

Epidemiology of on-line disease

G. Kesidis
CSE and EE Depts
Pennsylvania State University
kesidis@engr.psu.edu

BIONETICS, Avignon, Dec. 10, 2009

Outline

- introduction
- epidemics in living populations
- SIR models and inoculation dynamics
- discrete-time epidemics on graphs
 - graph models of social ties
 - disease spread via social engineering
 - branching process approximation of an SI epidemic
- peculiarities of on-line disease
 - spread in network and/or application layers
 - speed of spread (Slammer worm example)
 - selection of candidates for infection
- discussion
- references

Introduction

- A great variety of epidemic models have been proposed.
- Some models of the spread of epidemics are similar to those that model
 - the spread of gossip,
 - the formation of consensus,
 - and, more fundamentally, flooding algorithms.
- We will focus here on deterministic SIR models.
- On-line, the spread of malicious software is a serious threat to
 - privacy,
 - identity, and
 - access to communication and computing resources.

Disease spread among living populations

- Depends on:
 - The dynamics of physical contact by infected individuals required to pass the disease, in turn depending on
 - * how the disease was introduced and established into the population, and
 - * how the disease naturally affects the mobility of those infected, including the removal of infected individuals by death, in addition to pre-existing mobility factors such as geographical barriers.
 - Any incubation time the disease requires before it is contagious *i.e.*, infected but not yet infective.
 - The probability of disease transfer given the necessary physical transaction by an infective.
- A particularly lethal disease may quickly destroy the local population into which it was introduced before an opportunity presents itself to spread outside the local population.

Epidemic Countermeasures

- Countermeasures to the spread of disease can target any of these factors.
- For infected individuals, this could include:
 - medical treatments to cure the disease or reduce its contagiousness, and
 - physical interventions: from wearing of face-masks to reduce contagiousness; to limitations to mobility, *e.g.*, avoiding points of congregation such as bus terminals and schools; to complete quarantine.
- Also, uninfected individuals may take precautions, *e.g.*, inoculation to reduce or eliminate susceptibility, esp. those who tend to congregate.
- Often countermeasures are delayed, *e.g.*, by the need to test an inoculation for side-effects.
- Finally, diseases may mutate to circumvent acquired immunities, so individuals may continually transition at certain time scales among infective, recovered and susceptible states.

Epidemic models: Introduction

- Some models target
 - the details of individual contact between susceptibles and infectives,
 - the lifetime of individual infectives prior to their “removal” (including death or cure),
 - a dormant period between the time an individual is infected and when that individual becomes an infective, *etc.*
- Other models simply attempt to capture how *average* populations of susceptibles and infectives evolve depending on certain important parameters.
- Stochastic approximations [Kurtz’81] can be used to relate simpler deterministic models to scaled limits of more detailed stochastic ones.

Deterministic SIR model: closed population

- For closed (fixed) population of N (susceptible) peers, at time t let
 - $x_i(t)$ be the number of peers infected with the disease,
 - $x_s(t)$ be the number of uninfected peers, and
 - $x_r(t)$ be the number of “removed” peers (dead/offline or cured/patched).

- Assume $x_s(0), x_i(0) > 0$ and $x_r(0) = 0$, and note

$$x_s(t) + x_i(t) + x_r(t) = N \quad \forall t \geq 0.$$

- Time-evolution can be continuously approximated using susceptibles-infectives-removals (SIR) dynamics [Daley-Gani'99]: for times $t \geq 0$,

$$\dot{x}_s(t) = -\beta x_s(t)x_i(t)$$

$$\dot{x}_i(t) = \beta x_s(t)x_i(t) - \delta x_i(t) = [x_s(t) - \frac{\delta}{\beta}] \beta x_i(t)$$

$$\dot{x}_r(t) = \delta x_i(t)$$

with fixed parameters $\beta, \delta > 0$.

SIR model for a closed population (cont)

$$\dot{x}_s = -\beta x_s x_i; \quad \dot{x}_i = \beta x_s x_i - \delta x_i; \quad \dot{x}_r = \delta x_i.$$

- “Coagulation” term $x_s x_i$ models how the rate of contact between the infectives and susceptibles is an increasing function of the numbers of each.
- Thus, β would be larger for higher probability of infection given contact.
- The removal parameter δ represents the inverse of the lifetime of a peer as an infective.
- Note that summing the equations gives $\dot{x}_i(s) + \dot{x}_r(s) + \dot{x}_r(t) = 0$, consistent with a closed/fixed population.
- Also, $x_i(\infty) := \lim_{t \rightarrow \infty} x_i(t) = 0$.
- Finally, note assumed homogeneous population, *i.e.*, all nodes equally susceptible and proximal to each other for purposes of disease spread, and with common disease lifetime.
- $x_s(0) > \frac{\delta}{\beta} \Rightarrow$ infectives will grow initially, $\dot{x}_i(0) > 0$.

SIR model dynamics

- The limiting values of x_r, x_s depend on
 - $x_s(0)/N$ and the “relative” removal rate, δ/β [Daley-Gani’99],
 - particularly for the question: Will the disease infect all susceptibles, *i.e.*, will $x_r(\infty) = N$ and $x_s(\infty) = 0$?
- Combining the first and third equations to get $\dot{x}_i/x_i = -\dot{x}_r\delta/\beta$ gives $x_s(t) = x_s(0) \exp(-x_r(t)\beta/\delta)$.
- Thus, positive $x_r(\infty) < N$ can be numerically solved from $N - x_r(\infty) = x_s(\infty) = x_s(0) \exp(-x_r(\infty)\beta/\delta)$.
- So, only if $x_r \in \mathbb{R}$ exceeds $N - 1$ then the infection reaches the whole population.
- Note that the SIR model has only a few important parameters, thus it’s unlikely to “overfit” a given set of epidemiological data, a potentially significant problem of more complex models [Anderson-May’91, Mollison’95].

SIR model for an open population

- For exogenous arrivals at constant rate $\lambda > 0$ of susceptibles only:

$$\begin{aligned}\dot{x}_s(t) &= -\beta x_s(t)x_i(t) + \lambda \\ \dot{x}_i(t) &= \beta x_s(t)x_i(t) - \delta x_i(t)\end{aligned}$$

which we can write in vector form as $\underline{\dot{x}} = \underline{F}(\underline{x})$.

- The (stable) equilibrium point is $x_s^* = \delta/\beta$ and $x_i^* = \lambda/\delta$.

Inoculation (patching)

- Now consider removals due to vaccine uptake.
- Can rapidly inoculate susceptible machines by expedited dissemination of patches [Vojnovic'05] to remove vulnerabilities to a propagating attack, even so-called “white” worms.
- But patches that are not thoroughly tested prior to use can have serious side effects on the operation of the computer and the great variety of software it may be expected to support.
- Moreover, authentication overhead is required to ensure that a patch is legitimate and not itself another form of attack.

Inoculation model

- For simplicity, consider a homogeneous population with
 - initial infective population $x_i(0) > 0$,
 - initial susceptible population $x_s(0) > 0$, and
 - initial “removed” population $x_r(0) = 0$.
- Finally, at time t assume individuals decide whether to inoculate/patch with probability $p(x_s(t), x_i(t))$,
- *i.e.*, we are assuming that “global” information related to x_i and x_s is being at least approximately communicated to all susceptible peers, constituting the social dynamic here (no other inoculation coordination assumed).
- The inoculation uptake probability $p \in [0, 1]$ is naturally expected to be
 - a decreasing function of the probability and potential severity of side-effects of the inoculation/patch, and
 - an increasing function of the potential severity of disease once infected.

Inoculation uptake probability

- For $x_s > 0$, one can also expect p is a continuous, increasing function of the current proportion of the infective population

$$\frac{x_i}{x_i + x_s};$$

equivalently, as the proportion of infectives diminishes, there is less incentive to inoculate.

- Thus, $p(x_s, x_i)$ is increasing in x_i and decreasing in x_s .
- Moreover, since there is no incentive to inoculate when $x_i = 0$, we assume that for $x_s > 0$:

$$x_i = 0 \Leftrightarrow p(x_s, x_i) = 0.$$

- Such social dynamics of vaccine uptake can be modeled in greater detail as a game, and analyzed using Markov decision processes, *e.g.*, [Bauch'04].

Closed populations

- We have motivated the following SIR model:

$$\begin{aligned}\dot{x}_s &= -\beta x_s x_i - p(x_s, x_i) x_s, \\ \dot{x}_i &= \beta x_s x_i - \delta x_i, \\ \dot{x}_r &= p(x_s, x_i) x_s + \delta x_i,\end{aligned}$$

where the fixed positive constant

- β is the infection-rate parameter, and
 - δ is the infectives-removal parameter, *i.e.*, $1/\delta$ is the mean lifetime of an individual as an infective prior to removal by some means (death/crash or cure).
- For the unforced/closed stationary regime, *i.e.*, when \dot{x}_i and $\dot{x}_s \rightarrow 0$ as time $t \rightarrow \infty$, there are two cases.
 - if at equilibrium $x_i > 0$, then necessarily $x_s = \delta/\beta$ but then $\dot{x}_s < 0$.
 - Or if $x_i = 0$ then $p = 0$ so x_s, x_r can take on any values such that $x_s + x_r = N$.

Open population

- Now suppose that susceptibles arrive at constant rate λ .
- Assume the rate of departure (without return) by movement of individuals in all states x_i, x_s, x_r is governed by the same parameter $\mu > 0$.
- In this case, the SIR model is:

$$\begin{aligned}\dot{x}_s &= -\beta x_s x_i + \lambda - p(x_s, x_i)x_s - \mu x_s, \\ \dot{x}_i &= \beta x_s x_i - \delta x_i - \mu x_i, \\ \dot{x}_r &= p(x_s, x_i)x_s + \delta x_i - \mu x_r.\end{aligned}$$

- If we define $X \equiv x_s + x_i + x_r$, then $\dot{X} = \lambda - \mu X$ so that equilibrium $X = \lambda/\mu$.

Open population: equilibria

- For a constant-forced stationary regime:

$$\begin{aligned}0 &= -(\beta x_i + p(x_s, x_i) + \mu)x_s + \lambda, \\0 &= (\beta x_s - \delta - \mu)x_i.\end{aligned}$$

- If $x_i = 0$ then $x_s = \lambda/\mu$ (again, since $p(x_s, 0) = 0$ by assumption).
- Alternatively, if $x_s = (\delta + \mu)/\beta$ in then equilibrium $x_i > 0$ must satisfy:

$$x_i = \frac{\lambda}{\delta + \mu} - \frac{\mu}{\beta} - \frac{1}{\beta} p\left(\frac{\delta + \mu}{\beta}, x_i\right).$$

- Assuming $\frac{\lambda}{\delta + \mu} - \frac{\mu}{\beta} > 0$, an equilibrium solution $x_i > 0$ will exist since p is assumed continuous and monotonically increasing in x_i .

Modeling social networks as graphs

- Social ties between humans are complex, multi-faceted: profession, family, political leanings, ailments, race and ethnicity, religion, place of origin, education, hobbies, *etc.*
- By the mere existence of a one-dimensional social tie (or a measure of the strength of the tie), two humans are linked (weighted linked), giving a graph for each facet of social interaction.
- On-line, some users deliberately separate their communities of interest, so if all such graphs are “merged”, the result is *not* necessarily more fully connected for purposes of communication including epidemic spread and search.
- For the Facebook centralized social network: surfing either by clicking links and by Facebook search engine (as *www*).
- Recently, tremendous interest in random graphs with heavy tailed degree distributions (*e.g.*, small worlds, preferential attachment), representing social networks, *www*, *etc.*

Diseases that spread via social engineering

- Sometimes assumed that machines operated by peers that are socially proximal are likely to share vulnerability to the same infection ploy whether, *e.g.*, via:
 - hit-listing or socially shared software trojans (bond/edge percolation), particularly vulnerable super (well connected) peers, or
 - a compromised website or the social networking site itself (site percolation).
- That is, the disease will spread along social ties.
- *e.g.*, an attachment trojan or URL ploy of an email message more likely will work if there is a social connection between sender and receiver, otherwise it might simply be discarded as spam.

Discrete-time branching process models: No population limit (Galton-Watson)

- A time epoch $n \in \mathbb{Z}^+$, there are X_n infectives.
- Each infective will in the next epoch:
 - independently generate k infectives with probability p_k , for $k \geq 0$, and
 - will be removed.
- So $E(X_{n+1}|X_n) = \eta X_n$ where $\eta := \sum_{k \geq 0} k p_k$.
- Extinction probability $\varepsilon := P(\lim_{n \rightarrow \infty} X_n = 0)$ satisfies

$$\varepsilon = \sum_{k \geq 0} p_k \varepsilon^k.$$

- If $\eta < 1$ (resp. $\eta > 1$) then $\varepsilon = 1$ (resp. $\varepsilon < 1$).
 - Phase transition example [Eugster *et al.*, '04]:
if $p_0 = (1 - z)^2$, $p_1 = 2z(1 - z)$ and $p_2 = z^2$,
then $\varepsilon = (z^{-1} - 1)^2 \wedge 1$.
-

Discrete-time branching process model: limited population of N peers

- Infectives chosen at random from susceptible population.
- Infectives never removed (SI) vs removal after one epoch (infect and die SIR).
- For infect and die SIR with large N : with probability $1 - \varepsilon$, size of epidemic πN satisfies

$$\pi = 1 - e^{-\pi\eta}.$$

- Note that the solution π does not depend on N .
- Also note that $\forall \eta < \infty$, $\pi < 1$ as with continuous-time SIR model discussed previously.

Discrete-time epidemics on graphs

- Consider a discrete-time disease which spreads along the edges of the graph and infects vertices.
- The disease at a given single vertex at time zero.
- If a vertex v has the disease at time n , then at $n+1$: all or some (presumed susceptible) uninfected vertices at time n that are connected to v will be infected.
- Can show a branching process well approximates a discrete-time epidemic on certain random graph models, *e.g.*,
 - Erdos-Renyi, Sec. 2.2 of [Durrett'07], and
 - arbitrary degree distr'n, Sec. 3.1 of [Durrett'07].
- But the branching process approximation may be significantly erroneous,
 - owing to the increasing probability of re-infection or simultaneous infection attempts (loops) as no. infectives approaches no. vertices,
 - see [Zou'07] for a related model of email worms.

Discrete-time epidemics on graphs (cont)

- Consider a graph of N vertices, N large with mean vertex degree (fan-out) η .
- Say “atomic infection” probability, *i.e.*, prob. that an epidemic that reaches all N nodes, is equivalent to the probability that the graph is connected.
- For many random graph models with sufficiently high degree η (*e.g.*, Erdos-Renyi with $\eta = c + \log N$ for some constant c), can show for large N that atomic infection probability has a phase transition when $\eta / \log N = 1$.
- Assuming $\eta > \log N$, the *time* till atomic infection is at most logarithmic in N .
- The branching process approximation of [Newman *et al.*, '01] can be generalized to consider a random subset/thinning of edges through which disease spreads [Newman'02], Sec. 3.5 of [Durrett'07].
- Branching and graph-based epidemics can be approximated by simpler discrete-time SI/SIR/SIS ODEs, *e.g.*, SIS models of [Pastor-Satorras'01, Wilson'09] & Sec. 4.8 of [Durrett'07].

On-line diseases

- On-line diseases are based on malicious software (called called malcode or malware) with a reproductive strategy to spread among host computers.
- On-line diseases have many aspects similar to “natural” ones including:
 - certain types of malware can “mutate” to overcome defenses to their spread;
 - infected host machines can be cured by removal of malicious software (or not, resulting in endemic disease); and
 - a certain vulnerability can be patched (removed) thus, at least partially, inoculating/immunizing against future malware that attempts to exploit it.

On-line diseases: Potential speed of spread

- All computers on the Internet can communicate with each other with great speed.
- The spread of malware on the Internet can be orders of magnitude faster than the most contagious diseases in humans,
- even given that geographic hurdles to human disease spread have been overcome by modern methods of travel, particularly by airplane.

Speed of the Slammer worm

- In January 2003, the Slammer worm reached and infected 75,000 susceptible SQL servers around the world in less than 20 minutes, and this by random scanning in the IPv4 address space by the collective infectives [Moore'04].
- Slammer combined the exploit (the “scan” itself) and malware transmission into a single UDP packet.
- The enormous volume of scanning by Slammer infectives was conspicuous and had significantly disruptive “denial of service” effects on some routers handling the traffic of infected proximal machines.
- Instead of geographical obstacle to its spread, Slammer total scan rate was limited by the access bandwidths to the Internet of the infected domains/enterprises.
- Note: Slammer automatically spread in the network layer, not using any significant “social” associations.

Modeling specific Internet worms

- Propagation of the Blaster, Slammer and Witty [Kumar'05] worms:
 - congested network links thereby creating a temporary denial-of-access to the Internet for large population of end-hosts, and
 - resulted in a significant direct expenditure for patching and very significant aggregate loss of productivity.
- We now focus on “bandwidth-limited”, random UDP-scanning worms like Slammer and Witty that spread extremely rapidly in the wild.

Worm scanning traffic generation

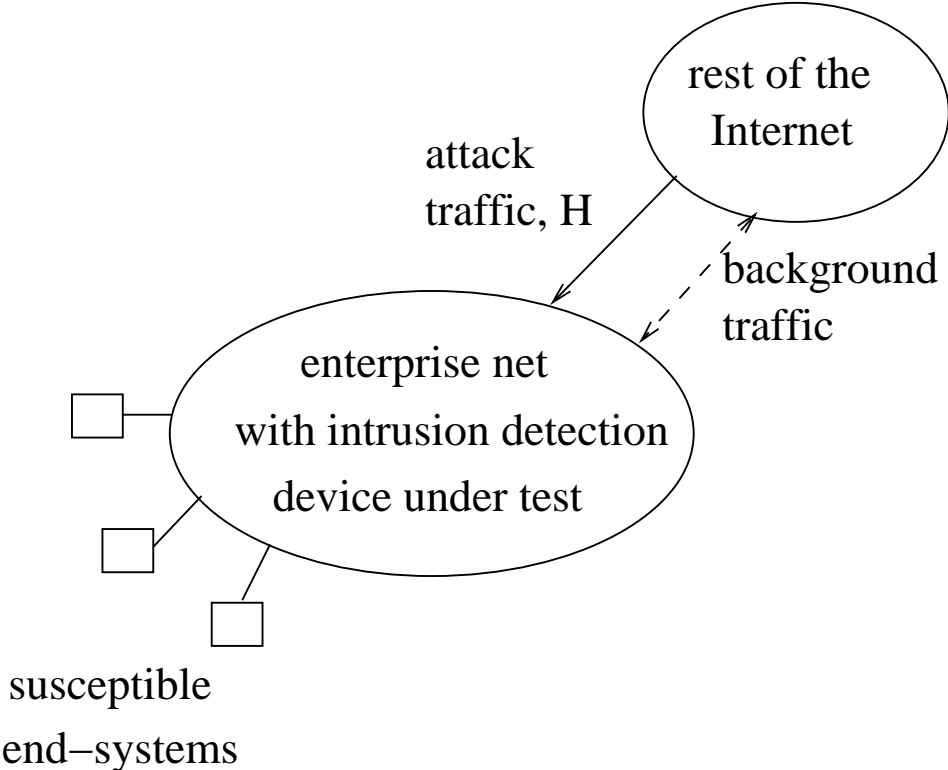
- Arguments have been made for worm defenses (detection and response) deployed in peripheral enterprise networks.
- So, need to realistically model the worm probing (scanning) activity *from* the Internet *to* the enterprise network under test.
- We assume that the scans generated from a given enterprise to the rest of the (much larger) Internet and the scanning activity directed at the enterprise from without are negligibly dependent.
- The scan-rate directed at the enterprise under simulation could be approximated as

$$\frac{A}{2^{32}}S(t)$$

where A is the size of its address space and $S(t)$ is the total (Internet-wide) instantaneous scan-rate of the worm at time t .

- Alternatively, a random thinning of S could be used.
-

Source of attack traffic from Internet

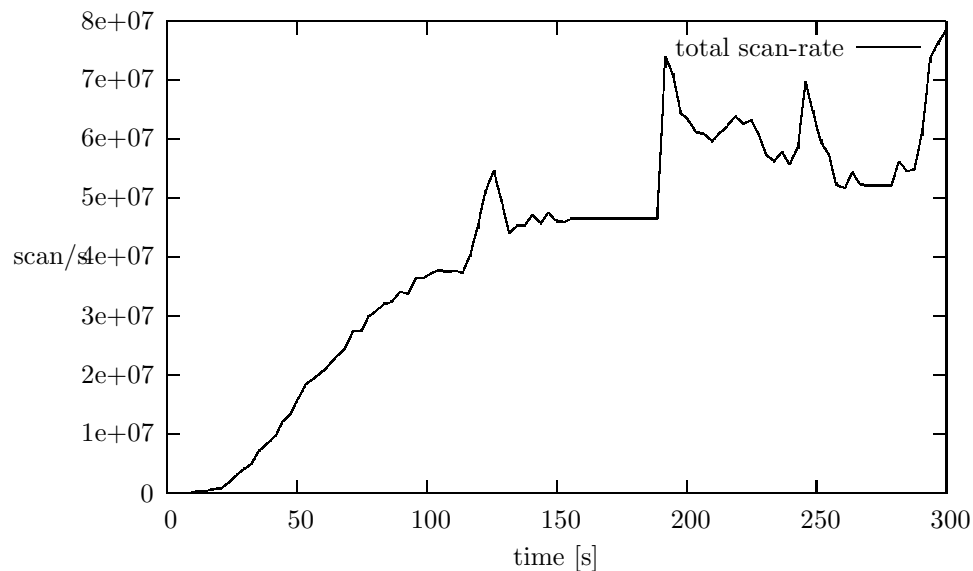


Worm scanning traffic generation (cont)

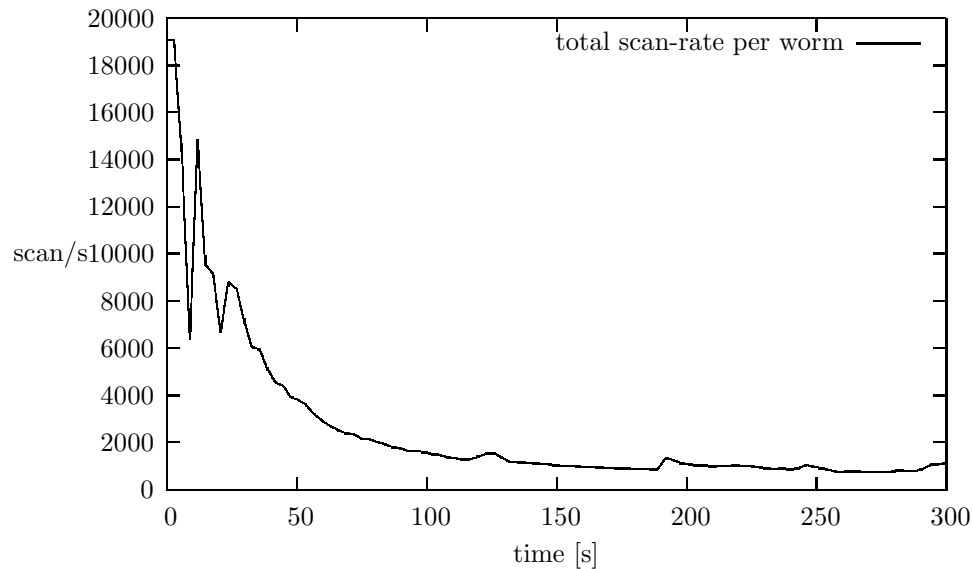
- The total scan-traffic generation S can be estimated from extrapolations of measured data for a particular worm *when this is available*, e.g., from a /8 “network telescope”.
- Alternatively, one could use a mathematical model whose parameters can be
 - fit to the salient data of a given worm (again, if that data is available) or
 - varied in an attempt to capture the behavior of actual worms for which measured Internet data is unavailable or set for hypothetical worms.
- A mathematical model also has insight and computational advantages over the potentially more accurate approach based on scale-down techniques [Weaver’04] and parallel simulation.
- Finally, a mathematical model allows for convenient study of hypothetical worms that are necessary to consider when evaluating defenses to be deployed.

Slammer's Internet spread

- The success of the simple SIR model for the Code Red worm has been demonstrated.
- Modeling Slammer and Witty is substantially more complex because network bandwidth limitations mitigated the spread of the worm.
- Beyond just spreading very quickly, Slammer was the first significant worm without a constant scanning rate:



Slammer's scan-rate per worm (infective)



- Note that the oscillations in the above curves are largely due to measurement error that is magnified by extrapolation.
- This graph demonstrates how Slammer was bandwidth limited.

Modeling Internet worms

- SIR models were used for Internet worms such as Code Red in [Staniford'02, Zou'02, Chen'03, Zou'03, Liljenstam'03].
- In [Kesidis'08], we showed how a stratified/generational SI model well fit the Slammer worm traces.
- Note that we are not considering countermeasures (inoculations or cures) or deaths: Slammer was a very rapidly spreading worm that was otherwise benign to its host (unlike Witty).

Simple SI model of scanning worms

- Suppose each infected host generates “contagion”, each targeted at a specific host, at a constant rate σ .
- Suppose that each contagion will select a single host (infected or otherwise) at random; the probability that a given host is selected is η .
- Therefore, over the interval of time $(t, t + dt)$, the expected change in the number of infected hosts, $dx_i(t)$ will be equal to
 - the total amount of contagion generated in the interval, $\sigma x_i(t)dt$,
 - times the expected amount of infection caused by a given scan, $\eta(N - x_i(t))$ (a small fractional quantity), *i.e.*, $\dot{x}_i(t) = \beta x_i(t)(N - x_i(t))$ where $\beta = \sigma\eta$.
- Recall [Daley-Gani'99]:

$$x_i(t) = \frac{x_i(0)N}{x_i(0) + (N - x_i(0)) \exp(-\beta Nt)}$$

where $x_s(t) + x_i(t) = N \forall t$ and $x_i(t) \rightarrow N$ as $t \rightarrow \infty$.

Homogeneous stratified SI model with instantaneously saturating links

- Assume the Internet core connecting peripheral enterprise networks only negligibly affects any scanning traffic they generate.
- Consider now a population of N enterprise networks (domains).
- For a homogeneous Internet model, assume each enterprise has the same number C of susceptible (SQL server) nodes.
- Each enterprise is in one of $C + 1$ states where state m connotes exactly m worms (infectives) for $0 \leq m \leq C$.
- For the *entire* network, define the state variables $y_m(t)$ representing the number of enterprises in state m at time t .
- Clearly, for all time $t \geq 0$, $\sum_{m=0}^C y_m(t) = N$ and $x_s \equiv y_0$.

Homogeneous model (cont)

- Define $Y(t) \equiv \sum_{m=1}^C y_m(t) = N - y_0(t)$ as the number of enterprises with one or more worms (infectives).
- Assume that each such infected enterprise transmits exactly σ scans/s into the Internet irrespective of the “degree” of its infection, *i.e.*, we assume that a single infective saturates the stub-link bandwidth of the enterprise.
- Finally, an implicit assumption of the following is that “local” infections (between nodes in the same enterprise) are negligible in number.
- Thus, the total rate of scanning (causing infection) into the Internet at time t is

$$S(t) = \sigma Y(t).$$

Homogeneous model (cont)

- The probability that a particular susceptible is infected by a scan is $\eta = 2^{-32}$ (purely random scanning in the 32-bit IPv4 address space).
- Thus, the probability that a scan causes an enterprise in state m at time t to transition to state $m+1$ is $(C-m)\eta$ because there are $C-m$ susceptible but not infected nodes in the enterprise at time t .
- Thus, define $\beta_m \equiv \sigma\eta(C-m)$.
- The y_m are governed by the following coupled SI equations: For times $t \geq 0$,

$$\begin{aligned}\dot{y}_C(t) &= \beta_{C-1}y_{C-1}(t)Y(t), \\ \dot{y}_m(t) &= (\beta_{m-1}y_{m-1}(t) - \beta_my_m(t))Y(t) \quad \text{for } 1 \leq m \leq C-1 \\ \dot{y}_0(t) &= -\beta_0y_0(t)Y(t).\end{aligned}$$

Homogeneous model (cont)

- The total number of worms at time t is clearly $\sum_{m=1}^C m y_m(t)$.

- Thus, the scan-rate per worm (per infective) is

$$\frac{\sigma Y(t)}{\sum_{m=1}^C m y_m(t)} = \frac{\sigma \sum_{m=1}^C y_m(t)}{\sum_{m=1}^C m y_m(t)}.$$

- Note that summing equations $m = 1$ to C yields the “standard” SI equation

$$dY/dt = \beta_0 y_0 Y = \beta_0 (N - Y) Y$$

whose solution is $Y(t) = Y(0)(1 + \exp(-\beta_0 N t))^{-1}$.

- The system has solution given by [Kesidis'08]:

$$y_m(t) = e^{-(C-m)s(t)} \sum_{j=0}^m \binom{C-j}{m-j} (1 - e^{-s(t)})^{m-j} y_j(0),$$

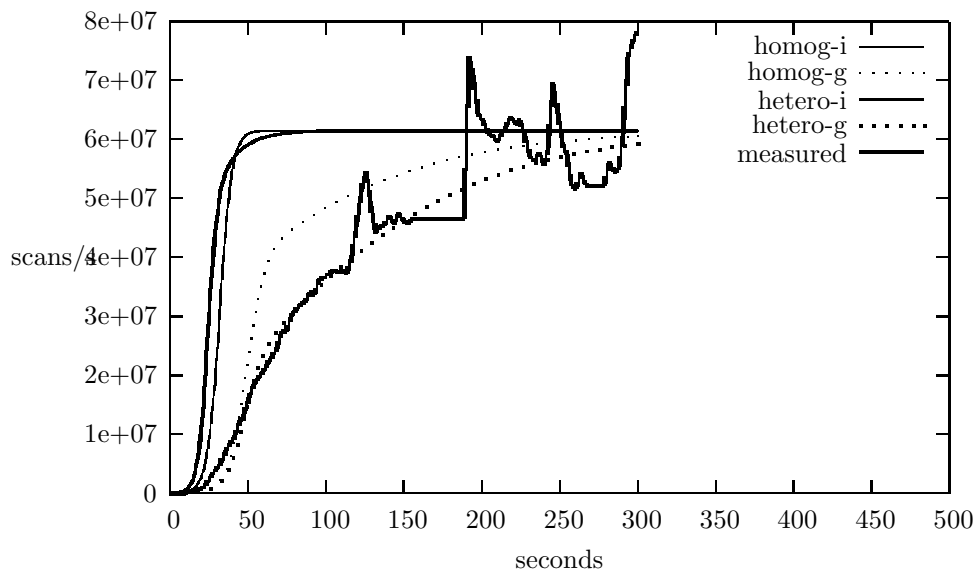
for $m = 0, 1, \dots, C$, where

$$e^{-s(t)} = (y_0(0) + (1 - y_0(0))e^{\beta C t})^{-\frac{1}{\beta}}.$$

Fitting to measured data

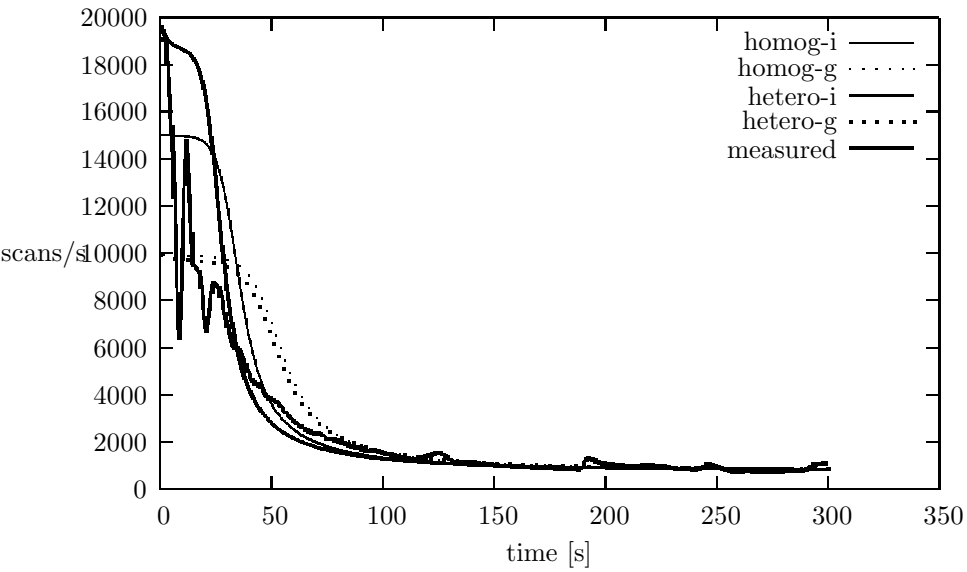
- We fitted just three parameters to measured data:
 - The initial value of scan-rate per worm: $\sigma = 15000$.
 - The ratio of initial to final value of scan-rate per worm: $C = 18$.
 - The final value of total instantaneous scan-rate, $NC\sigma$ (or simply from the total number of initially susceptible (and ultimately infected) end-systems, $NC = 73784$) giving $N = 4099$.
- Numerically solving starting from initial conditions $y_0(0) = N - 1$ and $y_1(0) = 1$ (*i.e.*, one initially infected server) yielded the following “homog-i” curves.
- This simple deterministic mathematical model of a homogeneous network with instantaneous link saturation yielded numerical results similar to those obtained by simulation of the “homogeneous clusters” model in [Weaver’04].

Total instantaneous scan-rate



- -g : gradual saturation model
- -i : instantaneous saturation model (with just one infective)

Scan-rate per worm (infective)



Heterogeneous network model with gradually saturating links

- Enterprise networks of class j :
 - have $C(j)$ susceptibles
 - have *maximum* scan-rate $\sigma_{j,C(j)}$
- $\sigma_{j,m}$ is the scan-rate to the Internet for class- j enterprise with $m \leq C(j)$ infectives where,
- under gradual link saturation, $\sigma_{j,m}$ is nondecreasing in m and $\sigma_{j,m} \leq i\sigma_{j,1}$ for all $m \geq 1$.
- $y_{j,m}(t)$ is the number of class- j enterprises at time t with m infectives
- $N(j)$ is the total number of class- j enterprises so that, for all times t ,

$$\sum_j N(j) = N \quad \text{and} \quad \sum_{m=0}^{C(j)} y_{j,m}(t) = N(j).$$

- The total instantaneous scan-rate is

$$S(t) \equiv \sum_{j,m} \sigma_{j,m} y_{j,m}(t).$$

Heterogeneous network model (cont)

- A more general set of coupled SI equations modeling worm spread than those used for a homogeneous network is as follows: For times $t \geq 0$ and all classes j :

$$\begin{aligned}\dot{y}_{j,C(j)}(t) &= \eta y_{j,C(j)-1}(t) S(t) \\ \dot{y}_{j,m}(t) &= \eta [(C(j) - m + 1) y_{j,m-1}(t) - (C(j) - m) y_{j,m}(t)] S(t) \\ \dot{y}_{j,0}(t) &= -\eta (C(j) - 1) y_{j,0}(t) S(t)\end{aligned}$$

where we recall $\eta = 2^{-32}$.

- The total number of infected end-systems (worms, infectives) is

$$\sum_{j,m} m y_{j,m}(t).$$

- The scan-rate per worm is the ratio of $S(t)$ to this quantity.

Fitting to data

- Fitting to data from the Slammer worm, we would require that the total number of susceptibles

$$\sum_j N(j)C(j) = 73782.$$

Fitting to the final value of the scan-rate per infective curve,

$$\frac{\sum_j N(j)\sigma_{j,C(j)}}{\sum_j N(j)C(j)} \approx 15000/18.$$

Fitting to the initial value of scan-rate per infective (“in mean”):

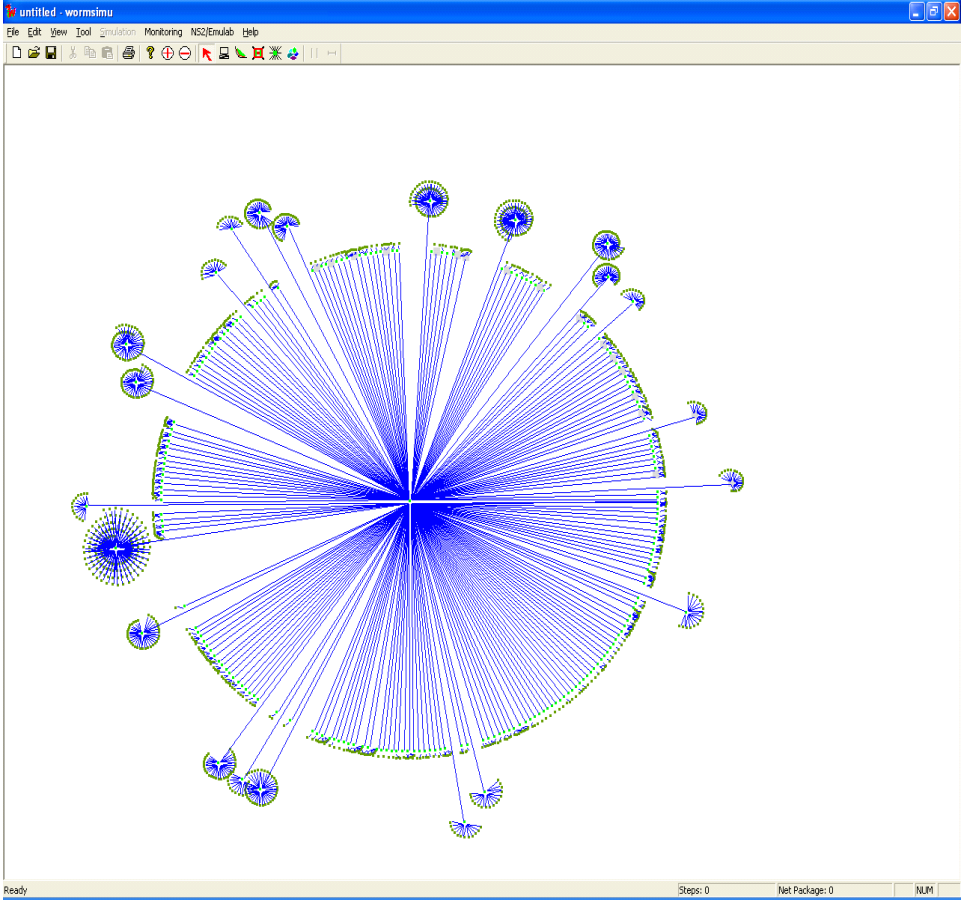
$$\frac{\sum_j N(j)\sigma_{j,1}}{\sum_j N(j)} \approx 15000.$$

- These three equations will determine three of the model parameters, leaving a number of parameters that can be used for potentially finer (and potentially over!) fitting to measured data.

Numerical examples with two classes

- We evaluated three other scenarios in the graphs above.
- Instantaneously saturating links with heterogeneous (2 class) enterprises: “hetero-i”.
- Gradual saturating links with homogeneous (1 class) enterprises: “homog-g”.
- Gradual saturating links with heterogeneous (2 class) enterprises: “hetero-g”.
- We found that varying the degree to which the stub-links gradually saturated had the most significant effects on the resulting total instantaneous scan-rate curves.
- See the previous two figures.

DETER/Emulab simulation set-up using Slammer's routeview data



Internet worm scanning strategies

- How can an infective y scan (network layer spreading) to determine the address of another computer x which is *likely* susceptible to y 's tactic of infection?
- One simple strategy is to
 - scan to addresses with the same most significant bits as the infective (*i.e.*, to proximal machines in the same domain),
 - under the assumption that if a vulnerability was found in one computer of a given domain, it's likely to exist in others,
 - and perhaps the local domain is simply more occupied than on average in the Internet.
- Vulnerability probings for infection attempts can also be conducted by vertically scanning (ports of a single machine) in addition to horizontally (address) scanning.

Monocultures and hit-listing

- It's well known that a particular automated exploit is often most effective for a "monoculture" of computers.
- *e.g.*, that use the same version of the same operating system and hence are all likely vulnerable to an exploit targeting it.
- The lack of operating system monocultures among cell phones was recently used to explain why viruses do *not* spread well through them [Wang'09].
- This factor can be roughly modeled by the β parameter of the SIR model.
- Malware can also harvest "hit lists" of potential victims stored on the host infective, *e.g.*, (application layer spreading) e-mail addresses or account identifiers on a social networking site visited by the host.
- Again, such "social" associations may indicate a greater propensity to be infectable by the same ploy.

References

- R.M. Anderson and R.M. May. *Infectious Disease of Humans: Dynamics and Control*. Oxford University Press, 1991.
- C.T. Bauch and D.J.D. Earn. Vaccination and the theory of games. *Proc. Nat'l Acad. Sci.*, 101:13391–13394, 2004.
- Z. Chen, L. Gao, and K. Kwiat. Modeling the spread of active worms. In *Proc. IEEE INFOCOM*, San Francisco, 2003.
- F. Comellas, M. Mitjana, and J.G. Peters. Epidemics in small-world communication networks. Technical Report SFU-CMPT-TR 2002-09, School of Computing Science, Simon Fraser University, Oct. 2002.
- Jahanian04 E. Cooke, M. Bailey, Z.M. Mao, D. Watson, F. Jahanian, and D. McPherson. Toward understanding distributed blackhole placement. In *Proc. ACM WORM, Washington, DC*, Oct. 29, 2004.
- D.J. Daley and J. Gani. *Epidemic Modeling: An Introduction*. Cambridge University Press, 1999.
- R. Durrett. *Random Graph Dynamics*. Cambridge University Press, New York, NY, 2007.
- P. Erdos and A. Renyi. The evolution of random graphs. *Magyar Tud. Akad. Mat. Kutato Int. Kozl.*, 5:17–61, 1960.
- P.T. Eugster, R. Guerraoui, A.-M. Kermarrec and L. Mas-soulie. From epidemics to distributed computing. *IEEE Computer*, **37**(5):p. 60–67, 2004.

References (cont)

- G. Kesidis, M. Vojnovic, I. Hamadeh, Y. Jin, and S. Jiwasurat. Model of the spread of randomly scanning internet worms that saturate access links. *ACM TOMACS*, May 2008.
- A. Kumar, V. Paxson, and N. Weaver. Exploiting underlying structure for detailed reconstruction of an internet-scale event. In *Proceedings of ACM IMC*, 2005.
- T. Kurtz. *Approximation of Population Processes*. SIAM, Philadelphia, 1981.
- M. Liljenstam, D. Nicol, V. Berk, and R. Gray. Simulating realistic network worm traffic for worm warning system design and testing. In *Proc. ACM WORM*, Washington, DC, 2003.
- D. Mollison. The structure of epidemic models. In *Epidemic Models: Their Structure and Relation to Data*, Cambridge, UK, 1995. Cambridge University Press.
- D. Moore, V. Paxson, S. Savage, C. Shannon, S. Staniford, and N. Weaver. Inside the Slammer worm. *IEEE Security and Privacy*, 2004.
- J. Nazario. *Defense and Detection Strategies Against Internet Worms*. Artech House, Norwood, MA, 2004.
- J. Nazario and T. Holz. As the net churns: Fast-flux botnet observations. In *Proc. 3rd Int'l Conf. on Malicious and Unwanted Software (MALWARE)*, 2008.

References (cont)

- M.E.J. Newman. Spread of epidemic disease on networks. *Phys. Rev. E*, 66, 2002.
- M.E.J. Newman, S.H. Strogatz, and D.J. Watts. Random graphs with arbitrary degree distributions and their applications. *Phys. Rev. E*, 64, 2001.
- R. Pastor-Satorras and A. Vespignani. Epidemic spreading in scale-free networks. *Phys. Rev. Lett.*, 86(14):3200–3203, 2001.
- Jamie Riden. Know your enemy: Fast-flux service networks. The HoneyNet project, 08/16/2008, Available at: <http://www.honeynet.org>
- S. Staniford, V. Paxson, and N. Weaver. How to own the Internet in your spare time. In *Proc. USENIX Security Symposium*, pages 149–167, Aug. 2002.
- K. Timm. IDS Evasion Techniques and Tactics. May 7, 2002. Available at <http://www.securityfocus.com/infocus/1577>
- M. Vojnovic and A. Ganesh. On the effectiveness of automatic patching. In *Proc. ACM Workshop on Rapid Malcode (WORM)*, pages 41–50, 2005.
- P. Wang, M.C. Gonzales, C.A. Hidalgo, and A.-L. Barabasi. Understanding the spreading patterns of mobile phone viruses. *Science*, Apr. 2009 (online).
- N. Weaver, I. Hamadeh, G. Kesidis, and V. Paxson. Preliminary results using scale-down to explore worm dynamics. In *Proc. ACM WORM, Washington, DC*, Oct. 2004.

References (cont)

- J.R. Wilson. Note on “influences of resource limitations and transmission costs on epidemic simulations and critical thresholds in scale-free networks”. to appear in *Simulation: Transactions of the Society for Modeling and Simulation International*, 2009.
- C.C. Zou, W. Gong, and D. Towsley. Code Red worm propagation modeling and analysis. In *Proc. 9th ACM Conference on Computer and Communication Security (CCS'02)*, Washington, DC, Nov. 2002.
- C.C. Zou, W. Gong, and D. Towsley. Worm propagation modeling and analysis under dynamic quarantine defense. In *ACM CCS Workshop on Rapid Malcode (WORM'03)*, Washington, DC, Oct. 2003.
- C.C. Zou, D. Towsley, and W. Gong. Modeling and simulation study of the propagation and defense of Internet email worm. *IEEE Trans. on Dependable and Secure Computing*, 4(2):105–118, April-June 2007.