Parallels Between Biological and Computer Epidemics

Tom Chen
SMU
tchen@engr.smu.edu

Outline

- Microscopic: How Biological and Computer Pathogens Spread
- Macroscopic: Biological and Computer Epidemiology
- Human and Artificial Immune Systems

Computer Pathogens

- Viruses and worms are characterized by capability for self-replication
  - Viruses: parasitic ability to self-replicate by modifying (infecting) a normal program/file with a copy of itself
  - Worms: stand-alone programs that exploit security holes to compromise other computers and transfer copies of itself through a network

Virus - Biological Parallels?

- Viruses named by Fred Cohen in 1983 after biological viruses
  - Biological viruses are strands of RNA or DNA in protein shell, not alive or complete by themselves
  - Parasitically infect a normal (host) cell
  - Hijack control of host cell’s reproductive machinery to reproduce more viruses

Viruses - What are They

Biological virus

DNA or RNA strand surrounded by protein shell

No life outside of host cell

Computer virus

Set of instructions

Incomplete program - not executable by itself

Viruses - How They Infect

Biological virus

Outer protein shell bonds to normal (host) cell

Virus RNA or DNA takes over control of host cell

Computer virus

Virus code attaches to or overwrites normal (host) program or file

Virus code takes over control when host program is executed

Viruses - Replication

Biological virus

Virus RNA or DNA hijacks host cell's reproductive machinery to produce more viruses

Computer virus

Virus code contains instructions to copy itself to other locations (programs, files, disks,...)

Viruses - Transmission

Biological virus

Transmitted to other individuals by various vectors - air, water, physical contact,...

Computer virus

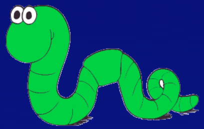
Transmitted to other computers by various vectors - email, disks, file sharing,...

Worms - Biological Parallels?

- Worms named by Shoch and Hupp (Xerox) in 1979 after electronic network-based “tapeworm” in John Brunner’s novel, “The Shockwave Rider”
  - Envisioned multi-segmented distributed program spread over many computers
  - Impervious to deletion of any segments
  - Not really how modern worms work

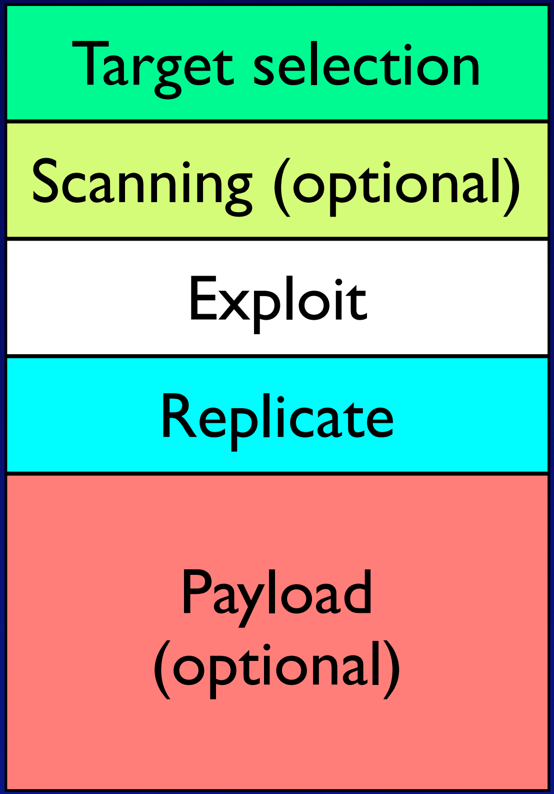
Biological Parallels?

Computer virus ↔ Biological virus

Worm  ↔ ~~Worm~~

What is a better analogy?

Worm Anatomy

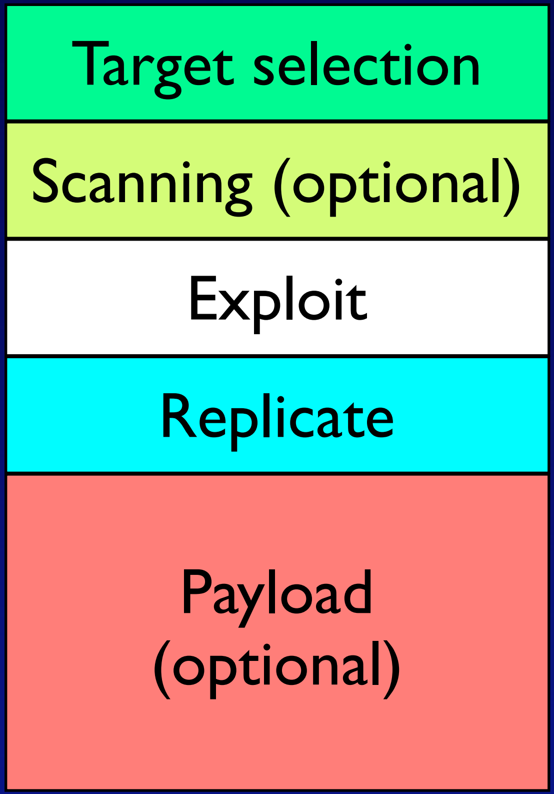


Target selection	- Chooses candidates to target
Scanning (optional)	- Learns suitability of target
Exploit	- Compromises security of target
Replicate	- Transmits worm copy to target
Payload (optional)	- Damage to target

SQL Slammer Example

- Starting January 25, 2003, SQL Slammer worm infected 200,000+
- Entire worm is 376 bytes carried in a single 404-byte UDP packet
- Exploited vulnerability in Microsoft SQL Server Resolution Service, included in MS SQL Server 2000 and MS Data Engine 2000

SQL Slammer Anatomy



Target selection

Scanning (optional)

Exploit

Replicate

Payload (optional)

- Chooses random IP addresses

- No scanning

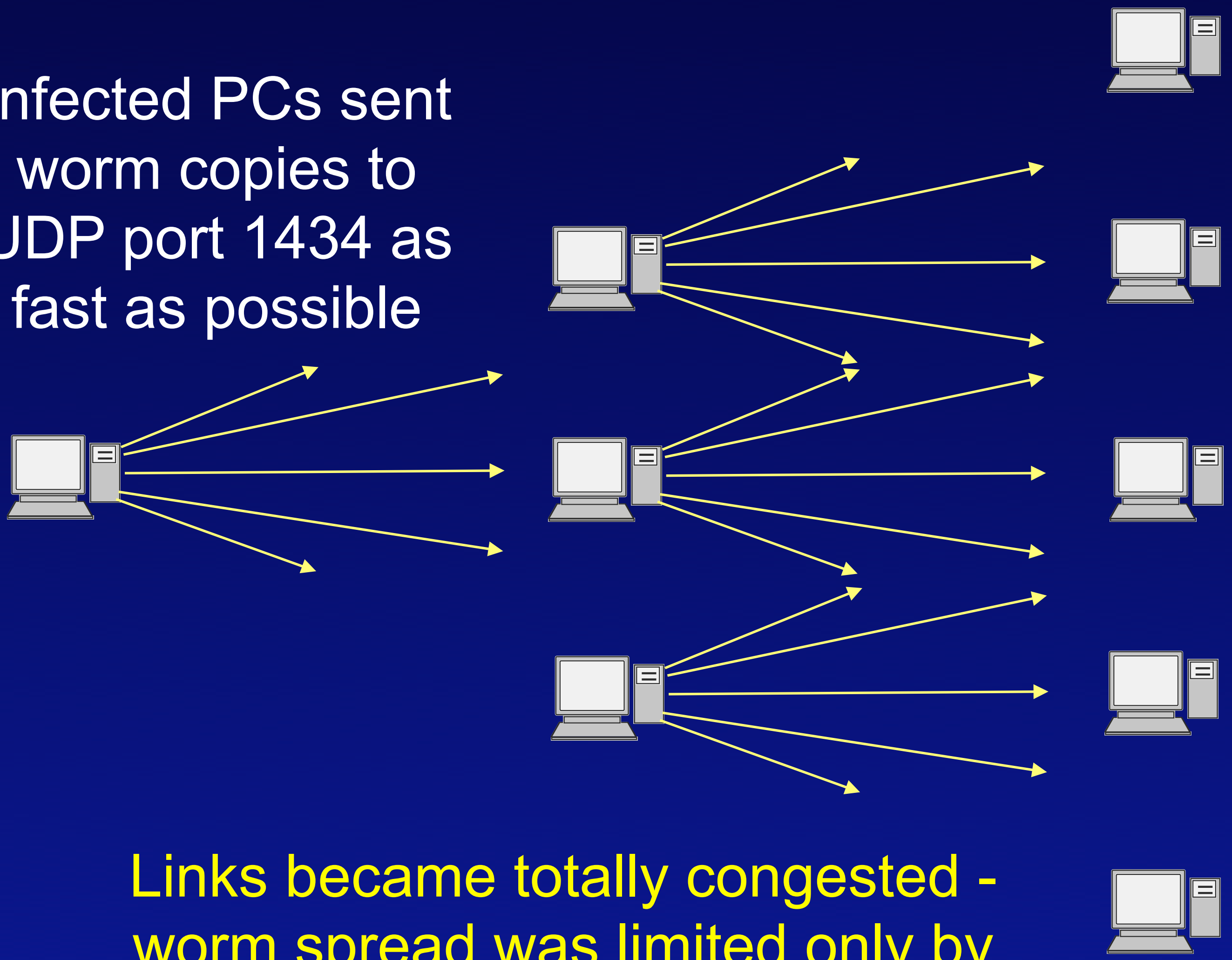
- Buffer overflow attack to UDP port 1434 (MS SQL Monitor port)

- UDP packet carries worm copy; infected targets are put into infinite loop to send out worm copies

- No payload

Slammer (cont)

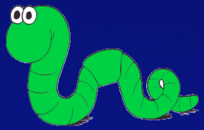
Infected PCs sent worm copies to UDP port 1434 as fast as possible



Links became totally congested - worm spread was limited only by available bandwidth

Biological Parallels?

Computer virus ↔ Biological virus

Worm  ↔ Cancer

Uncontrolled growth and metastasis

At Microscopic Level

- Despite obvious differences (electronic vs. biochemical), both computer pathogens and biological pathogens have found ways to (i) reproduce (ii) transmit themselves (iii) infect others
- Parallels in general behavior can be made, but no research done -- no practical benefit

At Macroscopic Level

- Epidemic modeling is concerned with spread of diseases among individuals in population
- Epidemic models make simplifying assumptions to gloss over the complexities at microscopic level
- Models are abstract enough for both computer pathogens and biological pathogens

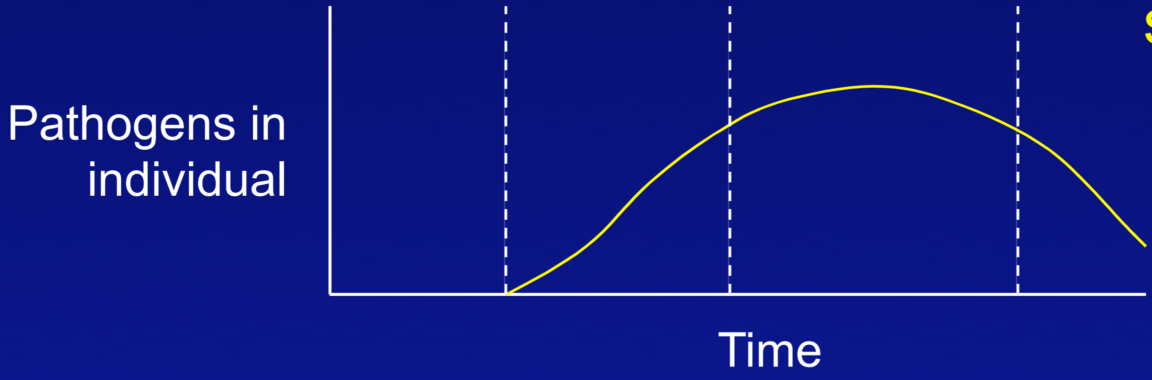
Epidemic Modeling

- Epidemic modeling helped devise vaccination strategies, eg, smallpox
- We would like to borrow the deterministic and stochastic models developed over 250 years of human diseases
- Little done so far -- only basic epidemic models used for viruses and worms

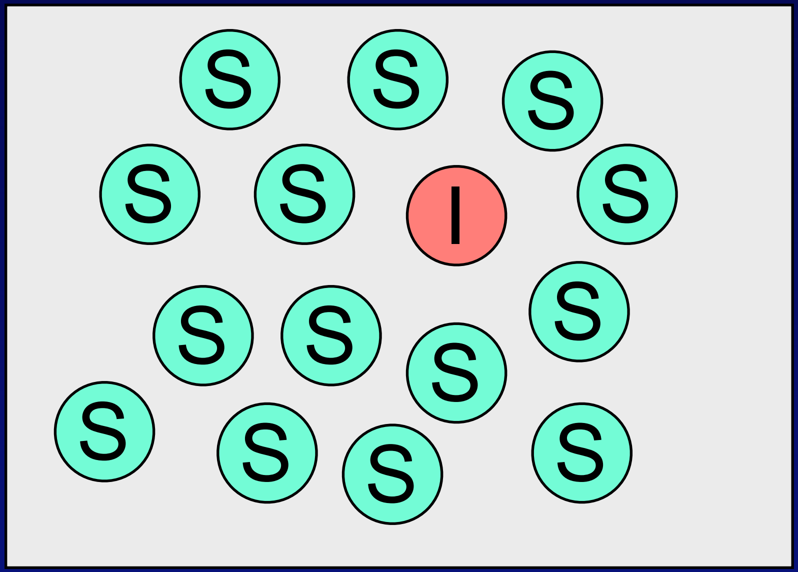
Usual Assumptions

- Individuals are assumed to progress through number of states

Susceptible → Latent → Infectious → Immune or dead or susceptible



Simple Epidemic (S-I) Model



- Individuals progress from Susceptible → Infected states (hence, "S–I model")

S = number Susceptibles

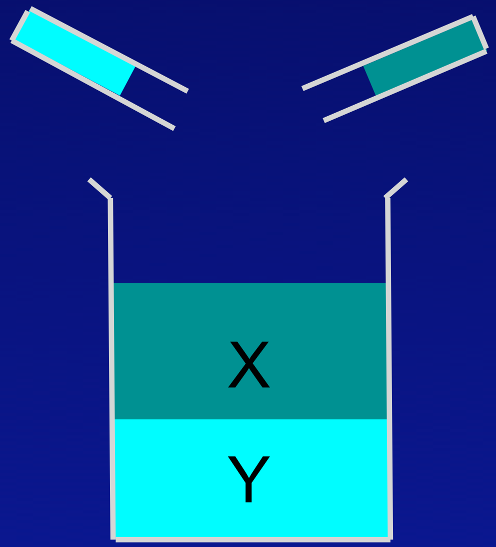
I = number Infecteds

N = S + I
= fixed population

– Susceptibles and Infecteds mix randomly

Law of Mass Action

- In chemical reactions, rate of reaction is proportional to product of masses (X·Y)
  - Fastest reaction when both X and Y large



Simple Epidemic (cont)

- Simple epidemic model applies law of mass action:
  - Rate of interactions between Susceptibles and Infecteds is proportional to product S·I

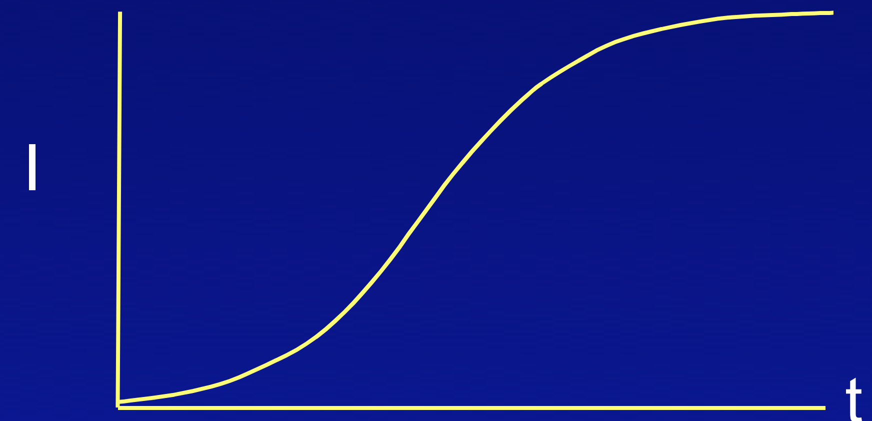
$$\frac{d}{dt}I = \beta SI$$

β= infection rate parameter

Simple Epidemic (cont)

- Solution: number of Infecteds shows logistic growth

$$I_t = \frac{I_0 N}{I_0 + (N - I_0)e^{-\beta N t}}$$



General Epidemic Model

- In addition, assume individuals progress from Susceptible → Infected → Removed (dead or immune)
  - Also called S-I-R model
  - R = number of Removed
- Assume Infecteds become removed at constant rate γ per capita

General Epidemic (cont)

- No closed solution to S-I-R model:

$$\frac{d}{dt}S = -\beta SI$$

$$\frac{d}{dt}I = \beta SI - \gamma I$$

$$\frac{d}{dt}R = \gamma I$$

General Epidemic (cont)

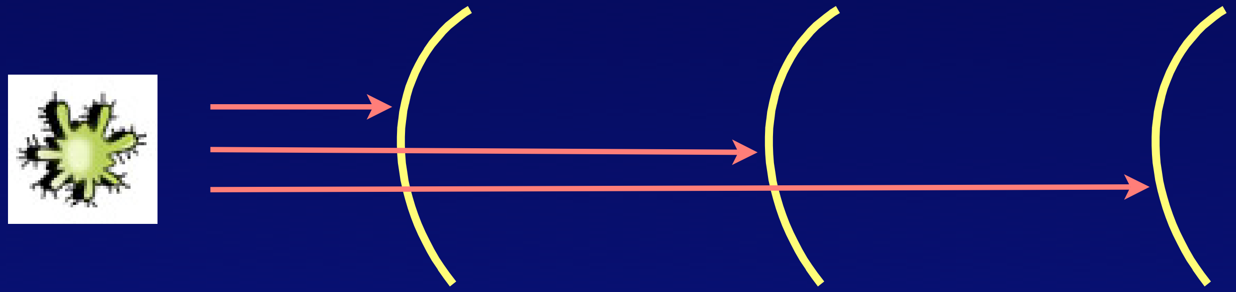
- Researchers have tried to apply S-I-R model to worm epidemics
  - Modifications include making β and γ parameters dependent on other factors, instead of constants
- Models need to take network characteristics into account, but not much progress

Artificial Immunity

- Researchers want to design artificial immune systems inspired by human immune system
  - Obvious differences (electronic vs. biochemical) but seek to borrow general principles
  - Human immune system is not perfect but amazingly effective against even new pathogens

Human Immunity

- 3 layers



Physical barriers (skin,...)

Innate immune system (common to all animals)

Adaptive immune system (prompted to action when needed)

Innate Immune System

- Innate immune system includes diverse weapons for fast defenses:
  - Phagocytes: white blood cells to "eat" cells
  - Complement system: proteins bind to chemical groups on common viruses, marks them for phagocytes
  - Natural killer cells: a mystery how decide which cells to kill, most potent when activated by interferon produced by infected cells

Adaptive Immune System

- When innate immune system struggles a while, it can trigger adaptive immune system including:
  - B cells producing antibodies
  - Killer T cells

Adaptive Immune System

- B cells:
  - 100 million different B cells are produced by various combinations of 120 different gene segments
  - When B cell binds to a matching virus, it produces masses of matching antibodies that mark viruses for phagocytes
  - Some B cells become “memory B cells” to remember a detected virus for later

Adaptive Immune System

- Killer T cells:
  - Diverse as B cells, constructed by various combinations of gene segments
  - Work by looking inside cells -- can detect cells already infected by virus
  - Kill infected cells to stop virus from replicating

Interesting Features

- Multiple layers -- for robustness
- Distributed detection -- detectors circulate around body
- Specific detectors -- antibodies bind only to matching viruses
- Diversity of detectors -- many, many different B cells created through combinatorics of gene segments

Interesting Features (cont)

- Adaptive -- antibodies finding a matching virus are replicated
- Learning and memory -- memory B cells remember detected viruses
- Detection of new viruses by anomaly detection -- detectors recognize “self” (normal cells) vs. “non-self” (pathogen)
  - Thymus deletes self-reacting B and T cells

Artificial Immune Systems

- Researchers have tried to borrow specific (not all) principles, with limited success
- Symantec's Digital Immune System
  - Suspicious files detected by antivirus software are automatically sent to Symantec
  - Symantec analyzes and creates signature
  - New signatures are automatically downloaded to update clients' antivirus software

Artificial Immunity

- Intrusion detection systems (IDSs) use anomaly detection
  - “Normal” traffic or system behavior is defined (”self”)
  - Anything else is classified as suspicious (”non-self”)
  - But definition of normal is problematic

Conclusions

- Parallels at microscopic level are not being pursued
- Epidemic modeling at macroscopic level is promising but unclear how to progress
- Human immunity is inspirational, but limited success in applying principles to artificial immune systems