

COS 116
The Computational Universe

Laboratory 8: Virus and Worm Propagation in Networks

You learned in lecture about computer viruses and worms. In this lab you will study virus propagation at the quantitative level. We can also study, to a certain extent, the spread of human disease since it has similar properties as computer viruses. You will first use a simple simulator application to gather data and then use Microsoft Excel to visualize it.

The application can model a network of interconnected computers similar to the Internet but with fewer (up to 100,000) computers. When you set all relevant parameters and click “Run Simulation,” it simulates the spread of a virus/worm in the network and outputs some numbers for you to analyze.

Hand in your lab report at the beginning of lecture on Tuesday, April 12.

Turn in only the Excel charts that are followed by a star (*). Do not turn in the charts without a star; they are just for you to visualize the simulation. Make sure each chart is labeled with the Experiment number and Question number, as well as any relevant additional information (e.g. particular options). You will not receive credit for charts that are unlabeled. Also, turn in answers to all questions posed in the body of the lab and in the “Additional Questions” section.

Introduction: Using the Simulator

1. Download this file to your Desktop:

http://www.cs.princeton.edu/courses/archive/spring10/cos116/lab8_files/vsim.exe

2. Double-click the file to run the simulator. Ignore any security warnings. The simulator has a variety of options to adjust. These options are explained below.

Note that the simulator uses randomness, especially when using the social network settings, so your results may not look exactly like your neighbor’s.

The image shows a software interface for a simulator, divided into two main sections: Population Settings and Virus Settings. Below these are simulation control options and a Run Simulation button.

Population Settings:

- Network Type: Fully-connected Network, Social Network
- Minimum Friends:
- Network Size:
- Initially Vulnerable: %
- Initially Infected:

Virus Settings:

- Time to Install: sec.
- Time to Spread: sec.
- Time to Repair: sec.

Simulation Control:

- Run Simulation For: secs.
- Report Status Every: secs.
-

Figure 1: Simulator interface

Here is a brief explanation of all adjustable parameters.

- **Network size:** The number of nodes in the network. This can be interpreted as the number of computers in a computer network or the number of people in a social network.
- **Network type:** In a fully-connected network, every node is directly connected to every other node. In a social network the connections are generated according to a randomized scheme so that the resultant network resembles actual social networks that have been studied. Though the details of the social network are interesting they are not relevant for this lab. You just need to know that each node is directly connected to at least the number specified in “Minimum Friends” though some nodes may be connected to much more than this minimum.
- **Initially vulnerable:** The percentage of computers that is initially vulnerable to infection. This models the possibility that some computers may already be patched against a weakness, or that some people are immune to a disease.
- **Initially infected population:** The number of computers that are initially infected.
- **Time to install:** The amount of time before an infected computer begins to spread the virus to other computers.
- **Time to spread:** The amount of time between infection attempts by an infected computer that is spreading the virus.
- **Time to repair:** The amount of time from infection until the virus is removed and the computer is immunized.

How infection works:

An infected computer spreads the virus by randomly choosing a computer to which it is directly connected, every "Time to Spread" seconds. If that computer is vulnerable, then it will begin spreading the virus "Time to Install" seconds later. However, the good guys may rush software patches to remove the virus/worm. The "Time to Repair" is the time after which the computer is considered repaired -- no longer spreading the virus/worm, nor susceptible to infection again.

Interpreting the results using Excel:

The simulator outputs periodic counts of vulnerable, infected, and repaired computers. You can save this output to a file of type CSV ("Comma Separated Values"), which is a standard format for data files. To graph the data from a CSV file in Excel:

1. Double-click the CSV file to open it in Excel.
2. Use the mouse to select all the columns and rows that have data (or, to do this with the keyboard, press Ctrl + A).
3. From the Insert menu, click the Scatter chart type "with only markers."
4. Select File ... Save As, and save the file as a *.xlsx file (not as a CSV file).

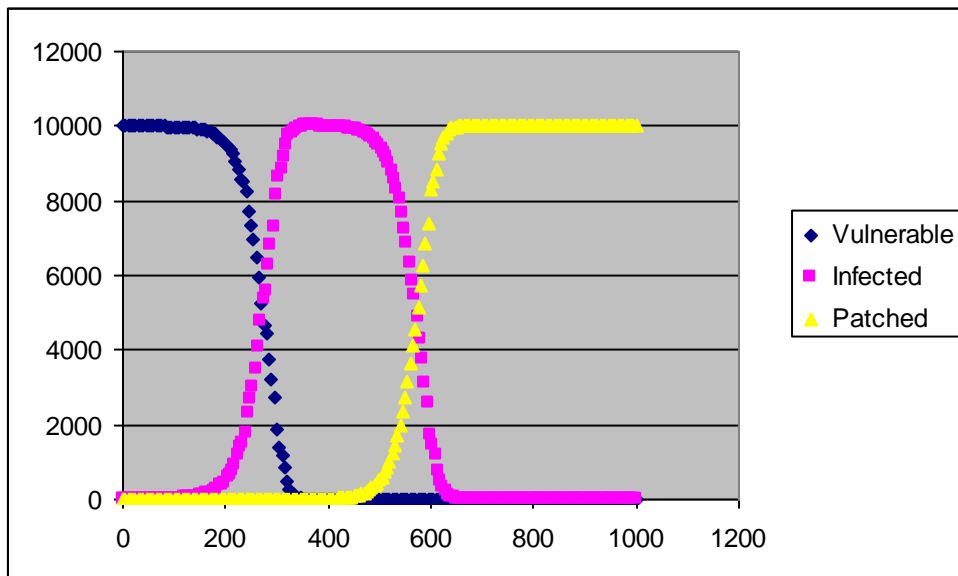


Figure 2: Example chart

Part 1: Worms vs. Email Viruses

You may wish to read the lecture notes to refresh your memory about the difference between viruses and worms. (Your TA is happy to answer questions.)

1. Run the simulator with the settings depicted in Figure 1 above (which are the default settings), except set “Time to Install” to 10 seconds. Plot the result in an Excel chart.
2. Now select “Social Network” and set “Minimum Friends” to 20. Also, change “Time to Install” to 150 seconds. Run the simulator, and plot the result in an Excel chart.
3. **Which of the previous simulations is a better model for the spread of worms, and which is a better model for the spread of email viruses? Explain your answer in terms of what you know about how worms and email viruses are propagated.**
4. **Report the total number of computers that were infected during each simulation.** To do this, look at the last line of output for the simulation, and note the number of infected and repaired computers. The total number of computers infected during the simulation is the sum of these two numbers.
5. The two totals you report in (4) should be similar to each other (if not, see the TA). Still, worms are considered more damaging than email viruses. **Explain why, based on the charts of the simulations.**

Part 2: Vulnerability of the Network

1. **Change back to the default settings (see Figure 1 above), and plot the result in an Excel chart*.**
2. Now change “Initially vulnerable” to 50%, and plot the result in an Excel chart. **Explain any difference between the two charts.**

Part 3: Infection Rate

1. Change back to the default settings, except set “Time to Spread” to 2 seconds. **Plot the result in an Excel chart*.**
2. Repeat Step 1 for “Time to Spread” values of 4, 6, 8 and 10 seconds. **Plot the results in separate Excel charts*.** Hint: It will probably be faster to generate all the data first, and then create the Excel charts. When saving the output to CSV files, use descriptive names so that you don’t forget what the output represents.

3. **Find the time point in each graph when the population becomes completely infected. How does this time point change as “Time to Spread” is increased? Why do you think this happens?**

Part 4: Disease Modeling

The simulator can also be used to model virus spreading in humans. When used this way, a “repaired” node represents a person who has either recovered from the illness or has been killed by it, so that he is no longer infecting others.

1. Change back to the default settings, except select “Social Network”, and set “Time to Install” to 10 seconds and “Time to Spread” to 10 seconds. Run the simulator, and plot the result in an Excel chart. This model will represent the spread of the flu.
2. Change “Time to Repair” to 10 seconds. Re-run the simulation, and plot the result in an Excel chart. This model will represent the spread of a more deadly virus, such as Ebola. (Note that in case of Ebola, people don’t get “repaired”; they die. But the net effect is still that they stop spreading disease any further.)
3. Now reduce “Time to Spread” for the deadly virus model and re-run the simulation until the total number of people infected is the same for both models (this may not be possible; just get as close as you can). **What does “Time to Spread” represent in our models? Why do you suppose that outbreaks of deadly viruses tend to be quite localized, even if they are highly contagious?**

Part 5: Additional Questions

1. **Which model is suited for studying the spread of the flu: fully connected network or social network? Write a ballpark estimate for “Minimum Friends” in this setting and briefly explain how you arrived at it.** Just guess -- no need for extensive Googling. (Hint: Whom does the person infect? Everybody she meets? Shakes hands with? Sneezes at? Some small subset of the above?)
2. **Explain briefly but clearly why the following measures are recommended during the cold season.** Use terminology from the lab, such as “Minimum Friends” and “Time to Spread.”
 - (i) Governments and employers should get a large fraction of the population vaccinated against the flu.
 - (ii) Cover your mouth while sneezing and frequently wash hands.

(iii) If you have a bad cold or flu, stay home and rest. (Note: Of course this is a good idea for the infected person, but why is it also a good public health measure?)

3. **By examining your data, why might an attacker prefer to write computer worms instead of viruses?**
4. **However, computer viruses are more common than worms. Why?** (Hint: The answer uses ideas you learned in class, such as social engineering, and the comparative difficulty of writing a virus versus writing a worm.)