

What to turn in: Submit hardcopy answers to the questions below. Please include your name, e-mail address, and the number of hours you spent working on the assignment.

1. Worms and Viruses: Propagation Models

Recall that the epidemic model of worm propagation developed in class yielded the following equation for determining the proportion of infected hosts at time t :

$$i(t) = \frac{e^{\beta(t-T)}}{1 + e^{\beta(t-T)}}$$

This model was based on the following definitions:

2^{32}	size of IP address space
N	size of the total vulnerable population
$I(t)$	infective/infected hosts at time t
β	Contact likelihood
$i(t) = I(t)/N$	proportion of infected population
T	Time to reach 50% infection rate

- a. Suppose a Code Red-like worm, which generates probe IP addresses uniformly at random, has a β value measured to be $\frac{1}{64}$ hosts/minute (roughly one host per hour). Assuming that there are 2^{19} vulnerable hosts (roughly half a million) on the Internet, calculate the number of probes per minute you expect to see from an instance of this worm.

From Lecture 15: A reasonable choice for β is $r * \frac{N}{2^{32}}$, where r is the probe rate:

$$r = \beta * \frac{2^{32}}{N} = \frac{1}{64} * \frac{2^{32}}{2^{19}} = \frac{2^{13}}{2^6} = 2^7 = 128 \text{ probes/minute}$$

- b. Suppose that a worm generates target IP addresses uniformly at random and that it attempts to infect $2^7 = 128$ such randomly chosen hosts per minute. If there are 2^{20} (roughly a million) vulnerable hosts on the Internet, what is a reasonable value for β according to the model?

$\beta = r * \frac{N}{2^{32}}$ yields an estimate of:

$$2^7 * 2^{20} / 2^{32} = 2^{27} / 2^{32} = 1/2^5 = 1/32$$

- c. Suppose that it was possible to create a “perfect” worm: such a worm would somehow (magically) choose a vulnerable victim at each probe and also (magically) coordinate all of the instances of the worms so that none of them try to infect the same target simultaneously. Suppose that the worm makes r probes per minute and that at time $t = 0$ there is one infected machine. Give a function $I(t)$ that indicates the number of infected hosts at time t as a function of r . Hints: (1) This model is unrelated to the formula described above. (2) Think about the number of infected hosts for small values of r (say, 1 and 2).

The key is to remember that each instance of the worm will make r copies of itself at each time step. This leads to an exponential growth in the number of worms:

For $r = 1$, we have: $I(0) = 1, I(1) = 2, I(2) = 4, I(3) = 8, I(4) = 16, \dots$

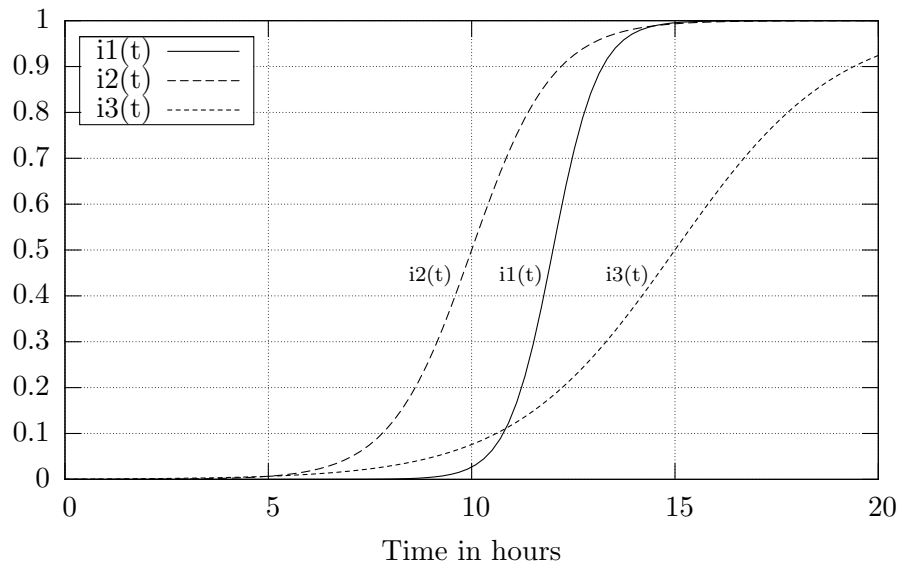


Figure 1: Infection models of three different Code Red-like worms.

For $r = 2$, we have: $I(0) = 1, I(1) = 3, I(2) = 9, I(3) = 27, I(4) = 81, \dots$

For $r = 3$, we have: $I(0) = 1, I(1) = 4, I(2) = 16, I(3) = 64, I(4) = 256, \dots$

The general pattern is thus given by:

$$I(t) = (r + 1)^t$$

- d. For this problem consider the graph in Figure 1 that depicts the infection models for three *unrelated* worms, Worm1: $i_1(t)$, Worm2: $i_2(t)$, and Worm3: $i_3(t)$. For each of the following questions, choose the appropriate answer:
- Which worm has the largest value of β ?

<input checked="" type="checkbox"/> Worm1	<input type="checkbox"/> Worm2	<input type="checkbox"/> Worm3	<input type="checkbox"/> Can't be determined
---	--------------------------------	--------------------------------	--
 - Which worm has the highest probe rate?

<input type="checkbox"/> Worm1	<input type="checkbox"/> Worm2	<input type="checkbox"/> Worm3	<input checked="" type="checkbox"/> Can't be determined
--------------------------------	--------------------------------	--------------------------------	---
 - Which worm will require the fastest response time to keep the fraction of the vulnerable population infected to less than 10%?

<input type="checkbox"/> Worm1	<input checked="" type="checkbox"/> Worm2	<input type="checkbox"/> Worm3	<input type="checkbox"/> Can't be determined
--------------------------------	---	--------------------------------	--
 - Which worm will infect the greatest number of hosts after 10 hours have elapsed?

<input type="checkbox"/> Worm1	<input type="checkbox"/> Worm2	<input type="checkbox"/> Worm3	<input checked="" type="checkbox"/> Can't be determined
--------------------------------	--------------------------------	--------------------------------	---
 - Which worm has the largest value for parameter T ?

<input type="checkbox"/> Worm1	<input type="checkbox"/> Worm2	<input checked="" type="checkbox"/> Worm3	<input type="checkbox"/> Can't be determined
--------------------------------	--------------------------------	---	--
 - Suppose that all three worms are modified to that they infect an initial hitlist consisting of 10% of their vulnerable populations instantaneously, and then behave as the original version of the worm. Which worm will achieve a 50% infection-rate the fastest?

- e. As part of your job at the National Internet Security Agency (NISA), you monitor the progress of a new worm called Blister whose behavior is described by the epidemic model above—it generates target addresses uniformly at random and propagates according to the infection function $i(t)$. At time $t = 180$ minutes Blister manages to infect 10,000 hosts (out of a possible 1,000,000 vulnerable hosts). From observing this worm in the wild, you are able to deduce parameters $\beta_{Blister}$ and $T_{Blister}$ that describe it accurately (for this question, their precise values don't matter).

A few weeks later, a new, more aggressive variant of Blister (Blister-v2) appears on the Internet. This version uses an initial hit list of 10,000 vulnerable hosts to start the spread of the worm much faster. After the initial hit list is infected, Blister-v2 spreads at random just like the original Blister worm. Assuming that infecting the hit list takes negligible time, what is the equation for $i(t)$ that models Blister-v2?

$$i(t) = \frac{e^{\beta_{Blister}(t+180-T_{Blister})}}{1 + e^{\beta_{Blister}(t+180-T_{Blister})}}$$

- f. If $T_{Blister} = 300$ minutes, sketch the graph of $I(t)$ for Blister-v2 that corresponds to your solution to part (b). Label the axes appropriately. Be as precise as you can given the information above, but you *do not* have to evaluate the function $I(t)$ a large number of times—instead, indicate the general shape of the curve and highlight any particular points of interest (e.g. the inflection point of the graph).

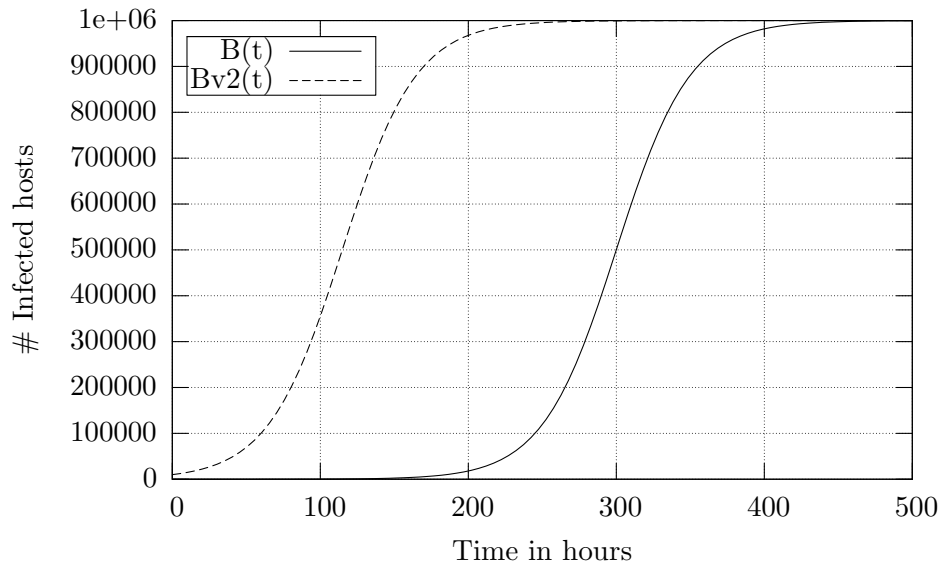


Figure 2: $I(t)$ for Blister-v2. The vertical line indicates the time $T = 120$ at which 1/2 of the vulnerable hosts are infected.

2. Worms and Viruses — Automatic Fingerprinting

Describe two assumptions about worm-generated network traffic that are used in the paper “Automatic Worm Fingerprinting” by Singh, *et al.* For each assumption, briefly describe a way that

a hacker could program a worm to violate the assumption, thereby making it harder to detect his worm.

- Invariant content: polymorphic worm could cause the copies it spawns to have different, randomly generated content.
- High frequency: the worm/virus could search slowly.
- Address dispersion: the worm/virus could search topologically to avoid random address dispersion.