Code Red Worm Propagation Modeling and Analysis

Zou, Gong, & Towsley

Michael E. Locasto

March 4, 2003

Paper # 46

Overview

- Code Red incident data & impact
- epidemiology models
  - traditional (biological) infection models
  - two-factor worm model
- related work & questions
  - (Weaver & Sapphire)

Motivation

- Internet great medium for spreading malicious code
  - Code Red & Co. renew interest in worm studies
- Issues:
  - How to explain worm propagation curves?
  - What factors affect spreading behavior?
  - Can we generate a more accurate model?

Epidemic Models

- Deterministic vs. Stochastic
  - Simple epidemic model (paper #45)
  - general epidemic model (Kermack-Mckendrick add notion of removed hosts)
- good baseline, need to be adjusted to explain Internet worm data
- any model must be deterministic (b/c of scale)

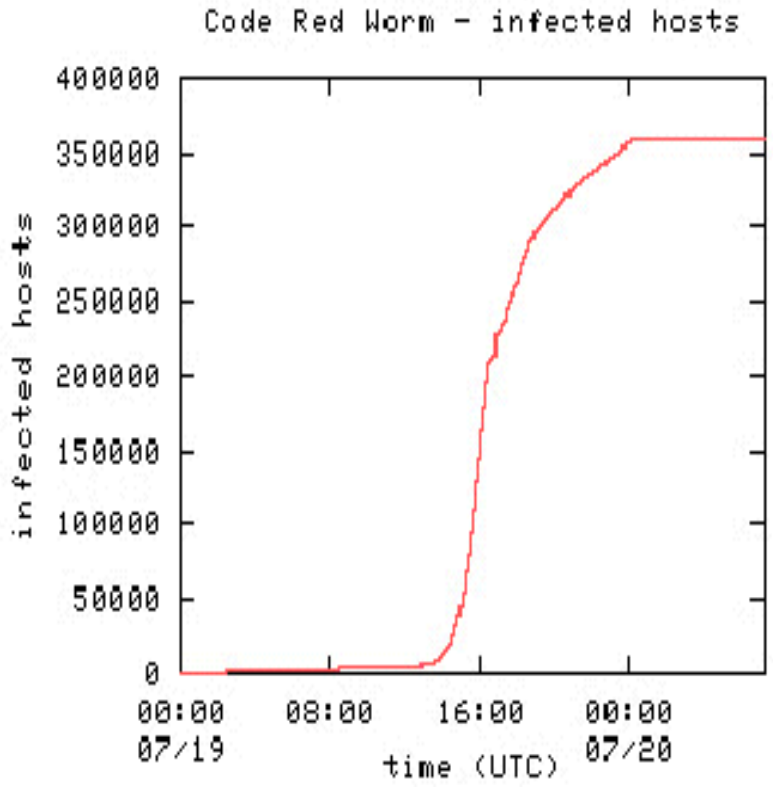
Two-Factor Worm Model

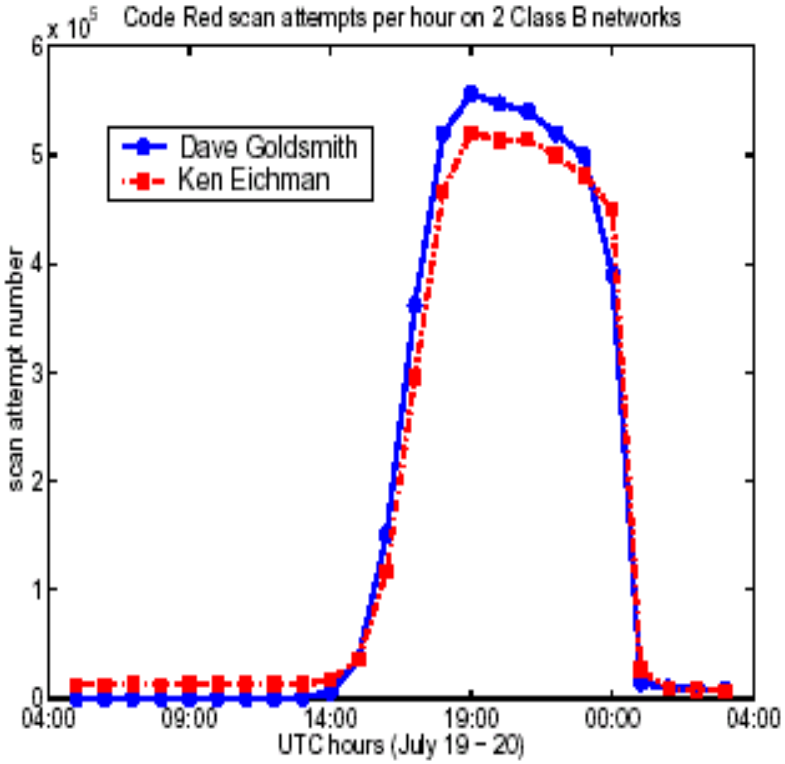
- Two major factors affect worm spread:
  - dynamic human countermeasures
    - anti-virus software cleaning
    - patching
    - firewall updates
    - disconnect/shutdown
  - interference due to aggressive scanning
- Rate of infection (ß) is **not** constant

Two-Factor Worm Model (con)

- Two important restrictions:
  - consider only “continuously activated” worms
  - consider worms that propagate w/ort topology

Infection Statistics

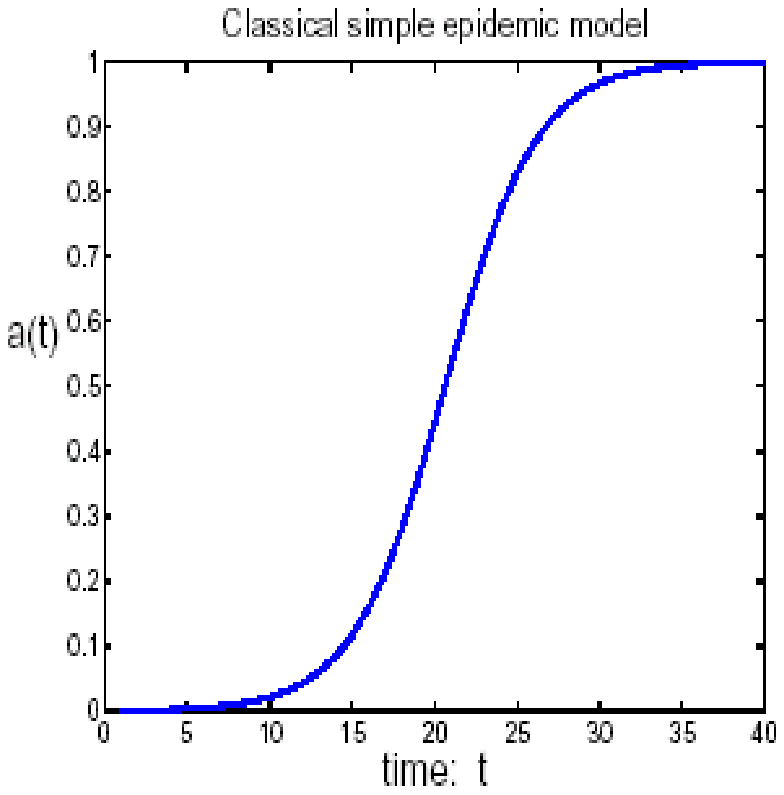




a. Code Red scan attempts

Classic Simple Epidemic Model

- Model presented in paper 45 (classic simple epidemic model, k=1.8, k=BN)
- a(t) = J(t) / N (fraction of population infected)
- Wrong! (compare to last slide)



Simple Epidemic Model Math

- Variables:
  - infected hosts (had virus at some point) = J(t)
  - population size = N
  - infection rate = ß(t)
- $dJ(t)/dt = \beta J(t)[N - J(t)]$

Two-Factor Model Math

- $dI(t)/dt = \beta(t)[N - R(t) - I(t) - Q(t)]I(t) - dR(t)/dt$
  - S(t) = susceptible hosts
  - I(t) = infectious hosts
  - R(t) = removed hosts from I population
  - Q(t) = removed hosts from S population
  - J(t) = I(t) + R(t)
  - C(t) = R(t) + Q(t)
  - J(t) = I(t) + R(t)
  - N = population (I+R+Q+S)

Two-Factor Fit

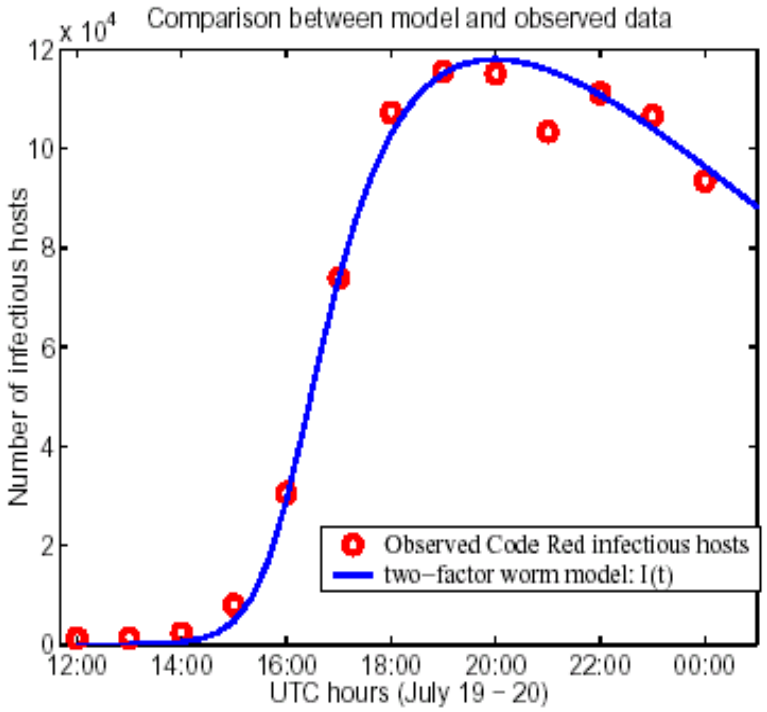


Figure 9: Comparison between observed data and our model

- Take removed hosts from both S and I populations into account
- non-constant infection rate (decreases)
- fits well with observed data

Results

- Two-factor worm model
  - accurate model without topology constraints
  - explains exponential start & end drop off
  - identifies 2 critical factors in worm propagation
- Only 60% of CR targets infected