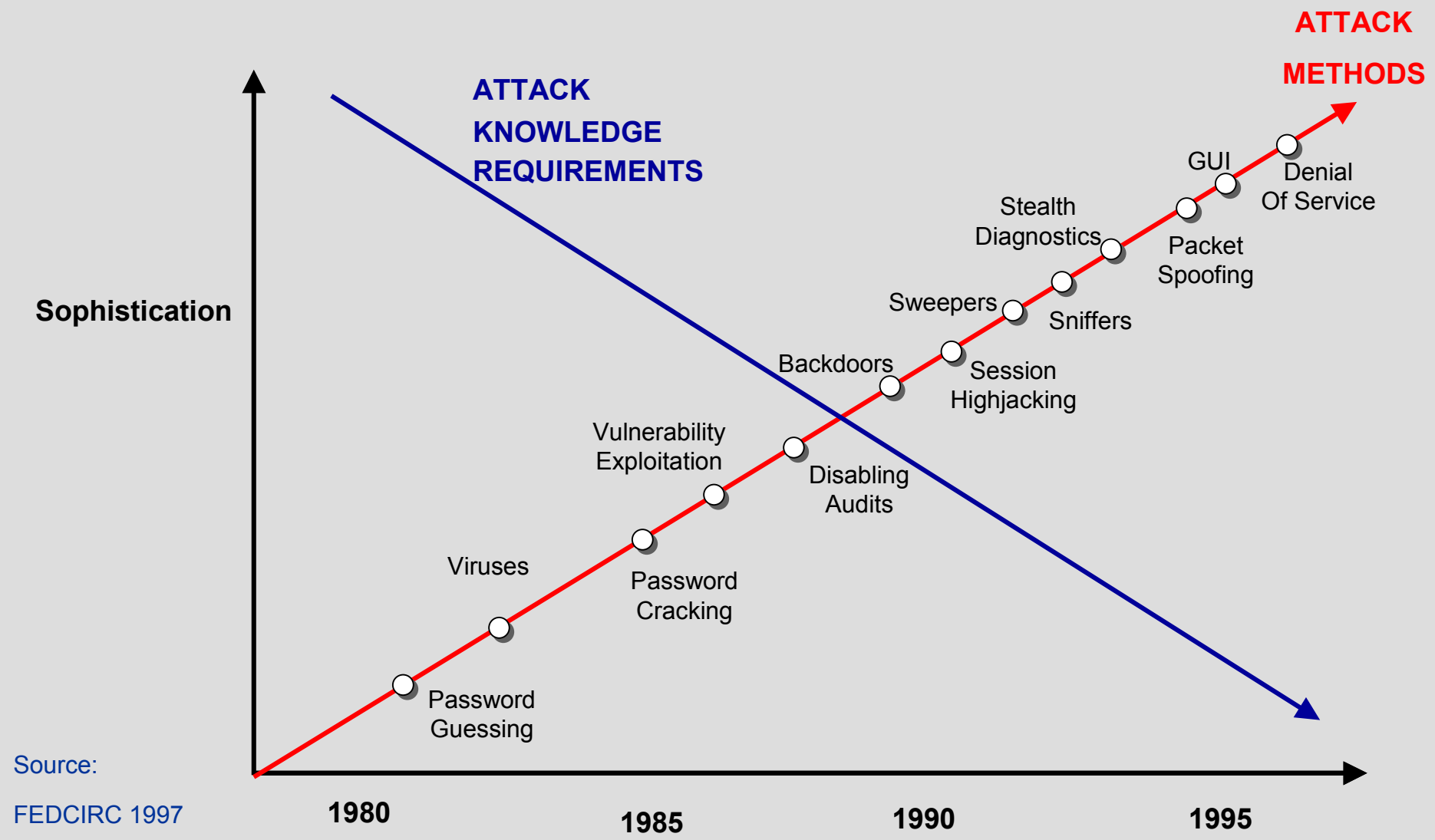
Number of infrastructure attacks increasing

- Viruses and worms,
- malicious mobile code
- denial of service

Major deficiencies

- Not secure
- Susceptible to information flow disruption

Attack Sophistication



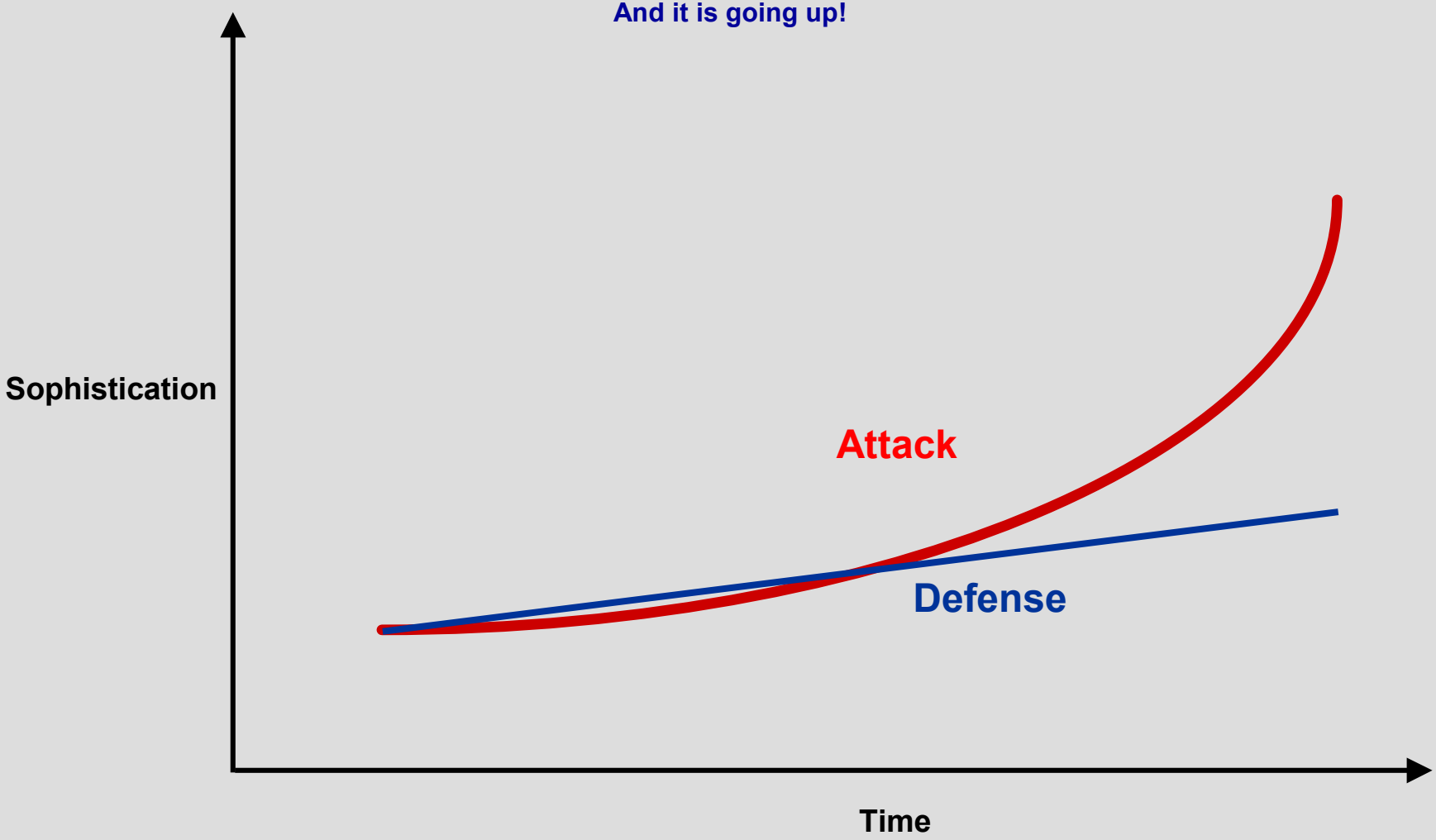
Source:

FEDCIRC 1997

The Widening Gap between Attack and Defend

NRL alone gets 1000 hits/day + 12 real attempted attacks/day

And it is going up!



Technical Problem

REAL THREAT: Sophisticated attacker-funded by nation states

Virus attacks created by sophisticated attacker will be harder to detect

- Current viruses have known structures
- Known manufacture and distribution mechanisms
- Visual Basic Worm Generator Kit

Problem: Defend against virus of unknown structure and/or self-distribution mechanism

Technical Approach

Overall goal: Virus spread prediction and control

Current Approaches:

- Information Flow Modeling
- Operation Network Modeling
- Study of Real World Virus Behavior
- Validation of Models – Data issues

Goals

Mathematical Goals:

- **Estimating the degree to which a virus will spread;**
- **Understanding the early history of a virus;**
- **Assessing the impact of proposed interventions;**
- **Optimizing the impact of prevention strategies;**
- **Assessing the effectiveness of partial protection**

Modeling Techniques

Continuum Models

Positive attributes:

Analyzable (based on a rich theoretical foundation)

- Sensitivity Analysis (forward and backward) is readily performed
- Model is defined in terms of available population based parameters
- Efficient prediction of mean behavior in the presence of low amplitude noise

Negative attributes:

Severe limitation on the dimension of the state space (risk, susceptibility…)

- Poor information about low probability events (tails of distributions)

Cannot handle individual level behavior

Modeling Techniques

Agent Based Models (Plusses):

- Few limitations on the dimension of the state space (risk, susceptibility…)
- Good information about low probability events (tails of distributions)
- Model directly incorporates individual level behavior

Agent Based Models (Minuses):

- Few analytical tools (nascent theoretical foundations)
- No backward sensitivity analysis
- Individual level data often has to be derived from population level data
- Computationally less efficient than continuum model

Information Flow Modeling-A Quick Example

Network of N nodes: S susceptible, I infected

probability of susceptible becoming infected, μ

probability of infected becoming recovered, δ

probability of transmitting virus, β

probability of connectivity, c

Markov

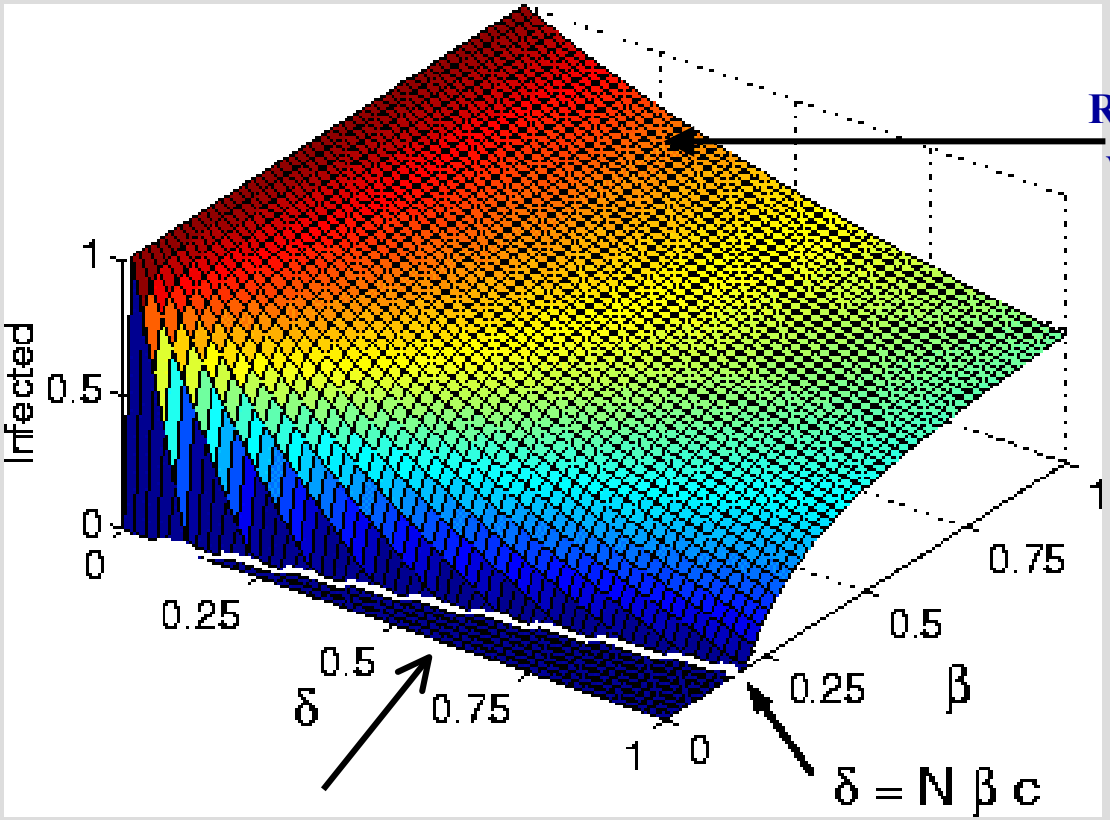
$$p_{I,I'} = \sum_{x} \binom{I}{x} (\delta)^{x} (1-\delta)^{I-x} \left(\binom{s}{x+I'-I} (\mu)^{x+I'-I} (1-\mu)^{S-(x+I'-I)} \right)$$

Deterministic

$$\frac{di}{dt} = (1-i)\mu(Ni,\beta) - i\delta$$

$$\mu(I,\beta) = \sum_{y=0}^{I} (1-(1-\beta)^{y}) \binom{I}{y} c^{y} (1-c)^{1-y}$$

Results of Compartment Model

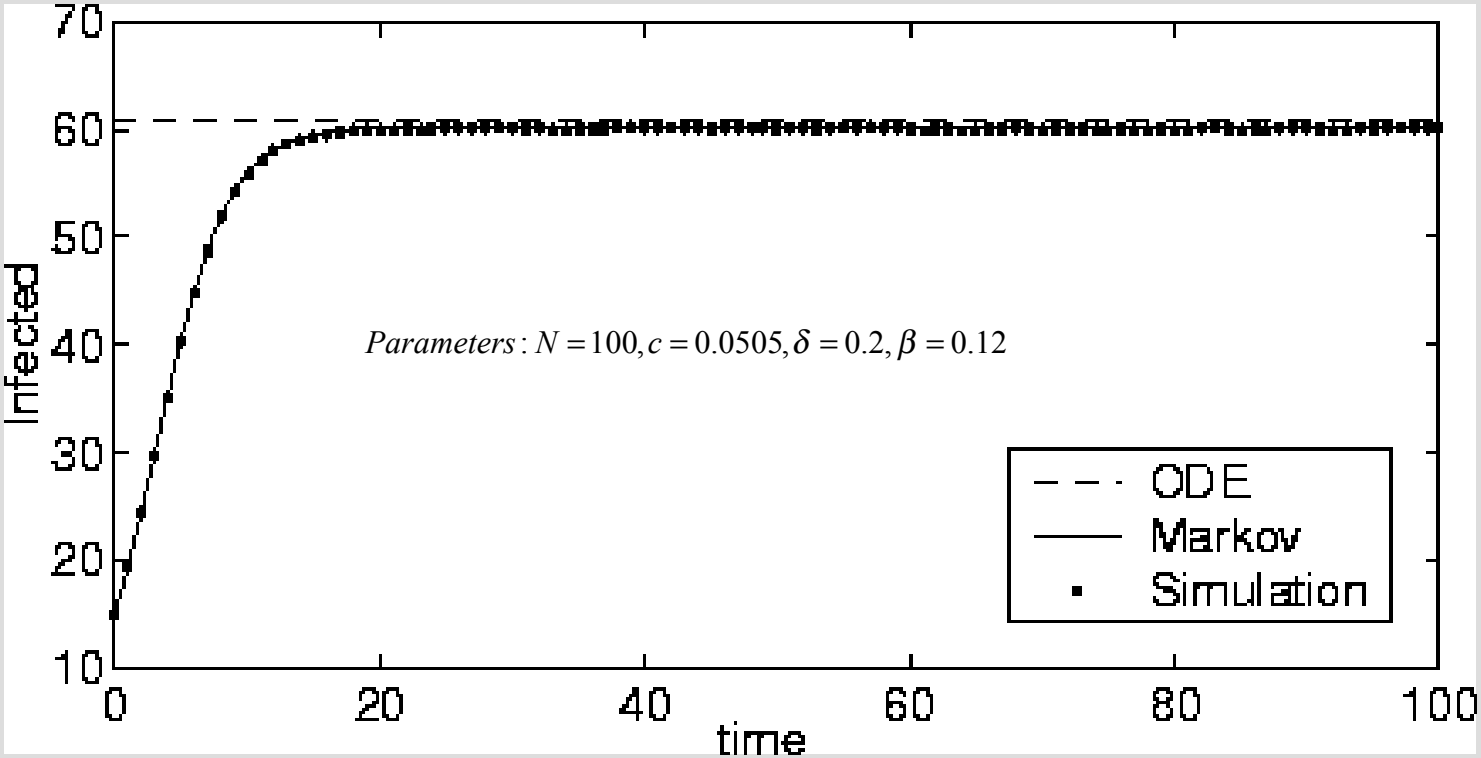


Region of Virus
Die Out

Predicting successful virus propagation as a function
of competing virus transmission and computer recovery

Comparison of Models

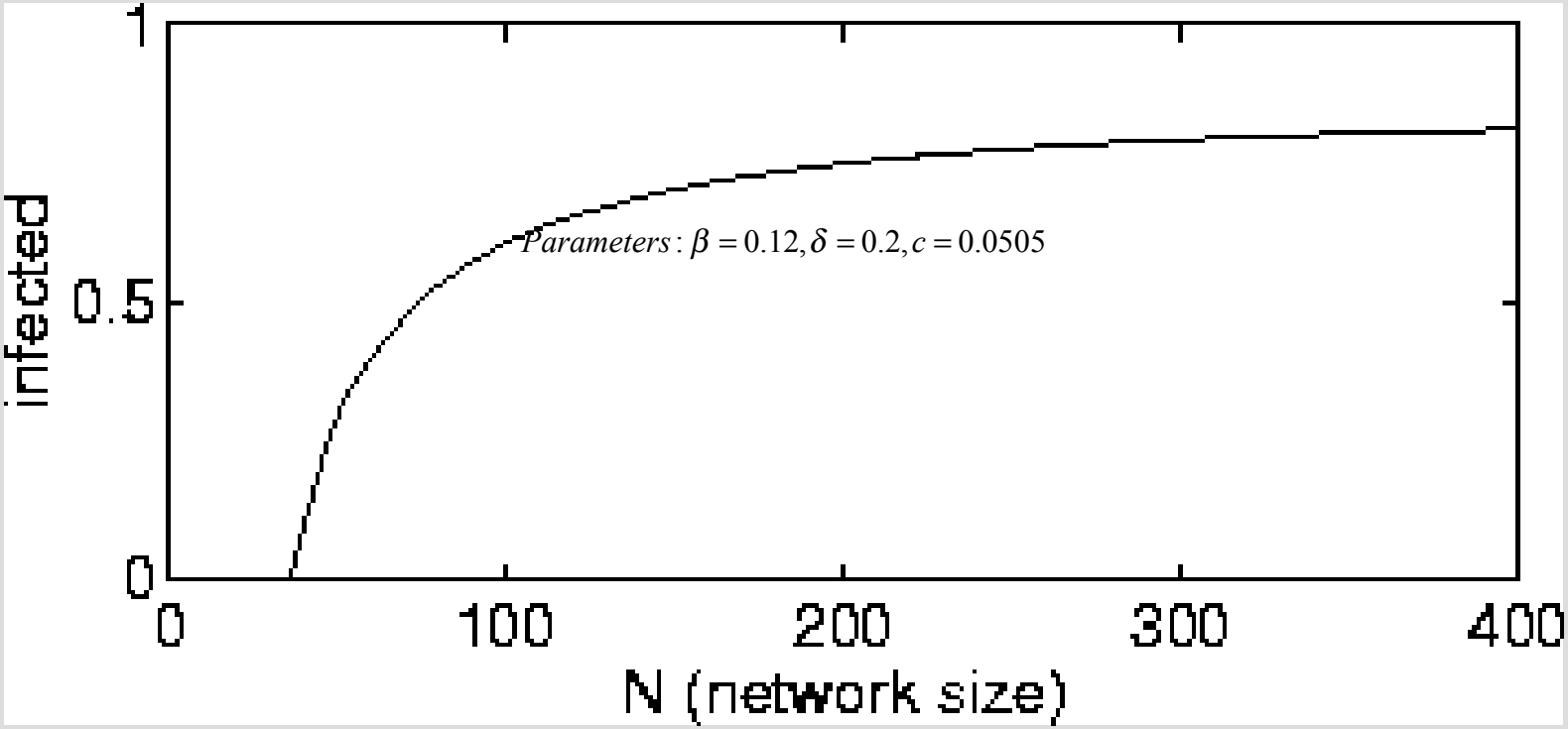
Model fidelity between agent (Markov) and compartment (ODE) simulations



$$Parameters : N = 100, c = 0.0505, \delta = 0.2, \beta = 0.12$$

Relationship Between Virus Spread and Network Size

Infected fraction decreases nonlinearly as network size decreases



$$Parameters: \beta = 0.12, \delta = 0.2, c = 0.0505$$

Anna Kournikova Virus

First detected at 1:21 pm, Feb 12, 2001

Microsoft Outlook worm

- written with Visual Basic Worm Generator Kit
- Email attachment (AnnaKournikova.jpg.vbs)
- Sends itself out to all entries in Address Book
- Appears to do no damage

Mail.com intercepted nearly 53,000 copies on 2/12

Approximately 100,000 Australian computers infected

McAfee's World Virus Map:

California most infected, Asia and Africa least infected

MessageLabs Virus Data

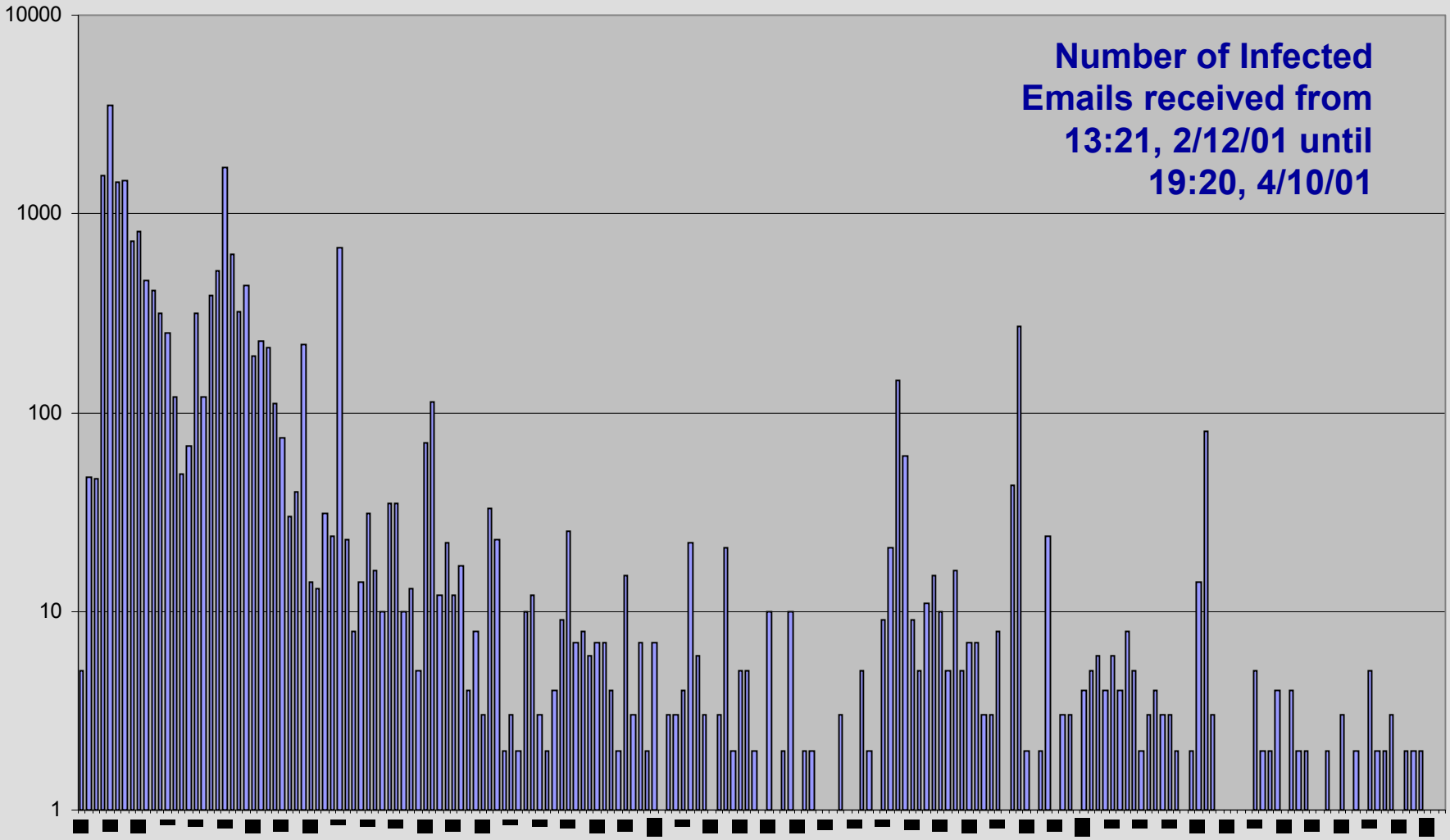
MessageLabs

- Internet Service Provider in UK
- All customer email scanned by four virus detectors
  - » McAfee, F-Secure, VFind, and Skeptic
  - » Skeptic - MessageLabs heuristics and rules based scanner.
  - » Signatures updated every ten minutes
- Email stopped and held for 30 days when virus detected.
- Sender & Administrator notified.
- Protected customers from Melissa, LoveBug & Anna viruses.
- **Internet virus "black hole"**

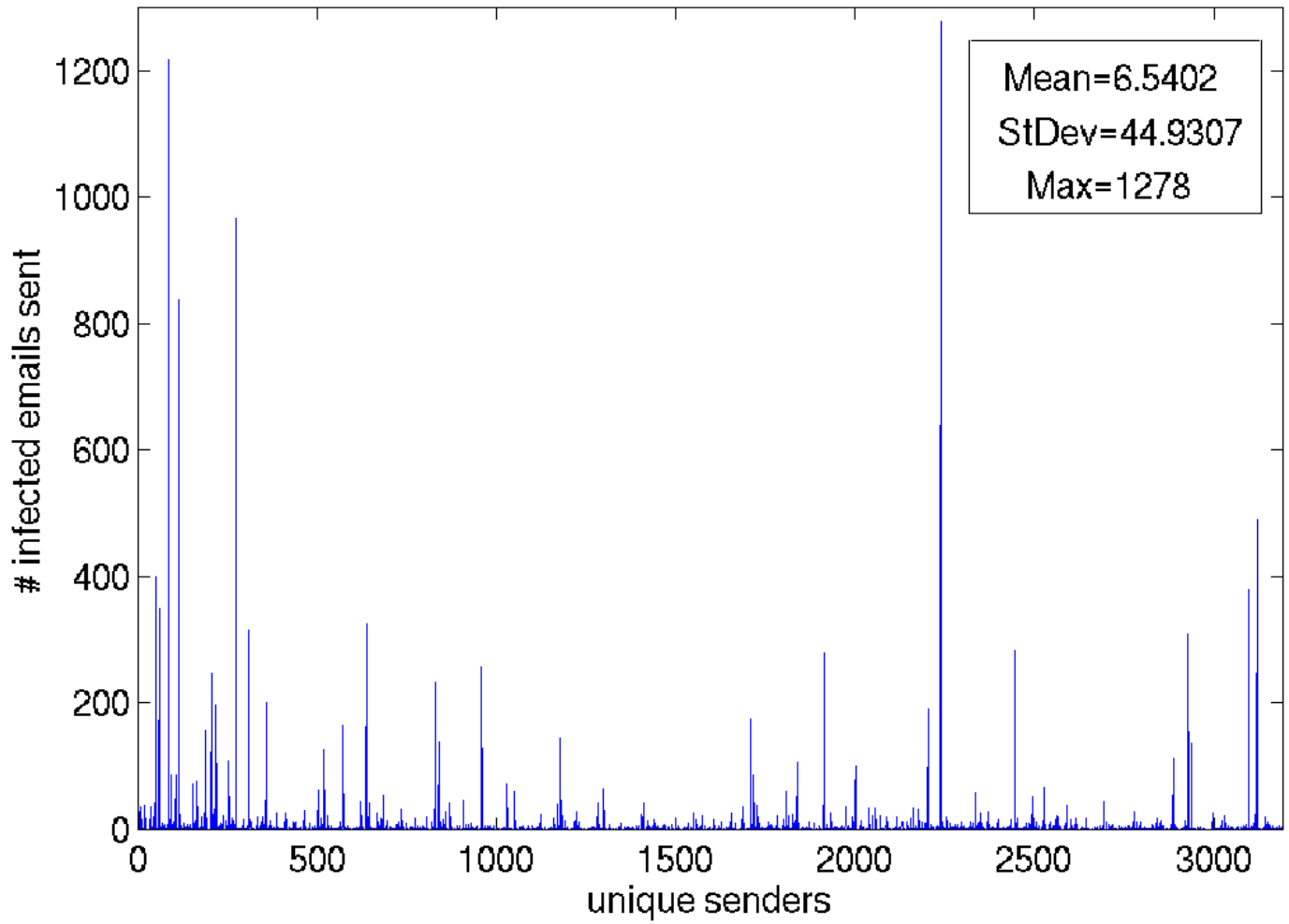
Received data files containing:

- Date and time of infected emails received by MessageLab
- Country of origin
- Unique number to identify source

MessageLabs: Infected Email per Hour



MessageLabs: Number of infected emails received from unique sources



Conference Themes

Math Bio:

- Population and epidemiological modeling
- Deterministic and stochastic viral spreading

Network modeling:

- Globally connected
- Locally connected
- Scale free

Control and prediction

- Continuous populations
- Discrete populations

Computer Science:

- Real experiences-from the trenches
- Immune modeling