Using Epidemiology to Model Computer Viruses

New Mexico

Supercomputing Challenge

Final Report

April 3, 2019

Team #27

Las Cruces High School

Team Members

Steven Fraga

Karson Million

Ben Widner

Breanna Widner

Teacher

Mrs. Lauren Curry

Project Mentor

Jesse Crawford

Executive Summary

We have created a model to simulate a computer virus using principles of epidemiology, the study of diseases and how they spread. We developed this program from scratch using the steps for an epidemiological simulation as a guide. The data we produced was produced based on the average network of an average household. This average network we used was based on research of the number of devices, cellphones, computers, and IoT devices; a household has on average. With this, we then ran our simulation on this network multiple times to find the effect of a computer virus attack on an average household. The resulting data from this was the cost, the number of devices infected, and the number of devices cured. We planned to check this data against data from historical attacks but were unable to complete that step this year. We plan to continue working on this project next year and to complete this step then.

Introduction

For this project, we chose to model how viruses spread by using existing epidemiology models and modifying this. We decided to use this as our project because of how prominent computer viruses have become recently. About 32% of the devices in the world are affected by some type of malware, and the cost of these attacks can be huge[2]. Several members of our group are interested in pursuing cyber-security as a career and suggested doing this subject for our project. We did lots of research on both epidemiology and computer viruses to find where they are similar and where they are incompatible. This research then shaped our approach to this project. We had to do additional research as we continued in our project and that research is explained as it appears later in this report.

We began our research by defining epidemiology, “Epidemiology is the study of the distribution and determinants of health-related states or events in specified populations, and the application of this study to the control of health problems”[3]. In simpler terms, epidemiology is the study of diseases. This is useful for us because we are studying the spread of computer viruses which spread much like human viruses through contact with other computers. Therefore, we can use the same strategies used in epidemiological models in modeling computer viruses. However, computer viruses have something vastly different as they are usually designed to get money so we must take this endgame into consideration in our model where we would not have to worry about this in an epidemiological model.

As we began to work on our project, we focused on the main parts of epidemiology which focuses mostly on the who, what, when, where, why, and how of a breakout of disease to figure out how to treat it and what it is likely to do next[3]. This is reflected in the basic features of an agent-based epidemiological model. These basic features are infected agents, susceptible agents, a method of distributing the virus, and some form of recovery or death[6]. As models get more complex, more elements are added on to this like immunity and more realistic movements etc.

The last thing we researched was computer viruses. A computer virus is defined as, “ a type of malicious code or program written to alter the way a computer operates and is designed to spread from one computer to another”[9]. These viruses can be introduced in many ways through websites, phishing, downloads, and connecting to other devices to name a few. This is problematic for many reasons because not everyone knows how to protect themselves from these viruses. This means people still open suspicious web pages, suspicious links in emails, download suspicious files, etc. Even if someone is aware of these things a person on their work network

might not be or their friend that connects to their wifi might not creating many pathways for viruses to enter their home network. Therefore, this simulation is important for households to plan financially for computer viruses and to show the importance of anti-virus software and constant vigilance to catch these viruses before they spread too far.

Description

We began this project with research into computer viruses and epidemiology so we could fully understand where they intersected and how best to build our model. Our initial plan was to use an epidemiology model Breanna had made in a computer science class but after talking to our mentor decided to start from scratch. The basic features of an epidemiology model are as noted above in the introduction. Using these as a guide we broke down our program into methods to make it easier to code. We use sliders to allow for different devices on a network since not all devices connect the same. Most people now link their cellphones to their computers and many IoT devices link to cellphones hence our decision to create these different breeds. We set labels to ensure these devices can be recognized. We also have a slider for the probability of finding(and removing) the virus since some viruses hide well while others are more easily recognizable.

The program works by selecting a device and infecting it. This introduces the virus into our network. We use red to represent the infected devices, green to represent the “cured” devices (devices which have had the virus removed), and grey to represent all other devices. The go procedure works as follows. First, it spreads the virus by iterating through all infected turtles and selects one of its neighbors to try and spread the virus to. Since on average devices have about an 85% vulnerability rate[5, 7] we choose a random number between 1 and 100, and if it is less than

85, we infect the device. After the virus is spread we then go through and cure a random portion of the infected devices. We do this by again taking a random number between 1 and 100 and check if it is less than the chance of finding the virus set by our slider. After all of this, we calculate the cost of this iteration and add it to the current total cost. The cost is calculated using this equation cost = (infected * 141) + (cured * 150). We got this equation because on average viruses cost $141 in loss of data[8] and removing viruses costs about $150 when done by Geek Squad a popular technology service[4].

We track the number of infected devices, number of cured devices and cost. To get a picture of the average cost of a virus infestation in the average home network, we found the average number of devices per house. Since most houses only have one router we set our number of routers slider to 1. Most houses have 15 IoT devices, so we set that slider to 15[10]. There are also on average two cellphones per household and one computer, so we set those sliders accordingly[1]. While we decided to use these averages for our trials this simulation is designed to work for any household by adjusting the sliders to fit the network you want to model.

We wanted to verify and calibrate our simulation using known data but ran out of time to do this so the current simulation may be inaccurate. We plan to continue this project next year to continue improving it.

Results

Runs	Cost	Total Infected	Total Cured
1	1746	6	6
2	6984	24	24
3	4656	16	16
4	582	2	2
5	3492	12	12
6	1164	4	4
7	3492	12	12
8	14841	51	51
9	582	2	2

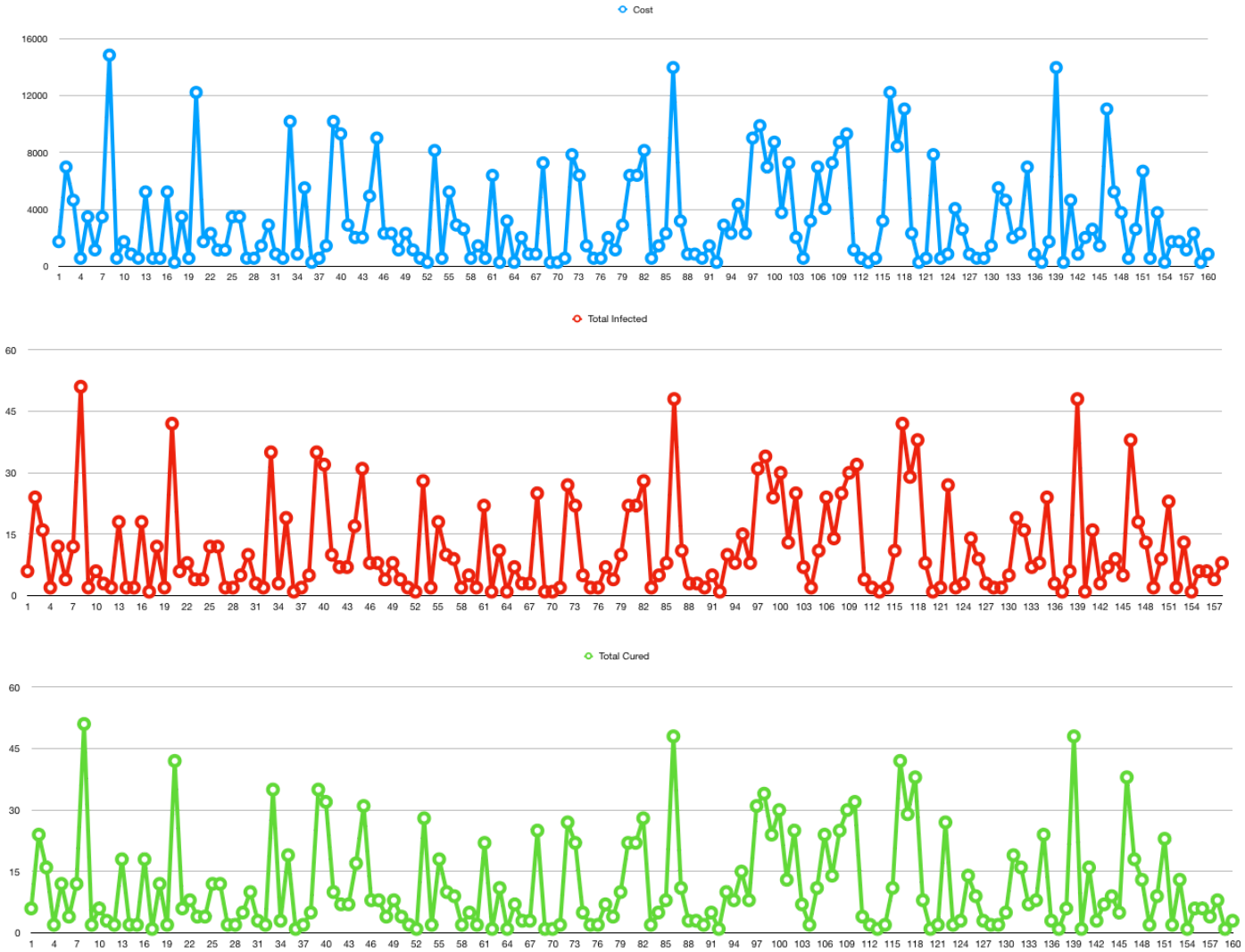
10	1746	6	6
11	873	3	3
12	582	2	2
13	5238	18	18
14	582	2	2
15	582	2	2
16	5238	18	18
17	291	1	1
18	3492	12	12
19	582	2	2
20	12222	42	42
21	1746	6	6
22	2328	8	8
23	1164	4	4
24	1164	4	4
25	3492	12	12
26	3492	12	12
27	582	2	2
28	582	2	2
29	1455	5	5
30	2910	10	10
31	873	3	3
32	582	2	2
33	10185	35	35
34	873	3	3
35	5529	19	19

36	291	1	1
37	582	2	2
38	1455	5	5
39	10185	35	35
40	9312	32	32
41	2910	10	10
42	2037	7	7
43	2037	7	7
44	4947	17	17
45	9021	31	31
46	2328	8	8
47	2328	8	8
48	1164	4	4
49	2328	8	8
50	1164	4	4
51	582	2	2
52	291	1	1
53	8148	28	28
54	582	2	2
55	5238	18	18
56	2910	10	10
57	2619	9	9
58	582	2	2
59	1455	5	5
60	582	2	2
61	6402	22	22

62	291	1	1
63	3201	11	11
64	291	1	1
65	2037	7	7
66	873	3	3
67	873	3	3
68	7275	25	25
69	291	1	1
70	291	1	1
71	582	2	2
72	7857	27	27
73	6402	22	22
74	1455	5	5
75	582	2	2
76	582	2	2
77	2037	7	7
78	1164	4	4
79	2910	10	10
80	6402	22	22
81	6402	22	22
82	8148	28	28
83	582	2	2
84	1455	5	5
85	2328	8	8
86	13968	48	48
87	3201	11	11

88	873	3	3
89	873	3	3
90	582	2	2
91	1455	5	5
92	291	1	1
93	2910	10	10
94	2328	8	8
95	4365	15	15
96	2328	8	8
97	9021	31	31
98	9894	34	34
99	6984	24	24
100	8730	30	30
101	3783	13	13
102	7275	25	25
103	2037	7	7
104	582	2	2
105	3201	11	11
106	6984	24	24
107	4074	14	14
108	7275	25	25
109	8730	30	30
110	9312	32	32
111	1164	4	4
112	582	2	2
113	291	1	1
114	582	2	2
115	3201	11	11
116	12222	42	42
117	8439	29	29
118	11058	38	38
119	2328	8	8
120	291	1	1
121	582	2	2
122	7857	27	27
123	582	2	2
124	873	3	3
125	4074	14	14
126	2619	9	9
127	873	3	3
128	582	2	2
129	582	2	2
130	1455	5	5
131	5529	19	19
132	4656	16	16
133	2037	7	7
134	2328	8	8
135	6984	24	24
136	873	3	3
137	291	1	1
138	1746	6	6
139	13968	48	48
140	291	1	1
141	4656	16	16
142	873	3	3
143	2037	7	7
144	2619	9	9
145	1455	5	5
146	11058	38	38
147	5238	18	18
148	3783	13	13
149	582	2	2
150	2619	9	9
151	6693	23	23
152	582	2	2
153	3783	13	13
154	291	1	1
155	1746	6	6
156	1746	6	6
157	1164	4	4
158	2328	8	8
159	291	1	1
160	873	3	3
Average	**3288.3**	**11.3**	**11.3**

This table is the data from all of our trials run in behavior space. We then took the average of the data collected to get a general picture of what an average attack would look like. As as shown above, the average computer virus attack infects 11 devices and costs about $3,288.30.



Here we have the graphs of the data presented in the table, and clearly, the values vary widely. This is because each trial acts a little differently. In some trials the virus is discovered quickly while in others it spreads before it is discovered. This will change as the likelihood of finding and removing a virus changes which we set at 50% for these trials. Also, all three graphs look the same because of two factors. First, in our simulation, cost depends directly on the number of

infected and cured agents. Also since our simulation has no death condition, all infected agents were subsequently cured. These are aspects we will take into account as we continue the project next year.

Conclusions

Looking at the data our simulation predicts a cost of $3,288.30 for a virus attack on an average household. This result needs to be checked with data from historical attacks; hence our continuation of this project next year. We can also see the average number of devices that were infected and cured. Eleven devices were infected, and eleven were cured.

Recommendations

This model while being complete can still be improved as we were not able to verify our results with historical attacks. We plan to do this as we continue the project next year.

Acknowledgments

We would like to recognize our mentor for providing us with much-needed advice on our program. We would also like to thank our parents for supporting us through this process. Lastly, we would like to especially recognize and thank our teacher for helping us through this process and just being amazing.

Bibliography

[1]A third of U.S. households have three or more smartphones. (2017, May 25). Retrieved from https://www.pewresearch.org/fact-tank/2017/05/25/a-third-of-americans-live-in-a-household-with-three-or-more-smartphones/

[2]Gaille, B. (2017, June 01). 37 Shocking Computer Virus Statistics. Retrieved from https://brandongaille.com/36-shocking-computer-virus-statistics/

[3]C. (2012, May 18). Lesson 1: Introduction to Epidemiology. Retrieved from https://www.cdc.gov/ophss/csels/dsepd/ss1978/lesson1/section1.html

[4]Jenkins, D., & Porter, J. (2019, February 01). 2018 Geek Squad Prices, Rates, Services & Alternatives. Retrieved from https://fitsmallbusiness.com/geek-squad-prices/

[5]Muoio, D. (2015, October 19). There's a good chance your Android phone is at risk. Retrieved from https://www.businessinsider.com/87-percent-of-android-devices-vulnerable-to-security-attacks-2015-10

[6]Perez, L., & Dragicevic, S. (2009, August 05). An agent-based approach for modeling dynamics of contagious disease spread. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2729742/

[7]Pusok, K. (2018, October 02). Opinion: Most Wi-Fi routers vulnerable to crime. Retrieved from

https://www.newsday.com/opinion/commentary/majority-of-wi-fi-routers-in-the-u-s-are-vulnerable-to-cybercrime-1.21370823

[8]Sobers, R. (2019, March 28). 60 Must-Know Cybersecurity Statistics for 2019. Retrieved from https://www.varonis.com/blog/cybersecurity-statistics/

[9]What is a computer virus? (n.d.). Retrieved from https://us.norton.com/internetsecurity-malware-what-is-a-computer-virus.html

[10]Willett, J. (n.d.). Minim's 3 IoT predictions for 2019. Retrieved from https://www.minim.co/blog/minims-iot-predictions-for-2019