Online Promiscuity: Prophylactic Patching and the Spread of Computer Transmitted Infections

Timothy Kelley
Indiana University Bloomington
107 S. Indiana Ave.
Bloomington, IN, 47405
kelleyt@indiana.edu

L. Jean Camp
Indiana University Bloomington
107 S. Indiana Ave.
Bloomington, IN, 47405
ljcamp@indiana.edu

ABSTRACT

There is a long history of studying the epidemiology of computer malware. Much of this work has focused on the behaviors of specific viruses, worms, or botnets. In contrast, we seek to utilize an extension of the simple SIS model to examine the efficacy of various aggregate patching and recovery behaviors. We use the SIS model because we are interested in global prevalence of malware, rather than the dynamics, such as recovery, covered in previous work. We consider four populations: vigilant and non-vigilant with infected or not for both sets. We show, using our model and a real world data set, that small increases in patch rates and recovery speed are the most effective approaches to reduce system wide vulnerabilities due to unprotected computers. Our results illustrate that a public health approach may be feasible, as what is required is that a subpopulation adopt prophylactic actions rather than near-universal immunization.

1. INTRODUCTION

Studying the spread of computer malware through the use of epidemiological models has been a useful tool in understanding the dynamics of individual outbreaks of malware, as well as giving some insight into possible mitigation policies. Kephart and White's early work focusing on system-wide prevalence examined effects of topology on virus spread as well the possibility of a social response to infection [18, 19]. Other work has focused on describing the dynamics of individual types of viruses, worms, or botnets.

In Kephart and White's examination of the social response, even a small social response was able to reduce significantly the total level of infection in the system. However, this result depends on a system where the recovered population could not become infected. For this simulation we wanted to examine the effects of social response when it led to recovery, but this recovery did not protect the user from reinfection.

We use the results from these individual models, as well as larger data on websites hosting phishing sites to model system-wide properties of malware spread. We use these system-wide properties to draw analogies from public health research regarding the spread of sexually transmitted infections (STIs) to examine organizational patching policies. From these results, we argue that thinking of security problems in terms of public health policy is a good addition to more traditional mental models of security.

Workshop on the Economics of Information Security (WEIS) 2012, June 25-26, Berlin, Germany

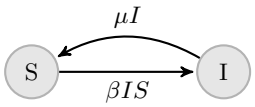

Figure 1: Permitted transitions in the SIS model.

2. BACKGROUND AND RELATED WORK

In the early work in adapting epidemiological models to computer viruses, the local nature of data transfer had to be taken into account. Computer viruses, in general, were spread very locally, and certain assumptions such as homogeneous population and the probability that an infected individual could infect any other individual in the susceptible population, did not hold [18]. In this environment Kephart and White (KW) adapted the SIS (Figure 1) model to a directed graph, to account for the non-homogeneous behavior of program sharing.

In their model, each computer is a vertex in a graph and an arc connects another computer in a program-exchange relationship. The arcs are associated with individual rates of infection and represent the set of vertices that can be infected by a given vertex, while each vertex is given an individual rate of recovery. Once a vertex has recovered, it is immediately capable of being reinfected. This, as the authors state, represents a very simple assumption that users will not become more vigilant after being infected. While this is a simple assumption, it seems to be a fairly good approximation for real world data [29].

Their deterministic calculations correspond to early results in prevalence driven epidemiological models [20], but failed to capture the social or organizational aspects of dealing with virus spreads. They modified their model to include a social response, or, as they call it, a kill switch model. That is, each computer, upon discovery and cleaning, alerts all

other computers it is connected to to alert them of possible infection [19].

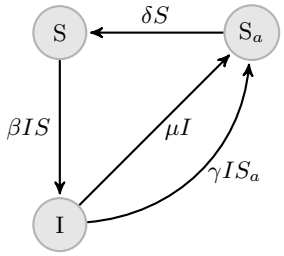

Figure 2: SIS model with Recovery and Social Response.

This extension (Figure 2) assumes that recovery corresponds to a temporary immunization from the virus [19]. Based on these model extensions, KW show that central reporting and response to an incident is important to containing the incident. With central reporting and response, even if an organization is above the epidemic threshold, an incident can be limited in size and duration [19].

Other early work in modeling computer worms and viruses used fairly standard epidemiological models, in particular, variations of the SIR model. They overcame the limitations of the well mixed assumption by incorporating the scanning behavior found in many worms by expanding on Kephart and White's work on the effects of topology on virus spread. Knight, Elder, and Wang, analyzed networks in hierarchical and cluster topologies to study the effects of immunization from viruses in theoretical email networks [40]. Newman, Forrest, and Balthrop expanded Knight *et.al.*'s work by incorporating actual email network data and studies of network structure from the realm of statistical physics [30]. Both Knight *et.al.* and Newman *et.al.* demonstrated that targeted immunization could have a drastic effect on the spreading of viruses spread by emails.

Newman *et.al.* draws on Albert, Jeong, and Barabási's work on describing the network topology of the Internet [8, 42], as well as Pastor-Satorras and Vespignani's work on the effects of that topology on the dynamics of epidemic models [31, 32]. Albert, Jeong, and Barabási's work demonstrated the scale-free topology of the World Wide Web [8], but also the difficulty in generating models that reflected the true topology [42]. Pastor-Satorras and Vespignani, on the other hand, demonstrated that a scale-free topology could lead to the possibility of infinite duration, though low-level, prevalence of a given epidemic spread [32, 31].

Zou *et.al.*'s work on modeling the Code-Red worm using the description and data provided by Moore *et.al.* modified the standard SIS model by incorporating a variation of Kephart and White's social response model, incorporating scanning rate, and allowing for infection rates to fluctuate in time [44, 28]. Including the social response in their model allowed them to take into account human responses to the onset of an infection [44].

Zou, Gong, and Towsley, also included a model that allowed systems to become quarantined, removing them from the susceptible and infectious populations [45]. They demonstrated that removing computers from both populations for some amount time was an effective mitigating factor [45]. However, as Serazzi and Zanero point out in their later work on Sapphire, quarantines would be difficult to implement, as infected hosts cannot be trusted to quarantine themselves [36]. Zou and Towsley revisited their earlier work to demonstrate that the increased range of addresses in IPv6 would effectively reduce the total prevalence of routing worms such as Sapphire. This, they show, is due to scanning worms inability to access significant parts of the IPv6 address space in a reasonable amount of time [43].

Moore *et.al.*'s data collection and description of the explosive growth of the Sapphire worm required further modifications to earlier models [27]. While Code-Red generally followed standard models, Sapphire spread fast enough to become bandwidth limited, which in turn, limited its total ability to spread [27]. Serazzi and Zanero designed a model that encoded network resources. Utilizing incoming and outgoing traffic rates into their model, they were able to capture the Sapphire's aggressive scanning. This scanning choked the Internet and greatly impeded Sapphire's rate of growth [36]. Serazzi and Zanero also point out the difficulty in implementing global security policies such as quarantines and hub immunizations.

Staniford, Paxson, and Weaver, contribute an excellent summary of many of the modeling attempts and call for a CDC for computer malware [39]. We agree with this model of thinking, and the data collected via their suggested sensors and analysis would be useful for further mitigation of online pathogens. However, the focus of this paper is more on the effects of risk takers on the total prevalence of contagion. Thus, we hope to show that a small group of users engaged in risky behavior creates a threat to the risk adverse population.

To this end we look primarily at August and Tunca's work on allowing users with illegal copies of software to patch [5] and Choi, Fershtman, and Gandal's work on cost of patching [14]. While August and Tunca focus primarily on whether or not firms should allow users of illegal copies to patch, Choi, Fershtman, and Gandal look at the costs associated with different user's and their willingness to patch. We combine both the pirates in August and Tunca's work, with the non-patching populations of Choi, Fershtman, and Gandal, to show that limitations on user's ability to maintain a secure system is dangerous to the risk adverse population.

Models of sexually transmitted diseases have become very complicated to deal the the multiple population interactions [12, 17]. However, most multiple population models do not couple the behavioral changes that occur do individual's perceptions of disease spread [33]. We build off of Perra *et al.*'s work to create a two population model with a social response that represents the ability of users to change behavior, and thus, their population group. This differentiates our model from more complicated models of STIs that use different characteristics of infection for individual population groups, but do not include behavioral responses to infection [11, 34, 37, 6].

3. METHODOLOGY

We first develop a simple model based on Kephart and White's initial social response model and Wang *et. al.*'s user vigilance model. We then use our model to examine the long term global prevalence of malware. Then we analyze the various parameters within this model to identify which ones are most effective at controlling systemic infection. We utilize anonymized data on websites used to support phishing attacks provided by Clayton and Moore, under a NDA, to demonstrate that the model can capture observed malicious behavior. We also attempt to answer questions about feasible responses to malware diffusion that could result in reduction in botnet prevalence.

3.1 Model Creation

Kephart and White's social response model (KW) demonstrates the effectiveness of social responses to computer infection. We extend their model to allow the possibility of infection in the inoculated population. This extension includes aspects of Wang *et. al.*'s vigilance model [41]. Similar to their approach, we view user vigilance as prevalence based response to the infectious population, with vigilant users returning to the more susceptible population at a constant rate.

Wang *et. al.* assume that user vigilance declines after an initial peak due to responses to new infections [41]. Rather than using delay differential equations to model the decrease in vigilance, we make vigilance reliant on the infectious population, and separate vigilant and non-vigilant users into different compartments. Users return to the non-vigilant population at a constant rate, which we feel represents the effects of cost incurred due to maintenance of a secure system. Another option would be to have the return rate inversely proportional to the infected population: The larger the infectious population, the slower vigilant users return to the non-vigilant population.

Since we are modeling the diffusion of malware, the individual behaviors are of limited use. We are more interested in equillibrium states. Thus, the more common SIR type models previously used for models of malware infection are not useful for our purposes. This assumption better captures the long-term behavior seen in malware such as the Blaster worm [7] as well as the persistent insecurities found in web servers that allow them to be reinfected [29].

$$\begin{aligned}
\frac{dS_r}{dt} &= -(\beta_r(I_r + I_a)S_r) - (\eta(I_r + I_a)S_a) + \delta S_a + \mu_r I_r \\
\frac{dI_r}{dt} &= \beta_r(I_r + I_a)S_r - \mu_r I_r - \mu_{a_1} I_r - \gamma_{a_1} I_r S_a \\
\frac{dS_a}{dt} &= -\beta_a(I_r + I_a)S_a - \delta S_a + \mu_{a_1} I_r + \mu p_2 I_a + \\
&\quad \gamma_{a_1} I_r S_a + \gamma_{a_2} I_a S_a + \eta(I_r + I_a)S_a \\
\frac{dI_a}{dt} &= \beta_a(I_r + I_a)S_a - \mu p_2 I_a - \gamma_{a_2} I_a S_a
\end{aligned} \tag{1}$$

The model we propose (Figure 3 and Equation 1) is a modified version of an SIS model with two interacting subpopulations. Our model does not assume immunity in the S_a population. This represents the fact that no security system is 100% effective at stopping all vulnerabilities. We do not find, in the course of our analysis, that the rates of infection in the resistant population are so low that they may be ignored.

Our model also assumes a well-mixed, homogeneous population. This is, in many ways, an unrealistic assumption, given the patterns of connection displayed by social networks and browsing behavior [26]. Moreover, it distracts from our metaphor of STIs, in that it assumes that all users are equally likely to interact with one another, rather than rely on contact patterns [15]. However, the dissemination of many online attacks is based on random scanning, which creates a scaled version of a well-mixed, homogeneous population [18]. Thus, this is a useful simplifying assumption, but can be expanded upon in future work.

3.1.1 Parameter Definitions

Table 3.1.1 briefly summarizes the various symbols we use in our model and analysis, which we describe here. S_r represents the susceptible population of non-vigilant users. These are systems that do not have a form of malware and can be infected. S_a represents susceptible systems within vigilant users. When an S_r or S_a system is infected, it transitions to the infected populations I_r or I_a, respectively.

η and δ govern the transitions between the two population groups. η represents the response of non-vigilant users to a given level of global infection. The higher η is, the faster non-vigilant users secure their systems. δ governs the response to the cost of maintaining a secure system. This is a constant rate, and the higher δ is, the less accepting users become of the cost, driving them to become insecure at a faster rate.

β_r, μ_r, β_a, and μ_{a_2} are the infection spread parameters for the non-vigilant and vigilant populations, respectively. β_r and β_a govern how fast an infection spreads, while μ_r and μ_{a_2} dictate how quickly a user recovers. Recovery could be a simple as deleting an infected file, or as complex as reinstalling an OS. We assume that $\beta_r > \beta_a$ and $\mu_r < \mu_{a_2}$ to represent the fact that users that are maintaining a secure system will be less likely to become infected and more likely to recover.

γ_{a_1} and γ_{a_2} embed the response to social pressure to recover in the non-vigilant and vigilant populations, respectively. Users responding to these parameters, but not to μ_r or μ_{a_2} do not scan their systems for potential threats, but respond when an entity they know alerts them to a possible threat. For example, a user may respond to a Firefox reminder to update their browser or the exhortation of a friend. A specific instance of this was Google's effort to alert users to possible infections on their computers [21]. This is less than ideal for maintaining a secure system, as, with limited contact, infections can persist.

μ_{a_1} defines the non-vigilant user's ability to clean or recover their system to a more secure state. This requires that non-vigilant users have access to the necessary patches and other up-to-date software to maintain a secure computer, at least

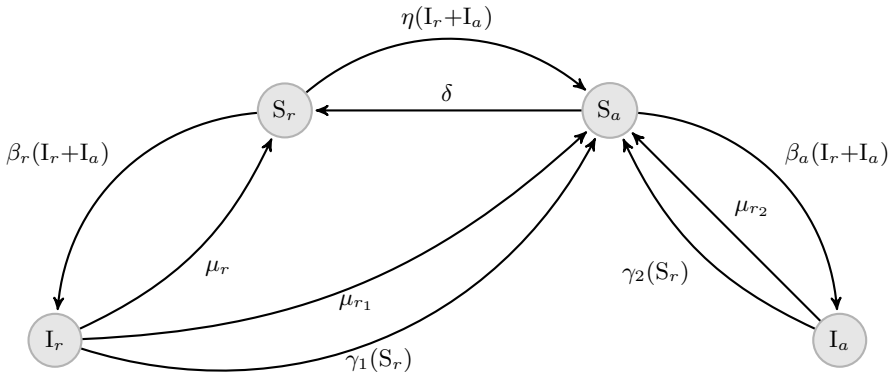

Figure 3: Two Population SIS model with Recovery and Social response.

Notation	Definition
S_r	Susceptible non-vigilant population
S_a	Susceptible vigilant population
I_r	Infected non-vigilant population
I_a	Infected vigilant population
η	Non-vigilant response to Infection
δ	Rate to return to non-vigilant population
β_r	Infection rate in non-vigilant population
β_a	Infection rate in vigilant population
μ_r	Non-vigilant recovery rate
μ_{a_1}	Non-vigilant to vigilant recovery rate
μ_{a_2}	Vigilant recovery rate
γ_{a_1}	Non-vigilant to vigilant social response rate
γ_{a_2}	Vigilant social response rate
R_∞	Equillibrium infected population
R_{∞_a}	Equillibrium infected vigilant population
R_{∞_r}	Equillibrium infected non-vigilant population

Table 1: Table Giving Definitions to included Symbols

until the cost of maintainence, δ, drives them back to the non-vigilant population.

3.2 Parameter Analysis

This model can be made equivalent to Kephart and White's "kill switch" model (Figure 2) by setting $\delta = 0.01$, $\beta = 0.5$, $\mu_{a_1} = 0.1$, and $\gamma_1 = 0.05$ and all other parameters to 0. We use both the Kephart and White (KW) model and a standard SIS model to compare our model under different parameter conditions. This allows us to evaluate which parameters may be realistic and useful.

3.2.1 Parameter Analysis in Vigilant Population Only

We first analyze the various effects of adjusting the parameters on the vigilant population to identify the most important parameters in controlling infection in that population. From there we move to analyzing the whole system, individually adjusting certain parameters to identify the key parameters in the system as a whole. For these simulations we vary one parameter and keep others constant. For each of the parameters we hold constant: β_a, μ_{a_2}, and γ_{a_2}, we set them to $0.5, 0.1, .01$ respectively.

These parameters are taken directly from KW and varied in later simulations. This sets a fixed social response at 1/10 the level of the cleaning response. This allows us to maintain consistency with our system-wide analysis below. We then vary the parameters of interest for each simulation from 0 to 1 by .01. Because we are only working with S_a, we initialize the populations to: $S_r = 0, S_a = 0.99, I = 0, I_a = 0.01$. Without an infected population I_r or I_a, no further infections are possible in this model.

3.2.2 Parameter Analysis with Both Populations

The system parameter analysis keeps the infection rate and cleaning rate in the non-security aware population at the same level as the standard SIS model used by KW ($\beta = 0.5$ and $\mu_{r_1} = 0.1$). This reduces the number of variables we must examine, and provides us with a reasonable worst case scenario of eighty percent of non-vigilant computers infected. However, we adjust the security aware population to reflect a greater vigilance.

We set the infection rate of S_a to half of the non-vigilant population's rate. Similarly, the cleaning rate of S_a is twice that found in the non-vigilant populations. In the vigilant population, there is a social response, but this is 1/10th the cleaning rate. This leads to the following parameter values: ($\beta_a = 0.25, \mu_{a_2} = 0.2$, and $\gamma_{a_2} = 0.02$). We normalize the initial populations to $S_r = 0.99, S_a = 0, I = 0.01, I_a = 0$.

Moreover, these parameter values are a reasonable estimation of actual global prevalence. Our intial parameter values in isolated populations lead to roughly 80% of the population falling into the non-vigilant population, and roughly 80% of that population infected. Within the vigilant population, the initial parameter values lead to roughly 13% of that population infected. With no interactions between the populations, this leads to a global prevalence of roughly 77%. These results correspond to estimates of global prevalence.

In their report to the House of Lords in 2007, the Science and Technology Committe reported on results from an earlier study that showed that roughly 80% computers lacked neccesary security measures, and roughly 72% of sampled systems had some type of malware [35]. However, the committee noted that this study only sampled 354 computers, so it probably was not an accurate portrayal of the actual prevalence of malware. The Anti-Phishing Working Group's (APWG) mean observed infection rate for 2010 and the 1st half of 2011, which is approximately 48% [3, 2, 1]. Thus, our initial parameter values align very closely with the earlier study, and represent approximately a 60% increase over the APWG's results.

3.2.3 Uncertainty and Sensitivity Analysis

The first two sets of analysis represent a very crude sensitivity analysis given the number of parameters. We analyze each parameter in light of a fixed system. This reduces the problem from a nine dimensional problem to a one dimensional one. However, it is not informative in terms of how the parameters interact with one another. To address this, we performed two types of sensitivity analysis in two situations.

We applied used Latin Hypercube Sampling (LHS) on the set of all parameters to measure both epistemic uncertainty and to perform a sensitivty analysis of the parameters given the measured value of total infectious computers [10]. LHS first samples from prior distributions of parameter values and generates sampled output for the number of samples. In our case, we used 1000 samples. This gives us our uncertainty analysis as we examine the variability of outputs as we vary the parameter values. From there LHS use a rank-transform correlation coefficient to measure the sensitivity of each parameter as it pertains to the measured output [10].

We used uniform priors for all parameter values, given our own uncertainty of acceptable distributions. Our initial test was performed over all parameters and used to identify the key bifurcation parameters [25]. These bifurcation parameters are key to differentiating the major equillibrium behavior in the system; mainly, whether a contagion is maintained or dies out. After we identified the key bifurcation parameters, we set them to ensure continued prevalence and performed the analysis again. This allowed us to do uncertainty and sensitivity analysis of the secondary parameters associated with reducing prevalence.

3.3 Model Fitting

To fit our model to data we used MATLAB's `lsqcurvefit` function to fit the cumulative sum of the infected risk takers and risk adverse populations to the cumulative sum of the observed data. All parameters, as well as the initial population values, had a lower bound of 0 and upper bound of 1. The population values were normalized to 1 before computation. We took our total population to be 150,000 based on the cumulative sum of the total number of attacks. We ran each fit 50 times to try to avoid local minima.

We fit our model to the top ten companies targeted by phishing scams. This data contains observed websites spoofing legitimate businesses such as banks and other online commerce. This does not represent the social aspect of the attack, but rather the infrastructure used to support such at-

tacks. We also fit our model to the total number of attacks for those companies.

To identify the top ten targeted entities, we cleaned the data to attempt to get unique identifiers. We found that, for the most part, our cleaning manage to capture most of the necessary data, but it can be refined for a more accurate picture. However, that being said, the top ten targeted entities account for 130559 out of 157355 observed attacks, roughly 83% of the total observed attacks.

4. RESULTS

We examined the effects of parameter variation in different phases. We first wanted to see if there was a way to reduce total infection prevalence by adjusting only the parameters associated with the vigilant population. After considering only the vigilant population, we investigated the effects of making the non-vigilant population respondent to the vigilant population.

We conducted this investigation by adjusting the μ_{a_1} and γ_{a_1} parameters to investigate the effect of a user's ability to recover to more secure behavior. We then adjusted the parameters that determined the speed of transition to and from the vigilant population (η and δ) in uninfected users. When the infection is less potent in vigilant users, we can reduce the total infected population by having more users become vigilant and having vigilant users stay vigilant longer. For these simulations we do not adjust the infection or clean rates, but keep the non-vigilant and vigilant population parameters at their fixed rates discussed above.

4.1 Effects of Adjusting Parameters in Vigilant Population Only

In KW's model of social response, they add a prevalence driven recovery effect on top of the standard, constant rate recovery. In order to investigate the effects of this recovery in the population we varied the social response in the vigilant population only, to see if it would lead to significant reduction in the equilibrium of total infected. Since we split the model into two subpopulations, we could also examine source of the infections.

4.1.1 Simulation 1

In Kephart and White's examination of the social response, even a small social response was able to reduce the total level of infection in the system. However, this relied on a system where the recovered population could not become infected. For this simulation we wanted to examine the effects of social response when it led to recovery, but this recovery did not protect the user from reinfection. We set the infection characteristics in the vigilant population to correspond to KW's model ($\beta_a = 0.5$ and $\mu_2 = 0.1$) and varied γ_{a_2}.

Looking at the results (Figure 4) we find that only when the combination of social response and cleaning rate is greater than the infection rate does the infection die off. That is, $\frac{\mu_{a_2}}{\beta_a - \gamma_{a_2}} = 1$ is the bifurcation point in this single population. When $\frac{\mu_{a_2}}{\beta_a - \gamma_{a_2}} < 1$ then $R_\infty = 1 - \frac{\mu_{a_2}}{\beta_a - \gamma_{a_2}}$, and when $\frac{\mu_{a_2}}{\beta_a - \gamma_{a_2}} > 1$, the infection disappears.

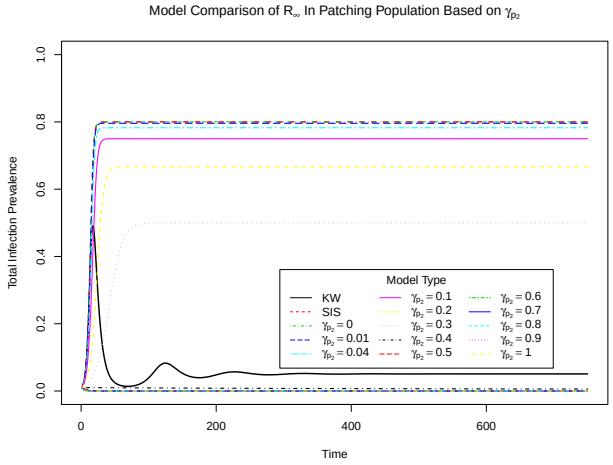

Figure 4: Comparison between SIS, KW, and Our Model, γ_{a_2} variable. Only when γ_{a_2} reaches high rates relative to β_a does R_{∞_a} fall.

These results mean that the total social response and the cleaning rate must effect the network at the same rate as the malware to be effective at eliminating its spread. For a single population then, social response is a useful measure in reducing the total infection rate, but is unlikely to be able to reduce the infection from a pandemic unless either the cleaning rate or the social response is unreasonably high. For example, the dynamic characteristics of Conficker.C described by Antonakakis *et.al.*, could be considered a pandemic infection [4]

As we will see, when we look at Simulation 9, the social response that allows non-vigilant to become vigilant, γ_{a_1} is an key ingredient to reducing the total R_∞. In our model a shift for the vigilant to the non-vigilant population is associated with an increase in the rate of social patching. An increase in the expense of recovery decreases the vigilant population and thus decreases the presumptive efficacy of social recovery. Moreover, since γ_{a_1} is prevalence driven and δ is constant, any amount of shifting is able to produce some amount of prevalent vigilant population.

4.1.2 Simulation 2

In Simulation 2 we varied the cleaning rate in the vigilant population. Without any cleaning rate, even in the presence of a social response, $R_\infty = 1$. In fact, when $\mu_a = 0$, the bifurcation point becomes $\frac{\beta_a}{\gamma_{a_2}} = 1$, and when $\frac{\beta_a}{\gamma_{a_2}} > 1$, $R_\infty = 1$. When $\frac{\beta_a}{\gamma_{a_2}} < 1$, $R_\infty = 0$. This means that, within a given population, the recovery rate is the most important aspect for reducing total infection.

Moreover, if we compare the results from Simulation 1 with Simulation 2 (Figure 5), as well as the bifurcation analysis, we note that without some sort of cleaning rate, social response is unable to reduce the total infected population on its own unless its rate of response is higher than the rate of infection. This means that only social responses without a systematic cleaning or recovery policy merely reduce the rate of spread, but not the overall infection equilibrium.

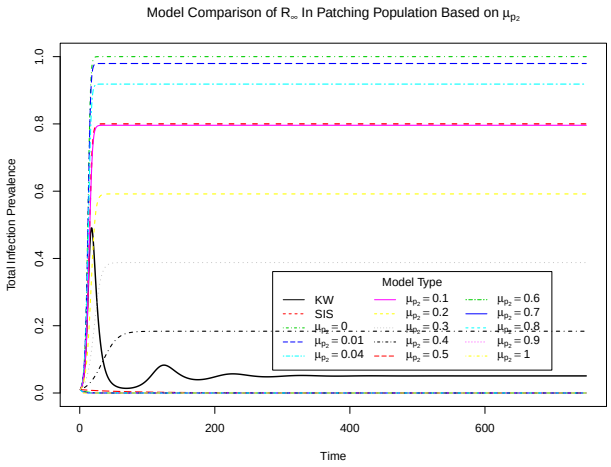

Figure 5: Comparison between SIS, KW, and Our Model, μ_{a_2} variable. Reductions in μ_{a_2} result in widespread prevalence. Moderate increases in μ_{a_2} can significantly reduce R_{∞_a}. When the rate of recovery exceeds β_a, $R_{\infty_a} = 0$.

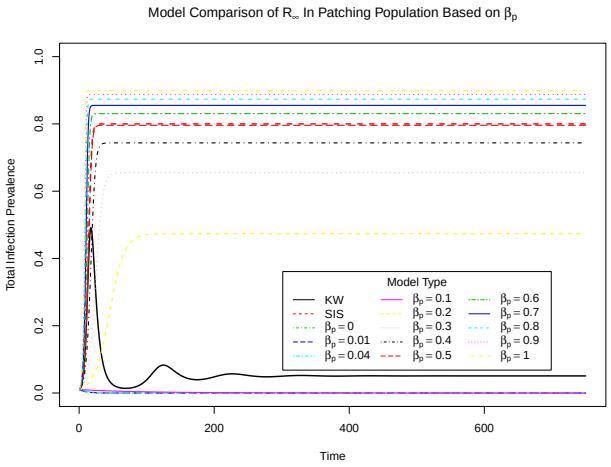

Figure 6: Comparison between SIS, KW, and Our Model, β_P variable. When β_a is reduced, global malware prevalence is reduced. Even at extreme values of β_a, μ_a is able to prevent R_{∞_a} from reaching 1.

This offers an explanation for the continuing existence of, for example, the long past Blaster worm [7]. Whether or not increasing the rate of social response to the level that it can be effective depends upon the topologies of the malware spread and social response. Enhancing social response is the focus of related human subjects research in progress. Exploring different malware diffusion and network topologies is the subject of future research.

4.1.3 Simulation 3

Simulation 3 is the complementary analysis to γ_{a_2}. There is no pandemic outbreak if $\beta_a - \gamma_{a_2} < \mu_{a_2}$. In practical terms, reducing β_a corresponds to user behavior, rather than cleaning/recovery of systems, as represented by μ_{a_2}. Examples of this include protected browsing, keeping a system up to date, or other security measures.

We see that, obviously, β_a is an important part of the infection. One thing that is noticable in our simulations is that β_a need not be very high to achieve a relatively high R_∞. When $\beta_a = 0.2$ (twice μ_{a_2}), $R_\infty = 1 - \frac{.1}{(0.2-.01)} = .474$, a value close to the 2 year average of infections provided by the APWG [3, 2, 1].

4.2 Effects of Adjusting A Single Parameters in Both Populations

For this set of analyses, we adjusted a single parameter in both vigilant and non-vigilant populations and investigated how it affected the R_∞ for the entire population. This allows us to study the global effect of a given parameter. The first three simulations examine the effects of behavior changes in the vigilant population. The final simulations study the interactions between the two uninfected populations governed by η and δ. We find that the principle way to reduce R_∞ is to allow infected individuals to recover to the vigilant population.

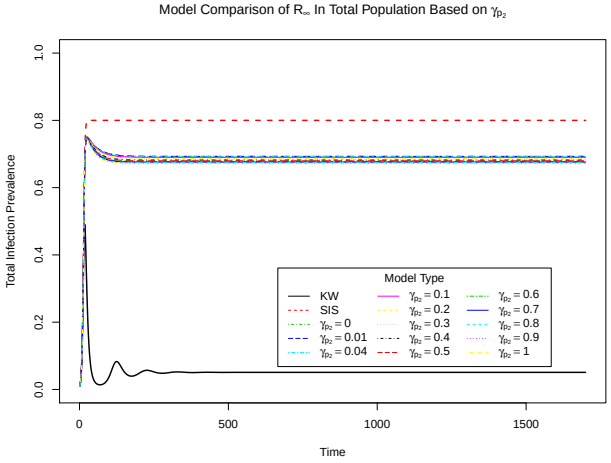

Figure 7: Comparison between SIS, KW, and Our Model, γ_{a_2} variable. Increasing the social response rate does reduce R_∞, but even at extreme values, γ_{a_2} is unable to significantly reduce R_∞ due to the effects of the large infected non-vigilant population.

4.2.1 Simulation 4

In this simulation we adjust the social response parameter in the vigilant population. This allows us to see how increasing the parameters in vigilant population has on the system-wide R_∞ (Figure 7). We notice, as in the following two simulations, increasing the responses in the vigilant population does little to reduce the total R_∞.

The dynamic relationship between γ_{a_2} and R_∞ is a bit more complicated in this simulation, as this simulation contains an I_r value that is non-zero, and transitions between the populations. We are holding $\mu_{a_1} = 0$ and $\gamma_1 = 0$, so we know that the relationship between $\eta = 0.05$ and $\delta = 0.04$ gives us

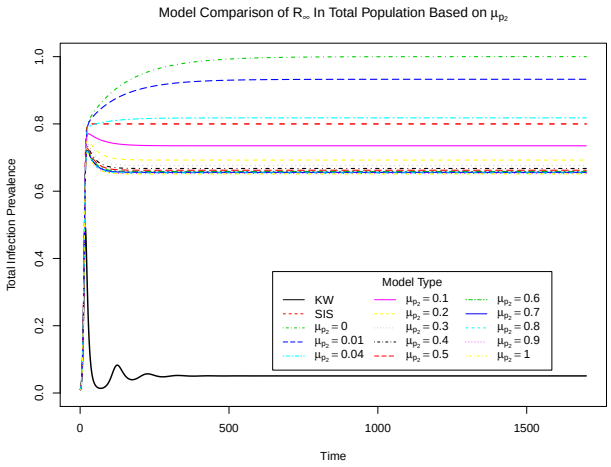

Figure 8: Comparison between SIS, KW, and Our Model, μ_{a_2} variable. Reductions in μ_{a_2} lead to increases in R_∞, as the global prevalence tends towards the behavior of the least secure population. Increases in μ_{a_2} are capable in reducing R_∞, but the reductions in R_∞ are mitigated by the behavior in the non-vigilant population.

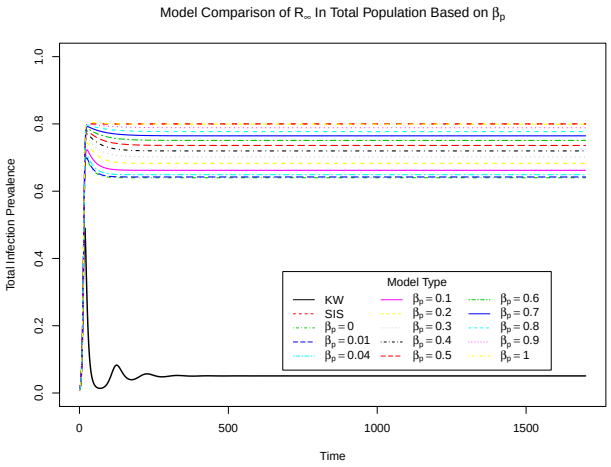

Figure 9: Comparison between SIS, KW, and Our Model, β_P variable. Increasing β_a increases R_∞, but due to the values of γ_{a_2} and μ_{a_2}, R_∞ performs as R_{∞_r}. If γ_2 or μ_{a_2} are reduced, R_∞ would also increase, as the vigilant population would be the least secure population, in that case.

an approximate 80-20 split between security conscious users and those that are unable or unwilling to engage in more secure behaviors. We also know that when $\gamma_{a_2} + \mu_{a_2} > \beta_a$, $R_{\infty_a} = 0$, in an isolated situation.

However, even when $\gamma_{a_2} = 1$, we still end up with an infected vigilant population. In this case, $R_{\infty_a} \approx 0.072$, while $R_{\infty_r} \approx .6$. $R_{\infty_a} = 0.072$ represents approximately 31% of the vigilant population, while $R_{\infty_r} \approx .6$ is approximately 77% of the non-vigilant population. Recall Figure 4 that illustrated that with only a vigilant population $\gamma_{a_2} = 1$ should remove all contagion within the vigilant population. Thus, the infections within the vigilant population are being driven by the non-vigilant population.

4.2.2 *Simulation 5*

In this simulation we adjusted the cleaning rate within the vigilant population. The key result here is if $R_{\infty_a} > R_{\infty_r}$, R_{∞_a} drives the total R_∞ (Figure 8). However, this is unlikely, as it is improbable that vigilant users will become infected at a greater rate than non-vigilant users. While, if $R_{\infty_a} < R_{\infty_r}$, but $\gamma_{a_2} + \mu_{a_2} < \beta_a$, the infection is driven by both vigilant and non-vigilant populations. In the case where $R_{\infty_a} < R_{\infty_r}$, and $\gamma_{a_2} + \mu_{a_2} > \beta_a$, the infections in the security aware population are due to the prevalence of infectious non-vigilant systems.

For example, when $\mu_{a_2} = 0$, there is no cleaning and the social response cannot reduce the spread of the infection within the vigilant population. Thus, $R_{\infty_a} = 1$. At the end of the 1700 time steps in our simulation, $R_\infty = 1$, with most of it (99.998%) being made up of the "vigilant" population. This suggests that the vigilant population cannot rely merely on protective measures to avoid infection, but must also be diligent in actively monitoring and maintaining their systems.

4.2.3 *Simulation 6*

This simulation adjusted the β_a parameter to investigate how allowing the vigilant population to reduce, or increase its infection rate would affect the system-wide R_∞. Given the parameter values for μ_{a_2}, as β_a increases to 1, the vigilant populations dynamics approach those of the non-vigilant population. Hence the convergence to $R_\infty = 0.8$, as β_a goes to 1 (Figure 9).

What is interesting about this simulation is, if our recovery and social response parameters were such that the reproduction rate of the vigilant population were greater than the non-security aware population, it would pull the system to a total R_{∞_a}, as users would flee the infectious environment of the non-vigilant group, to the even more hostile vigilant group.

4.2.4 *Simulation 7*

For this simulation, we varied η to see how increases in the response rate of non-security users in the face of infection impacted R_∞. Recall that η represents a user's ability to transition from non-vigilant, to vigilant, to reduce the likelihood of infection, in the face of an impending infection. Obviously, when $\eta = 0$, there is no transition to the security aware population, and the model behaves as a standard SIS model as shown in Figure 10.

However, when $\eta > 0$, the model behaves in an interesting manner. It is possible to see that increasing η reduces the system-wide R_∞ (Figure 10). But, there is a complex relationship going on between η and R_{∞_a}. As seen in Table 4.2.4 and Figure 10, while the total R_∞ is decreasing, the R_{∞_a} increases until $0.6 < \eta < 0.7$, when it begins to decrease. It is also possible to see that R_{∞_a}, while increasing

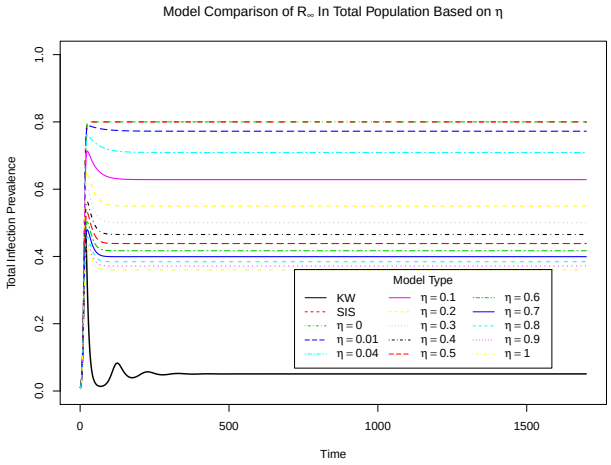

Figure 10: Comparison between SIS, KW, and Our Model, η variable. Increases in the ability for the susceptible non-vigilant population to become vigilant reduces R_∞.

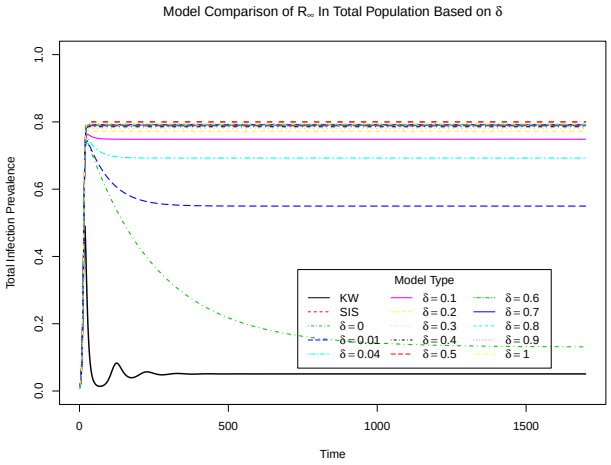

Figure 11: Comparison between SIS, KW, and Our Model, δ variable. Reducing users willingness to become non-vigilant reduces R_∞.

η	% Population$_a$	% Population$_a$ Infected	R_{∞_a}
0	0	-	0
0.1	40.1	43.4	.174
0.2	54.9	39.9	.219
0.3	63.1	37.5	.237
0.4	68.5	35.7	.245
0.5	72.3	34.3	.248
0.6	75.2	33.1	.249
0.7	77.4	32.1	.249
0.8	79.2	31.3	.248
0.9	80.8	30.5	.246

Table 2: Interaction between η and R_{∞_a}

in those intervals, is always decreasing as a percentage of the vigilant population.

η pulls more of the total population into the vigilant population, but, until it is able to overcome the increasingly small non-vigilant population, that population still exerts a growing cost on the vigilant population. This result is important, since it indicates that even a small population engaged in risk behavior, with limited opportunity to reduce their risk, threatens a larger, risk averse population.

When $\eta \gg \delta$, it is unable to pull R_∞ to R_{∞_a} in the isolated system case. Yet even an η as low as 0.1, is capable of reducing R_∞ more than any of the test values of β_a, μ_{a_2}, or γ_{a_2}. This suggests that if modifying η is feasible, it would have a significant impact on global malware presence.

4.2.5 *Simulation 8*

In this simulation we varied the other part of the transitions from non-vigilant to security. δ represents the constant rate of relapse where users view the costs of maintaining security may not be worth it. Reducing δ represents increasing a users willingness to engage in more secure behavior, while increasing δ represents users that are only willing to be vigilant in the face of large outbreaks.

What becomes immediately apparent is that when $\delta = 0$, R_∞ approaches R_{∞_a} (Figure 11). However, $\delta = 0$, while ideal, is unlikely. It represents a population that is fully vigilant, irrespective of cost. We can see that reducing δ from 0.04 to 0.01, results in $R_\infty \approx .549$, which is lower than the R_∞ achievable by extreme values in β_a, μ_{a_2}, or γ_{a_2}. It is unlikely that such lack of sensitivity is realistically achievable, though it is probably reasonable to assume that η and δ are of the same order of magnitude.

4.2.6 *Simulation 9*

In this simulation we investigate the ability of users to recover to a vigilant population through social response, rather than merely recovering to the standard susceptible population. γ_{a_1} represents non-vigilant users' ability to respond to social pressure applied by non-infected vigilant users, not just to clean their machines, but to also, at least for some time, to become vigilant users.

In KW's social response model, γ_{a_1} is kept to 1/10 of standard cleaning rate, but is effective at reducing R_∞ due to the lack of infection rate in the recovered population, and the inability to recover directly back to the susceptible population [19]. We concur with their assumption, in terms of limiting the social response rate. However, it is important to note how effective increases in γ_{a_1} are at controlling width of the infection peak curve, and mitigating R_∞. Arguably, γ_{a_1} should be limited in regards to β and μ, it may be, that given certain network topologies, even a relatively low γ_{a_1} will still be effective at reducing R_∞.

4.2.7 *Simulation 10*

In our final simulation, we vary μ_{a_1}, the parameter representing cleaning a computer and adapting vigilant behavior. For example, a user reinstalling an OS and applying patches and installing AV software, rather than just removing malware and hoping to avoid infection in the future. When

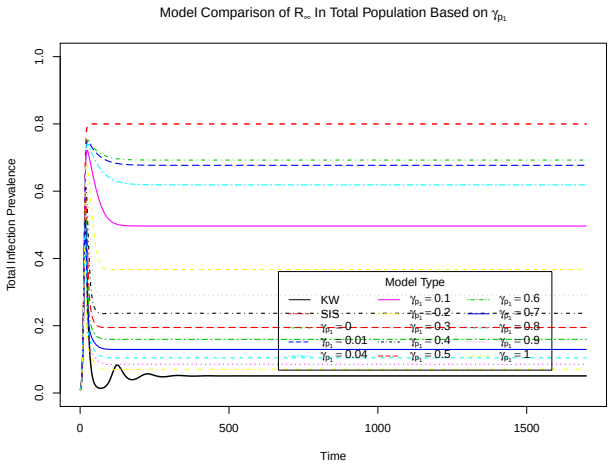

Figure 12: Comparison between SIS, KW, and Our Model, γ_{a_1} variable. Increases in a users' ability to become vigilant in response to social pressure is effective at reducing R_∞.

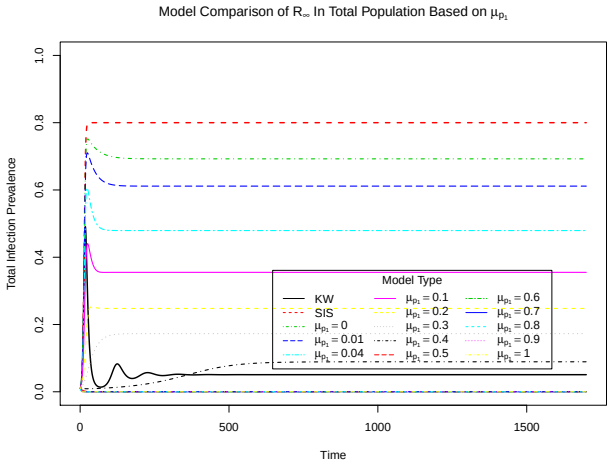

Figure 13: Comparison between SIS, KW, and Our Model, μ_{a_1} variable. Allowing users to recover their systems into the vigilant population is effective at reducing R_∞.

$\mu_{a_1} = 0$, users are unable to become vigilant users until they clean their computers and respond to the infection through η. $\mu_{a_1} > 0$ means that users have some method to recover directly to vigilant behavior.

μ_{a_1} is not as effective as γ_{a_1} at limiting the duration of the infection peak, but it does limit the peak's height, limiting the total infections. Moreover, μ_{a_1} is effective at low parameter values. $\mu_{a_1} = 0.05$, or 1/10 β, reduces $R_\infty = .451$, nearly half the R_∞ of the standard SIS model, and without adjusting any other parameters. This suggests that providing users with the ability to recover to updated and secured software/machines, should be a key component in any campaign to limit global prevalence of malware.

4.3 Uncertainty and Sensitivity Analysis

We used Latin hypercube sampling to examine both the epistemic uncertainty of the model, as well as the sensitivity of output variation to parameter variation [25]. The first step in LHS is to sample the parameter space to create a collection of measured outputs based on those samples. We did this sampling twice: first with all parameters sampled, followed by fixed values for the identified bifurcation parameters. In both cases we sampled the parameter space 1000 times. Our output of interest was total infection prevalence (Figures 15 and 16).

Since the output of interest varies over time, so too does the uncertainty. Thus, we created output probability distributions for each time step that we observed. The data that we later analyzed spans 100 days, so we looked at the first 100 time steps in our model.

The key result of our uncertainty analysis is how often the model settles into an equillibrium state with no infection. In the case when all parameters are varied, this occurs roughly 70% of the time by day 20 (Figure 17), and when the bifurcation parameters are fixed, that percentage hovers around 60%, though it takes a bit longer to reach that level (Figure 18). This means that, even with the key bifurcation parameters fixed to ensure prevalence outside the impact of other parameters, the vast majority of parameter combinations lead to no prevalence. It is easy for the recovery parameters to overcome the infection parameters in most cases due to the number of recovery parameters.

The results from the sensitivity analysis are equally interesting. Figure 19 shows the changing sensitivity of each parameter as time progresses. In the initial stages of the infection, social response and recovery from risk takers to the risk adverse population is more important than recovery within the risk adverse population. However, it rapidly loses its importance on overall prevalence, while risk adverse recovery increases its importance as time progresses.

All three of the standard recovery parameters (μ_x) are of approximately the same importance in the long term reduction of prevalence. However, the infection rate in the risk adverse group (β_P) loses its sensitivity gradually. The transmissions between susceptible risk takers and susceptible risk adverse (η and δ) are not significant in terms of affecting the global prevalence of a contagion, just as social response within the risk adverse community (γ_2).

When we fix the main bifurcation parameters ($\beta_r = 0.5$, $\beta_a = 0.25$, $\mu_{r_1} = 0.1$, $\mu_{a_1} = 0.01$, and $\mu_{a_2} = 0.02$), however, we get a better view of the effects of the social parameters. When all parameters are varied, γ_1 is significant parameter for reducing prevalence in the initial stages of a contagion, while γ_2 is never significant. However, when a contagion exists, we find that both of the social response recovery rates are important, at least until the later stages of a contagion, moreso than the transfer from susceptible risk takers to susceptible risk adverse (Figure 20).

4.4 Fitting the Model to Data

In our examination of the data we sorted each attack based on what online entity a given website was spoofing. We

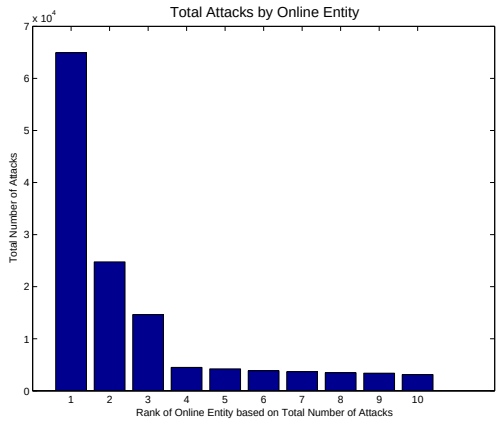

Figure 14: Plot of top-ten targeted entities by rank and total number of observed attacks.

tabulated the total number of attacks on each entity to find the top ten targets of observed attacks. These attacks have a heavy tail distribution (Figure 14, and seem to confirm Maillart and Sornette's work [24], but lower ranked entities still create large jitters in the overall behavior.

We normalized the SSE to the minimum error across all fits and companies when we plotted the fits. Due to the complex interactions of the parameters, while the model can fit the data, often times the parameter values do not make sense. Thus, the uncertainty and sensitivity analyses are more important to this analysis, than the fit. However, the fits are useful to see how the model can represent real world data, and demonstrate the long term prevalence of attacks.

The model fits are found in Section A.3. The key results found in our model fits is that, while our model can handle the baseline prevalence of attacks in many cases (Figure (21), it cannot handle abrupt spikes in contagion behavior (Figure 35). This suggests that we need to have some further refinement to our model. Ideally, we would like to consider better representations of contact patterns, but also the exploration of birth/death rates, representing new computers entering the network and older computers being shut off.

5. DISCUSSION

In Section 6 we reify the conclusions of the ten simulations described in this work. In this section we discuss the possible implications of our findings. That extreme changes in β_a have little effect in the equilibrium state of an infection is an encouraging result. The rate of spread of an infection is one variable completely subject to the control of the attacker. Therefore great efficacy in changes in beta would imply that defense could be ultimately futile.

Increasing the roughly equivalent variable, μ_a, is found to be as ineffective as β_a in decreasing the global prevalence of infection. However, there are a significant caveats. The outcome assumes that the malware will remain endemic with a roughly constant β and that recovery does not result in immunity to a particular malware component. Yet given the existence of multiple malware attacks, the use of multiple vectors for a single malware variant, the lack of broad immunity upon recovery, and the potential for malware to evolve these are not unreasonable assumptions.

Individuals choosing the recover due to social pressure (which includes automated pressure, such as Firefox exhortations to upgrade) must be faster than the rate at the virus is spreading. This is an extremely unlikely case. Yet the social recovery rate, γ_{a_1}, is one of the most effective measures in altering the equilibrium when there are two populations (vigilant and otherwise). However, increasing the response rate in the vigilant population has little effect on the global equilibrium. This is a mixed result given that it is arguably easier to alter a response rate in an aware population, but even modest gains in response of the unaware population can significantly reduce the global prevalence.

Transfer rates between the two populations is the most efficacious strategy for reducing long-term equilibrium. This argues that small increases in vigilance can result in significant increases in outcomes. Thus the title of the paper where increased use of healthy behaviors (e.g., contraception use or smoking cessation) can greatly reduce unintended consequences over the population as a whole. Compared this to situations where the entire population must engage in healthy behaviors (e.g., immunization) to result in significant outcomes. This argues for an approach that is closer to risk communication than mandates. Luckily, risk communication is feasible while global mandates are not.

Users must be able to act upon available information, e.g., δ should be quite low. The requires an ease of access to the resources necessary to engage in more secure behavior. Within the public health sector, barriers to treatment and preventative measures have been shown to greatly increase overall costs. For example, Franzini *et.al.*, estimated a likely additional cost of $43.6 million in a one year period in Texas, if adolescents were required to notify parents when they received reproductive health care [16]. This suggests that allowing access to security patches, even in the case of illegal copies, would be effective in lowering system-wide costs, though offering those patches may not be profit maximizing for a given firm [22].

Moreover, risk communication, when combined with access to treatment resources, has been effective in reducing prevalence in the public health sector. Spain *et.al.* demonstrated the effectiveness of at risk communication at recruiting at risk groups to utilize reproductive and preventative health care [38]. Several studies demonstrate the effectiveness of

Youth Peer Education services at referring at risk populations to appropriate clinics [23, 13]. When coupled with a voucher system for care, use of clinics increases dramatically [9]. Thus, there are extant systems of response and information that we can take advantage of in regards to encouraging more secure behavior.

The difference between the two populations are rate of recovery (μ), responsive to social pressure (γ) and decreased rate of infection (β). Therefore the findings above, of lack of efficacy of contact rate in the risk adverse population (β_a), social recovery rate (γ_{a_2}), and recovery rate (μ_{a_2}) are due primarily to the infectous interaction that the risk adverse population exerts as well as the interactions between these variables and the ability to become risk adverse before infection (δ) and difficulty to remain risk adverse (η). In future work we will extend the model to include this feedback.

6. CONCLUSIONS

In this paper we created and examined the parameters of two-population SIS epidemiological model in regards to global prevalence of malware. The two populations, vigilant, and non-security aware, interact in many different ways (Figure 3), which affects R_∞, to equilibrium infected population. We examined single parameter variations within the vigilant population and the system as a whole to identify key components to addressing the spread of malware.

In our first set of simulations, we examined the vigilant population in isolation, seeking to identify the most effective parameter for reducing or removing malware in that population. We found that, within the single population it was possible to completely eliminate malware spread by setting $\mu_{a_2} + \gamma_{a_2} > \beta_a$. We also showed that adjusting the recovery rate μ_{a_2} is the most effective way to reduce R_∞ in the vigilant population.

In our second set of simulations, we looked at the entire system and tried to find which parameters were effective at reducing global R_∞, while keeping the infection and recovery rates (β and μ) in the non-vigilant population constant. Here we find that, while we could eliminate the spread of infection within the vigilant population by overcoming the infection rate, adjusting the vigilant parameters had little effect on R_∞, and infections in the non-vigilant population drove the infections. However, when we examined the parameters governing transitions from non-vigilant to vigilant, we discovered several possibilities for infection control.

When we evaluated the transitions between uninfected non-vigilant and uninfected vigilant populations, we found that, while η was effective at making more users vigilant, even a small population of infected non-vigilant users could negatively impact the vigilant population. Similarly, when we prevented users from returning to the non-vigilant population, we could limit the infection spread to R_{∞_a}. Yet, this represents an unrealistic expectation of inelasticity (i.e., all users demanding secure behavior, regardless of cost).

Examing the parameters governing the recovery of infected non-vigilant to uninfected security aware, we find that allowing users to clean and repair their systems with updated and secure software is the most effective way at managing global prevalence of malware infection. Even at low levels, the ability to recover to the risk adverse population (μ_{a_1}) greatly reduces the global infection prevalence. Additionally, while not as effective as the risk adverse recovery rate (μ_{a_1}), the social response recovery to the risk adverse population (γ_{a_1}) is the next most effective parameter. This suggests that coupling social response, along with access to updates for all users, would be an effective measure for reducing the prevalence of global malware.

7. ACKNOWLEDGEMENTS

We would like to thank Richard Clayton and Tyler Moore for access to their data. We would also like to thank Alessandro Vespignani for his suggestions on model building and analysis. We would further like to thank the Volkswagen Foundation for their support of our research by providing funds to travel to WEIS 2012.

8. REFERENCES

[1] Anti-Phishing Working Group. Phishing Activity Trends Report 2nd Half 2010. Tech. Rep. December, Anti-Phishing Working Group, 2010.

[2] Anti-Phishing Working Group. Phishing Activity Trends Report 2nd Quarter 2010. Tech. Rep. June, 2010.

[3] Anti-Phishing Working Group. Phishing Activity Trends Report 1st Half 2011. Tech. Rep. June, Anti-Phishing Working Group, 2011.

[4] Antonakakis, M., Perdisci, R., Dagon, D., Lee, W., and Feamster, N. Building a dynamic reputation system for DNS. In *19th Usenix Security Symposium* (2010).

[5] August, T., and Tunca, T. I. Let the Pirates Patch? An Economic Analysis of Software Security Patch Restrictions. *Information Systems Research 19*, 1 (Mar. 2008), 48–70.

[6] Baggaley, R. F., Garnett, G. P., and Ferguson, N. M. Modelling the Impact of Antiretroviral Use in Resource-Poor Settings. *PLoS Medicine 3*, 4 (2006), e124.

[7] Bailey, M., Cooke, E., Jahanian, F., Watson, D., and Nazario, J. The Blaster Worm: Then and Now. *IEEE Security and Privacy Magazine 3*, 4 (July 2005), 26–31.

[8] BARABASI, A., ALBERT, R., and JEONG, H. Scale-free characteristics of random networks: the topology of the world-wide web. *Physica A: Statistical Mechanics and its Applications 281*, 1-4 (June 2000), 69–77.

[9] Bellows, N. M., Bellows, B. W., and Warren, C. Systematic Review: the use of vouchers for reproductive health services in developing countries: systematic review. *Tropical medicine & international health : TM & IH 16*, 1 (Jan. 2011), 84–96.

[10] Blower, S., and Dowlatabadi, H. Sensitivity and uncertainty analysis of complex models of disease transmission: an HIV model, as an example. *International Statistical Review/Revue 62*, 2 (1994).

[11] Boily, M.-C., Bastos, F. I., Desai, K., and Mâsse, B. Changes in the Transmission Dynamics of the HIV Epidemic After the Wide-Scale Use of Antiretroviral Therapy Could Explain Increases in

Sexually Transmitted Infections. *Sexually Transmitted Diseases 31*, 2 (Feb. 2004), 100–113.

[12] Brown, T., Bao, L., Raftery, A. E., Salomon, J. a., Baggaley, R. F., Stover, J., and Gerland, P. Modelling HIV epidemics in the antiretroviral era: the UNAIDS Estimation and Projection package 2009. *Sexually transmitted infections 86 Suppl 2* (Dec. 2010), ii3–10.

[13] Burke, H. M., Pedersen, K. F., and Williamson, N. E. An assessment of cost, quality and outcomes for five HIV prevention youth peer education programs in Zambia. *Health education research* (Nov. 2011).

[14] Choi, J., Fershtman, C., and Gandal, N. Network Security: Vulnerabilities and Disclosure Policy. 2007.

[15] Ferguson, N. M., and Garnett, G. P. More Realistic Models of Sexually Transmitted Disease Transmission Dynamics : Sexual Partnership Networks , Pair Models , and Moment Closure. 1–10.

[16] Franzini, L., Marks, E., Cromwell, P. F., Risser, J., McGill, L., Markham, C., Selwyn, B., and Shapiro, C. Projected economic costs due to health consequences of teenagers' loss of confidentiality in obtaining reproductive health care services in Texas. *Archives of pediatrics & adolescent medicine 158*, 12 (Dec. 2004), 1140–6.

[17] Gray, R. T., Beagley, K. W., Timms, P., and Wilson, D. P. Modeling the Impact of Potential Vaccines on Epidemics of Sexually Transmitted Chlamydia trachomatis Infection. *The Journal of Infectious Diseases 199*, 11 (June 2009), 1680–1688.

[18] Kephart, J., and White, S. Directed-graph epidemiological models of computer viruses. *Proceedings. 1991 IEEE Computer Society Symposium on Research in Security and Privacy 0* (1991), 343–359.

[19] Kephart, J., and White, S. Measuring and modeling computer virus prevalence. *Proceedings 1993 IEEE Computer Society Symposium on Research in Security and Privacy* (1993), 2–15.

[20] Kermack, W. O., and McKendrick, A. G. A Contribution to the Mathematical Theory of Epidemics. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character (1905-1934) 115*, 772 (Aug. 1927), 700–721.

[21] Krebs, B. Google: Your Computer Appears to Be Infected, 2011.

[22] Lahiri, A. Revisiting the Incentive to Tolerate Illegal Distribution of Software Products. In *44th Hawaii International Conference on System Sciences* (2011).

[23] Liambila, W., Askew, I., Mwangi, J., Ayisi, R., Kibaru, J., and Mullick, S. Feasibility and effectiveness of integrating provider-initiated testing and counselling within family planning services in Kenya. *AIDS (London, England) 23 Suppl 1* (Nov. 2009), S115–21.

[24] Maillart, T., and Sornette, D. Heavy-tailed distribution of cyber-risks. *The European Physical Journal B 75*, 3 (Apr. 2010), 357–364.

[25] Marino, S., Hogue, I. B., Ray, C. J., and Kirschner, D. E. A methodology for performing global uncertainty and sensitivity analysis in systems biology. *Journal of theoretical biology 254*, 1 (Sept. 2008), 178–96.

[26] Meiss, M. R., Menczer, F., and Vespignani, a. Structural analysis of behavioral networks from the Internet. *Journal of Physics A: Mathematical and Theoretical 41*, 22 (June 2008), 224022.

[27] Moore, D., Paxson, V., Savage, S., Shannon, C., Staniford, S., and Weaver, N. Inside the slammer worm. *IEEE Security & Privacy Magazine 1*, 4 (July 2003), 33–39.

[28] Moore, D., Shannon, C., and Brown, J. Code-Red: a case study on the spread and victims of an Internet worm. In *Proceedings of the 2nd ACM SIGCOMM Workshop on Internet measurment* (2002), ACM, pp. 273–284.

[29] Moore, T., and Clayton, R. Evil searching: Compromise and recompromise of internet hosts for phishing. *Financial Cryptography and Data Security* (2009), 256–272.

[30] Newman, M., Forrest, S., and Balthrop, J. Email networks and the spread of computer viruses. *Physical Review E 66*, 3 (Sept. 2002).

[31] Pastor-Satorras, R., and Vespignani, A. Epidemic dynamics and endemic states in complex networks. *Physical Review E 63*, 6 (May 2001), 1–8.

[32] Pastor-Satorras, R., and Vespignani, A. Epidemic Spreading in Scale-Free Networks. *Physical Review Letters 86*, 14 (Apr. 2001), 3200–3203.

[33] Perra, N., Balcan, D., Gonçalves, B., and Vespignani, A. Towards a characterization of behavior-disease models. *PloS one 6*, 8 (Jan. 2011), e23084.

[34] Renton, a. M., Whitaker, L., and Riddlesdell, M. Heterosexual HIV transmission and STD prevalence: predictions of a theoretical model. *Sexually Transmitted Infections 74*, 5 (Oct. 1998), 339–344.

[35] Science and Technology Committee. Personal Internet Security. In *5th Report of Session 2006-2007* (2007), vol. I of *5th Repo*, House of Lords, House of Lords, p. 121.

[36] Serazzi, G., and Zanero, S. Computer Virus Propagation Models. In *Performance Tools and Applications to Networked Systems*. Springer, 2004, pp. 26–50.

[37] Shiboski, S., and Padian, N. S. Population- And Individual-Based Approaches To The Design And Analysis Of Epidemiologic Studies Of Sexually Transmitted Disease Transmission. *Journal of Infectious Diseases 174*, Supplement 2 (Oct. 1996), S188–S200.

[38] Spain, J. E., Peipert, J. F., Madden, T., Allsworth, J. E., and Secura, G. M. The Contraceptive CHOICE Project: recruiting women at highest risk for unintended pregnancy and sexually transmitted infection. *Journal of women's health 19*, 12 (Dec. 2010), 2233–8.

[39] Staniford, S., Paxson, V., and Weaver, N. How to own the internet in your spare time. In *Proceedings of the 11th USENIX Security symposium* (2002), vol. 8, pp. 149–167.

[40] Wang, C., Knight, J., and Elder, M. On computer viral infection and the effect of

immunization. In *Proceedings 16th Annual Computer Security Applications Conference (ACSAC'00)* (2000), IEEE Comput. Soc, pp. 246–256.

[41] Wang, Y., and Wang, C. Modeling the effects of timing parameters on virus propagation. In *Proceedings of the 2003 ACM workshop on Rapid Malcode - WORM'03* (New York, New York, USA, 2003), ACM Press, p. 61.

[42] Yook, S.-H., Jeong, H., and Barabasi, A.-L. Modeling the Internet's large-scale topology. *Proceedings of the National Academy of Sciences of the United States of America 99*, 21 (Oct. 2002), 13382–6.

[43] Zou, C., and Towsley, D. Routing Worm: A Fast, Selective Attack Worm Based on IP Address Information. In *Workshop on Principles of Advanced and Distributed Simulation (PADS'05)* (2005), IEEE, pp. 199–206.

[44] Zou, C. C., Gong, W., and Towsley, D. Code red worm propagation modeling and analysis. In *Proceedings of the 9th ACM conference on Computer and communications security - CCS '02* (New York, New York, USA, 2002), ACM Press, p. 138.

[45] Zou, C. C., Gong, W., and Towsley, D. Worm propagation modeling and analysis under dynamic quarantine defense. In *Proceedings of the 2003 ACM workshop on Rapid Malcode - WORM'03* (New York, New York, USA, 2003), ACM Press, p. 51.

APPENDIX
A. ADDITIONAL FIGURES
A.1 Uncertainty Analysis

A.2 Sensitivity Analysis

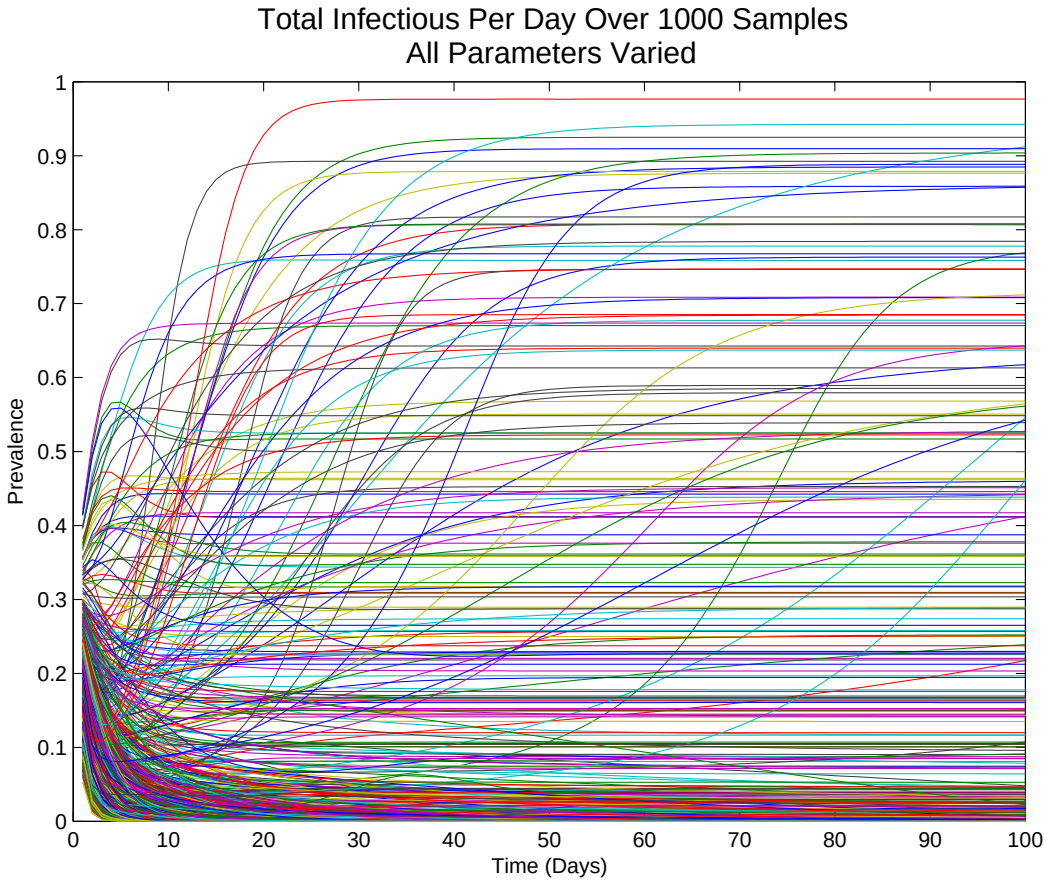

Figure 15: Output from 1000 samples from the parameter space with all parameters varied.

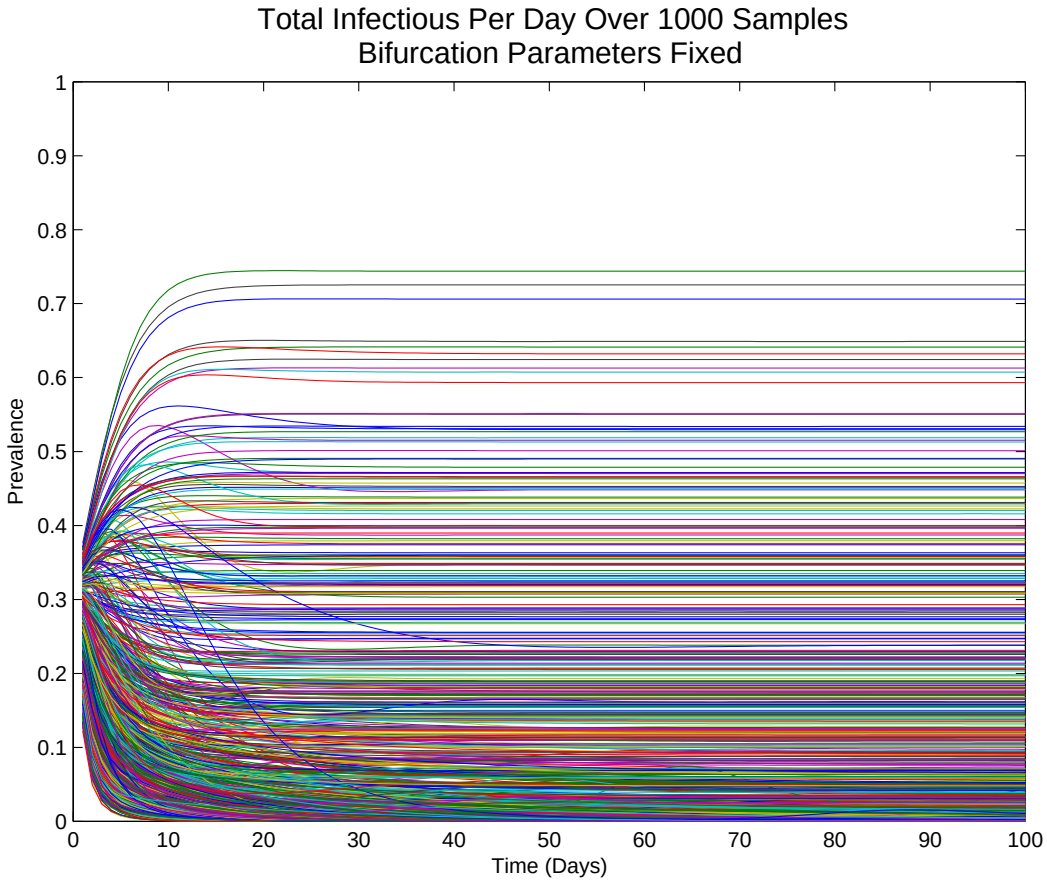

Figure 16: Output from 1000 samples from the parameter space with bifurcation parameters fixed.

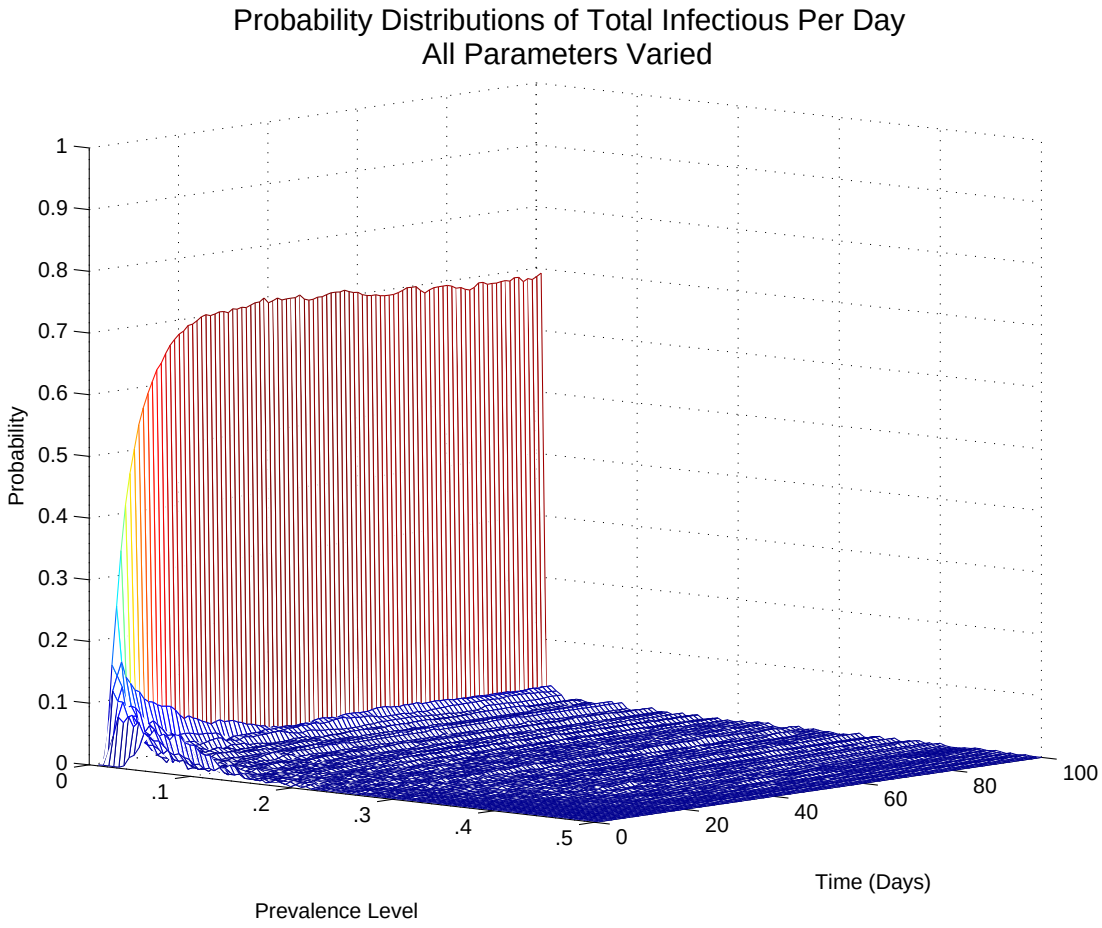

Figure 17: Probability distributions of outputs for each time used generated from 1000 runs using randomly sampled parameters.

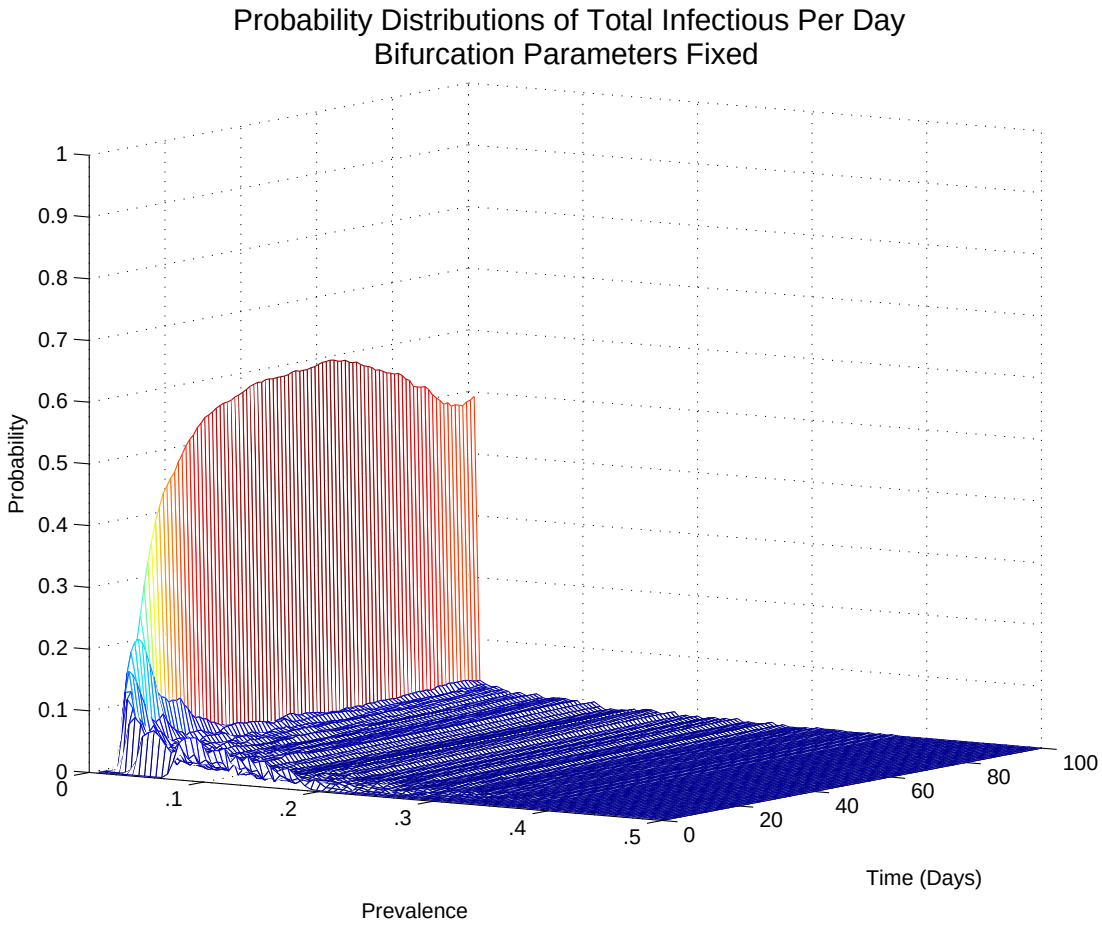

Figure 18: Probability distributions of outputs for each time used generated from 1000 runs using randomly sampled parameters with bifurcation parameters fixed.

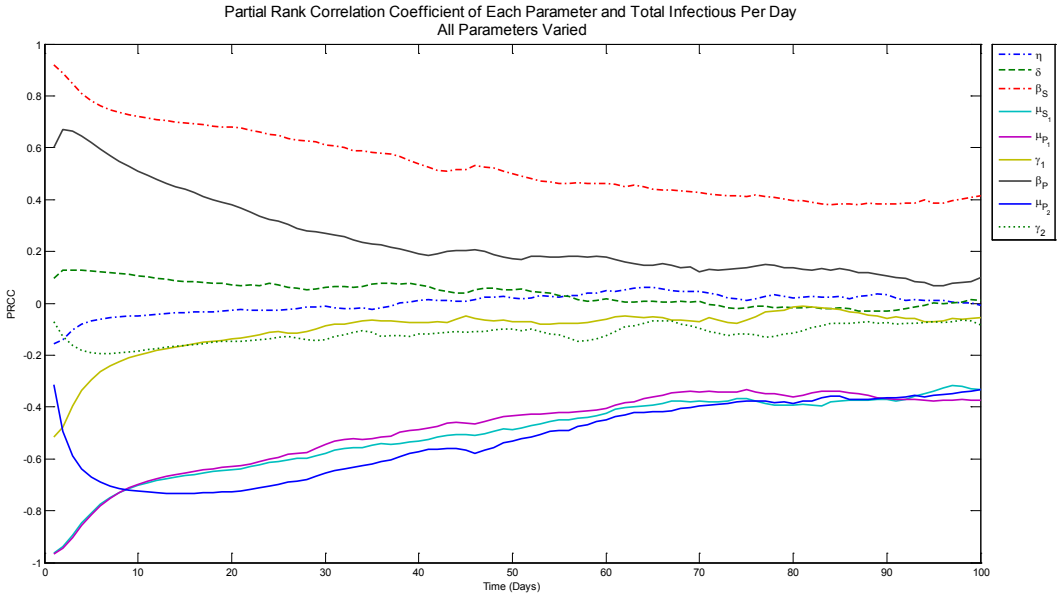

Figure 19: Partial ranked correlation coefficients for all parameters calculated for total infections at time=t.

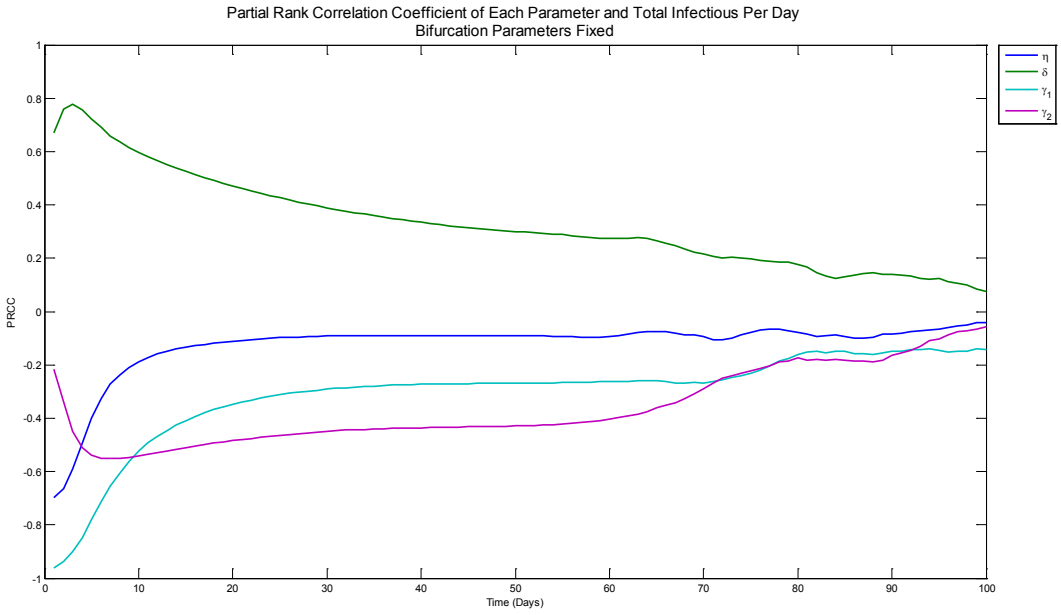

Figure 20: Partial ranked correlation coefficients for non-bifurcation parameters calculated for total infections at time=t.

A.3 Model Fitting

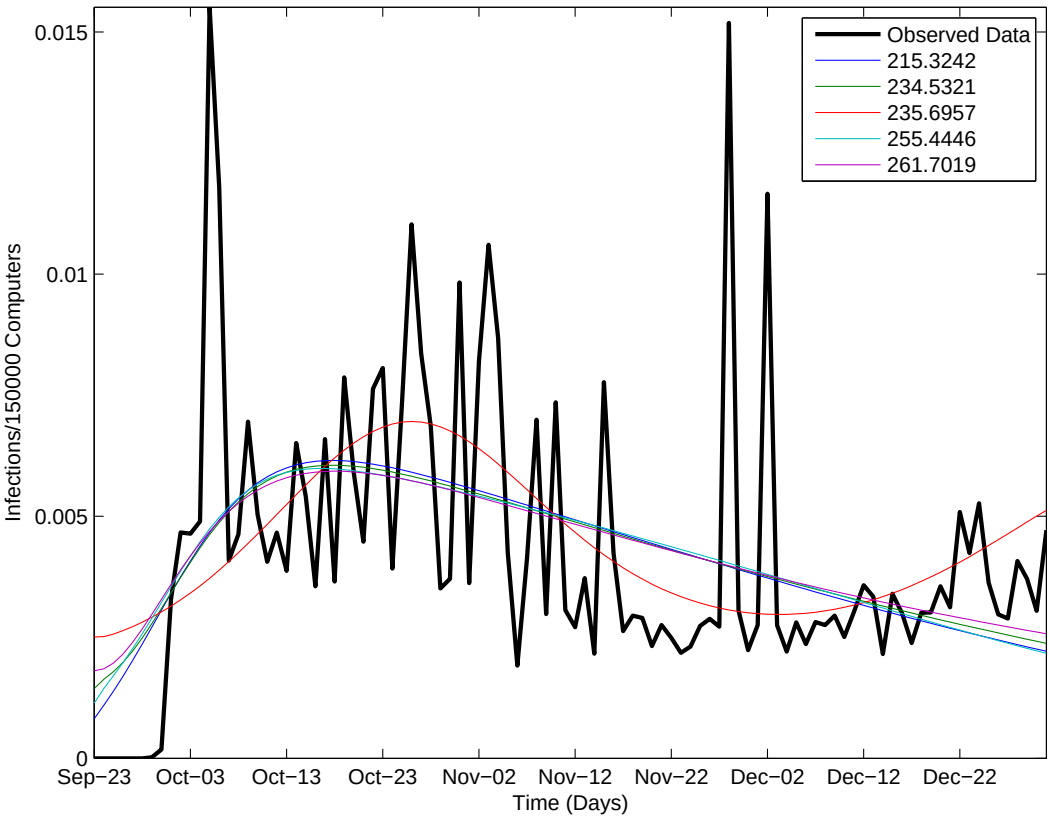

Figure 21: Top five model fits for the top targeted online entity

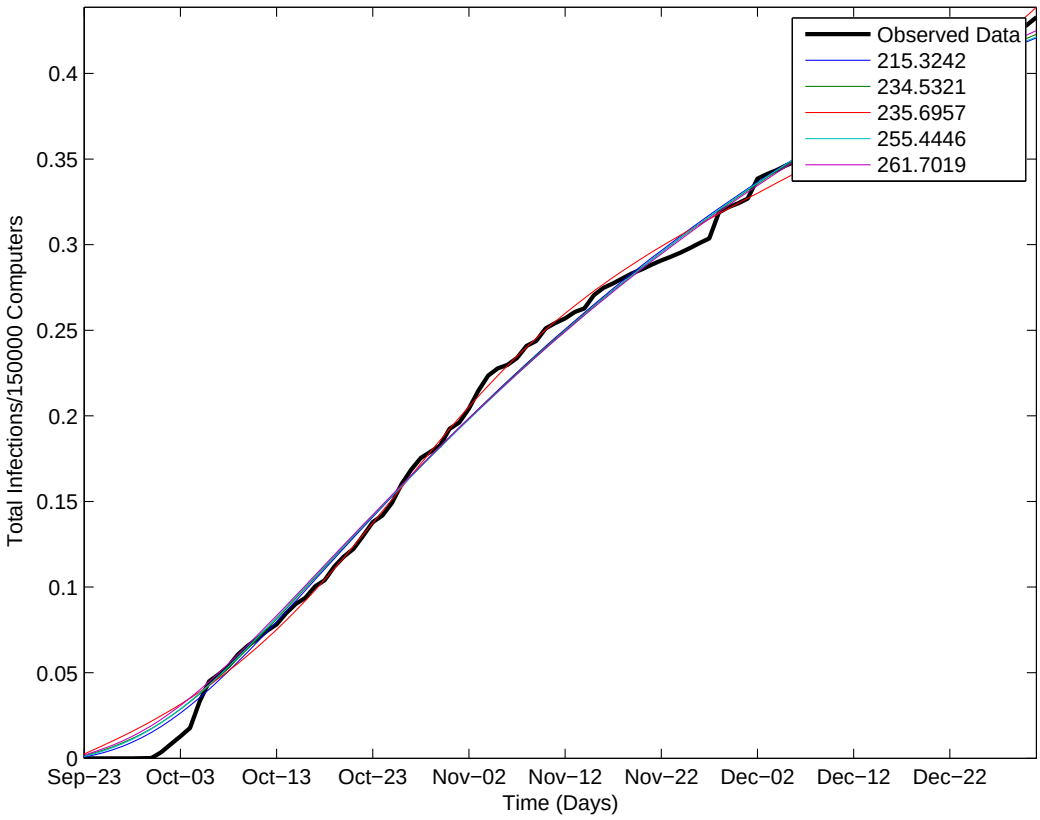

Figure 22: Top five model fits for the cumulative sum of observed attacks on the top targeted online entity

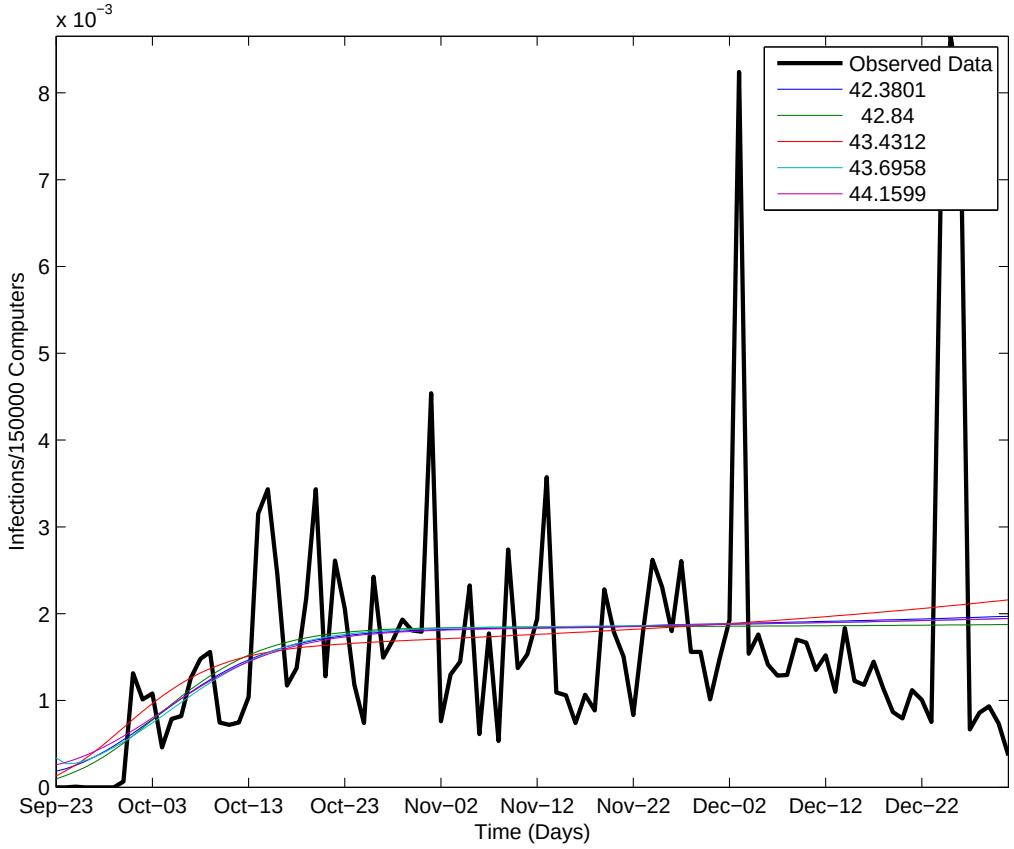

Figure 23: Top five model fits for the 2nd ranked targeted online entity

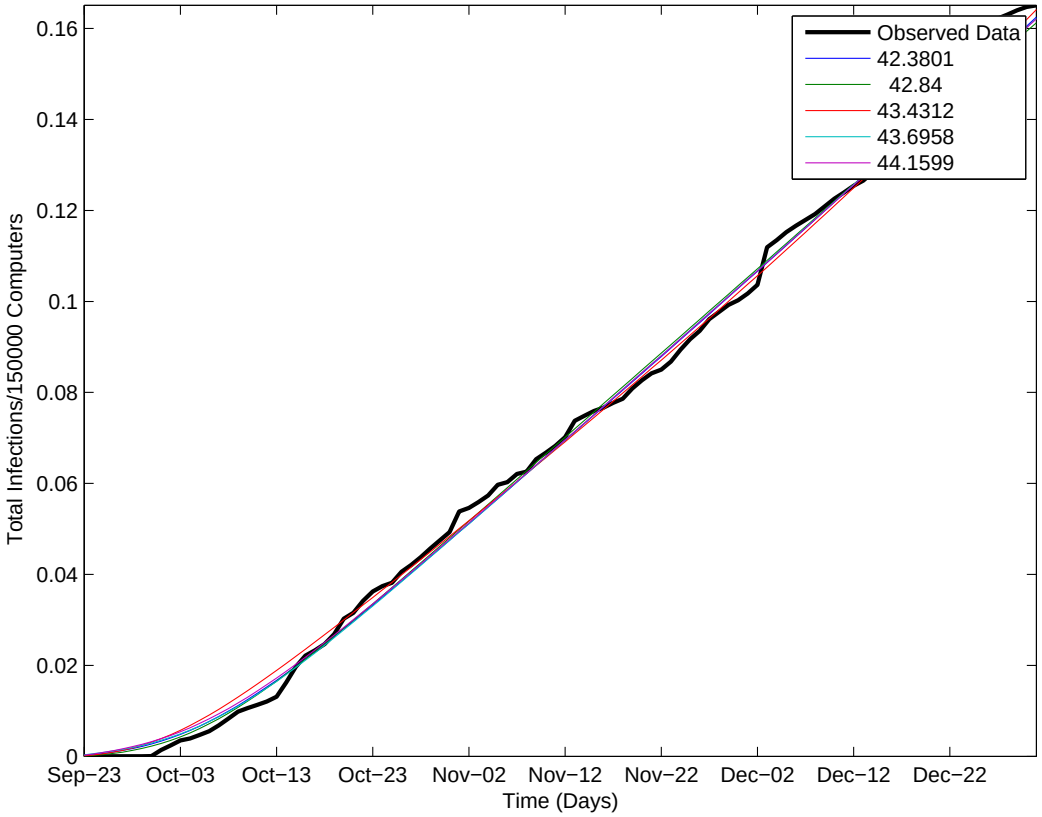

Figure 24: Top five model fits for the cumulative sum of observed attacks on the 2nd ranked targeted online entity

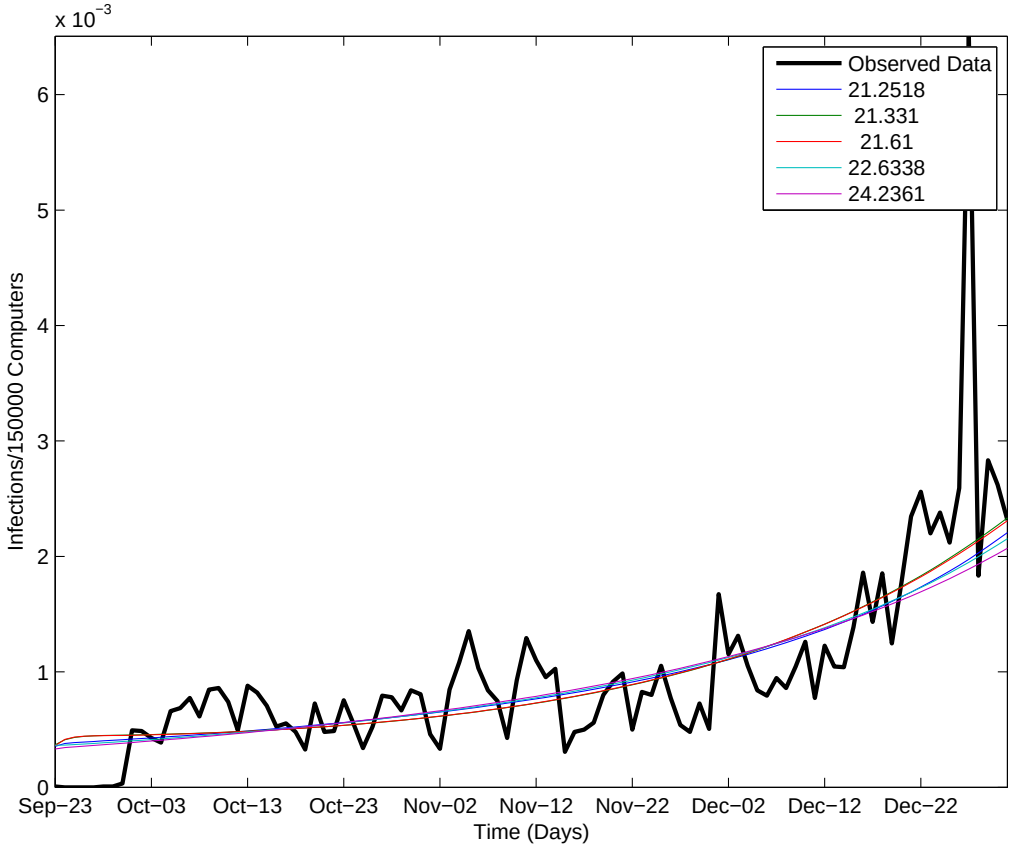

Figure 25: Top five model fits for the 3rd ranked targeted online entity

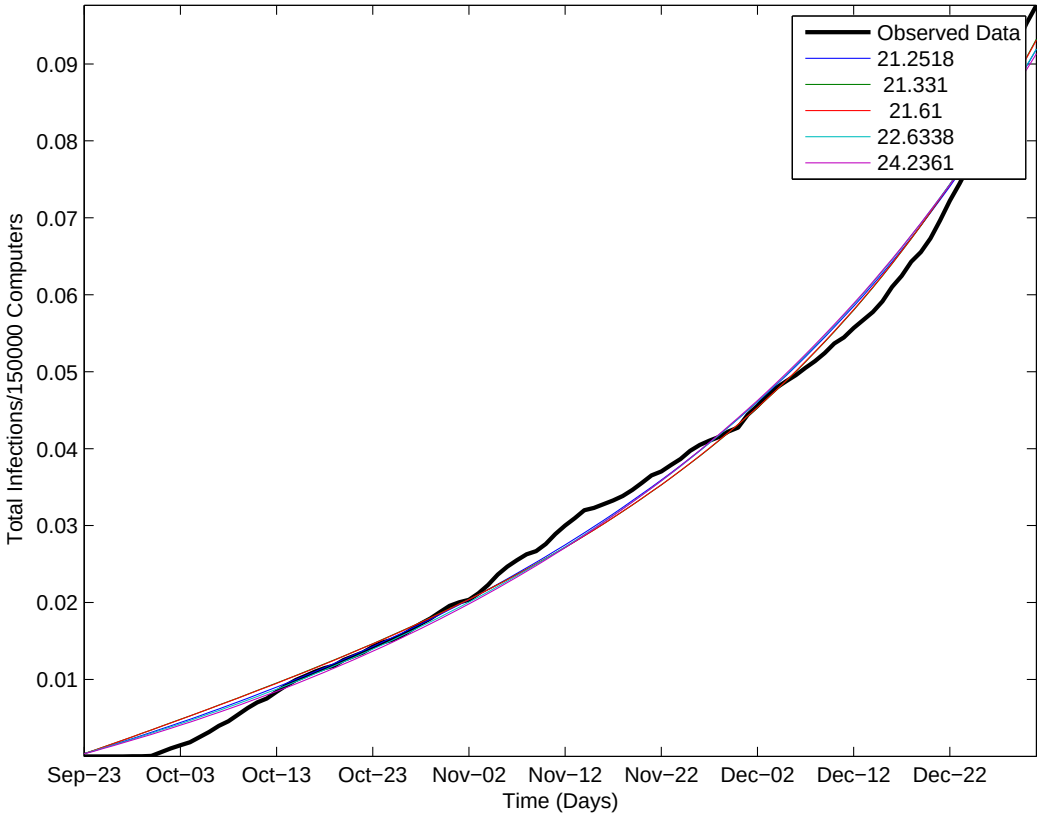

Figure 26: Top five model fits for the cumulative sum of observed attacks on the 3rd ranked targeted online entity

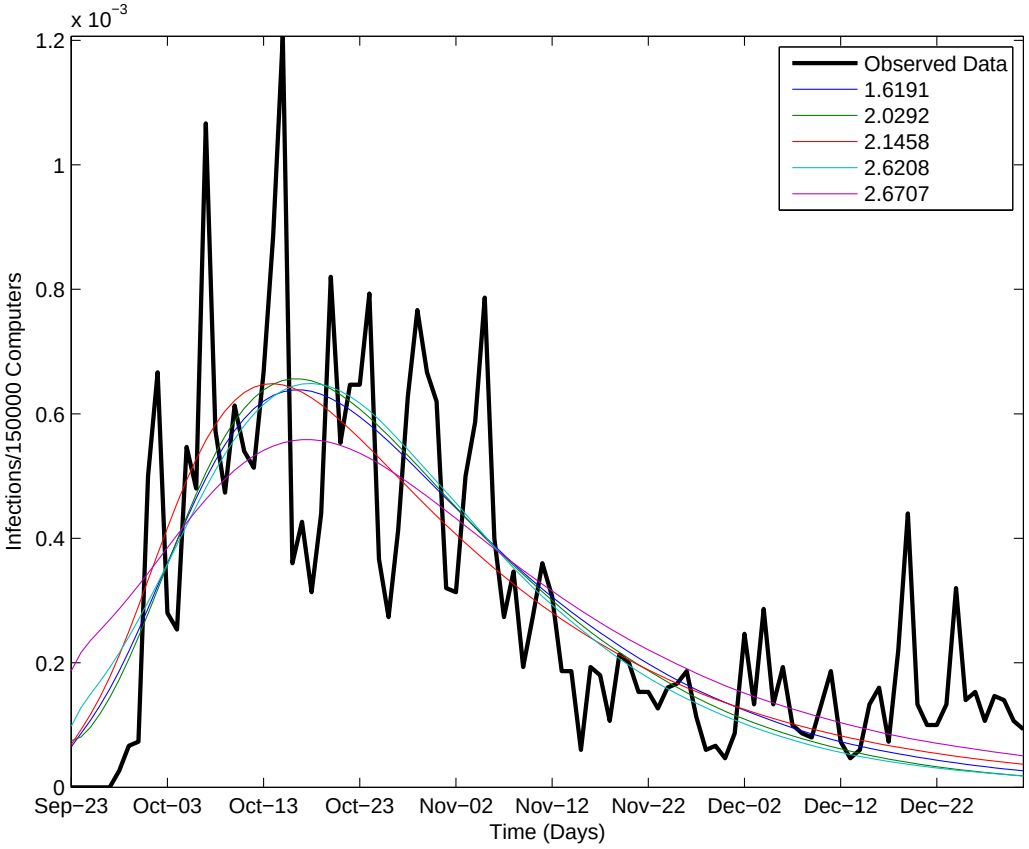

Figure 27: Top five model fits for the 4th ranked targeted online entity

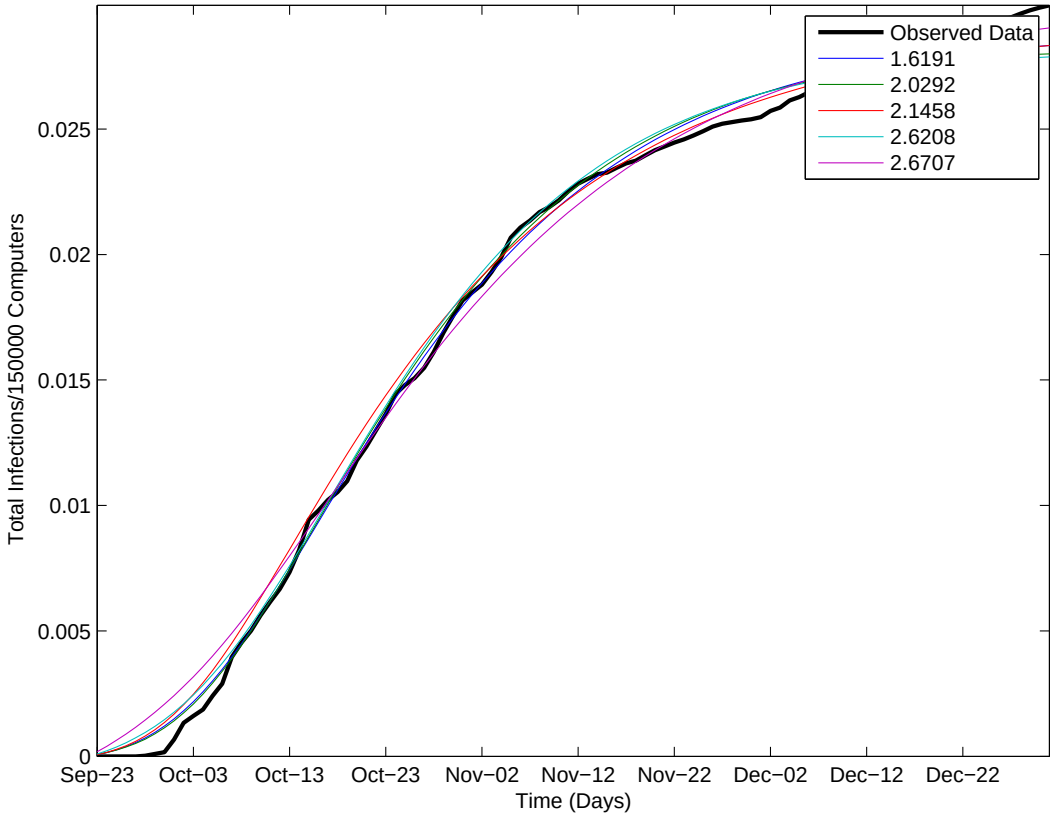

Figure 28: Top five model fits for the cumulative sum of observed attacks on the 4th ranked targeted online entity

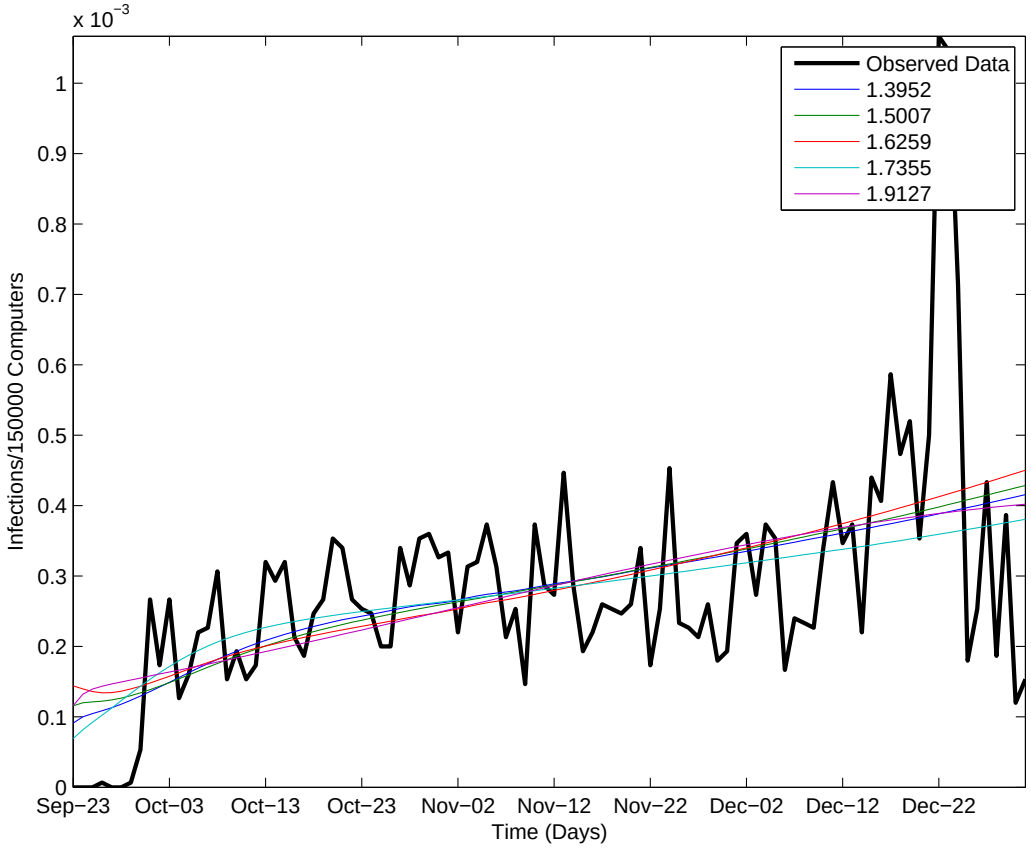

Figure 29: Top five model fits for the 5th ranked targeted online entity

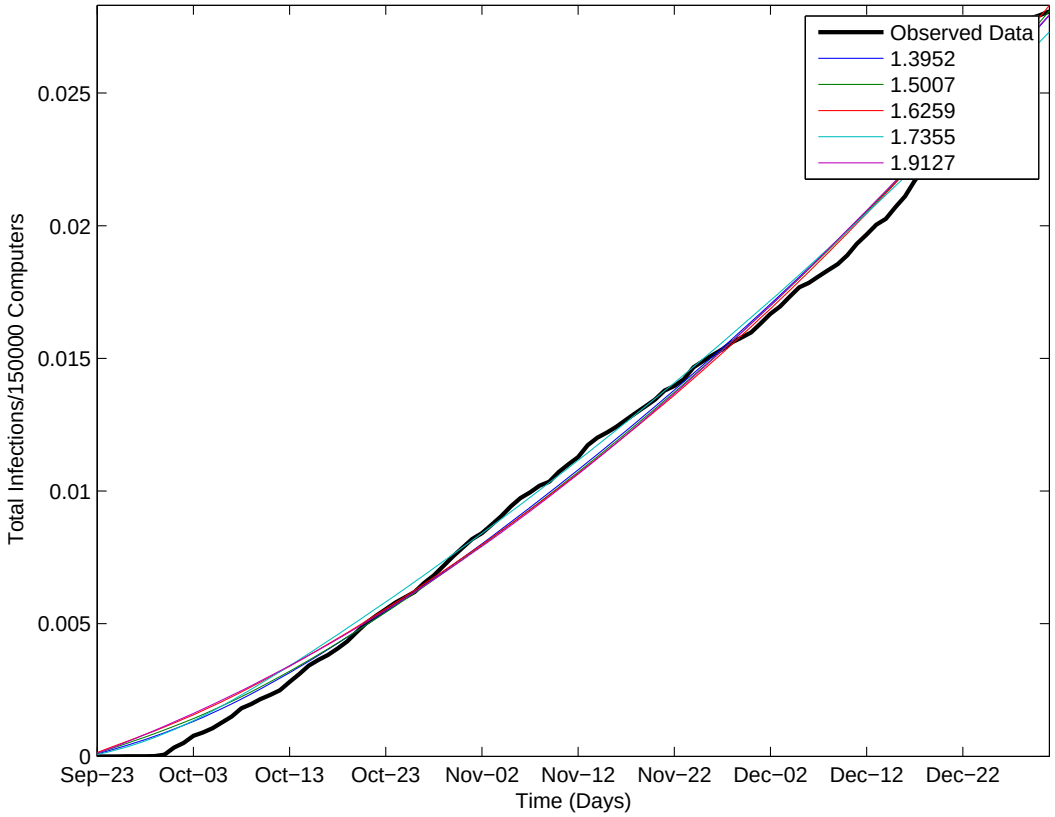

Figure 30: Top five model fits for the cumulative sum of observed attacks on the 5th ranked targeted online entity

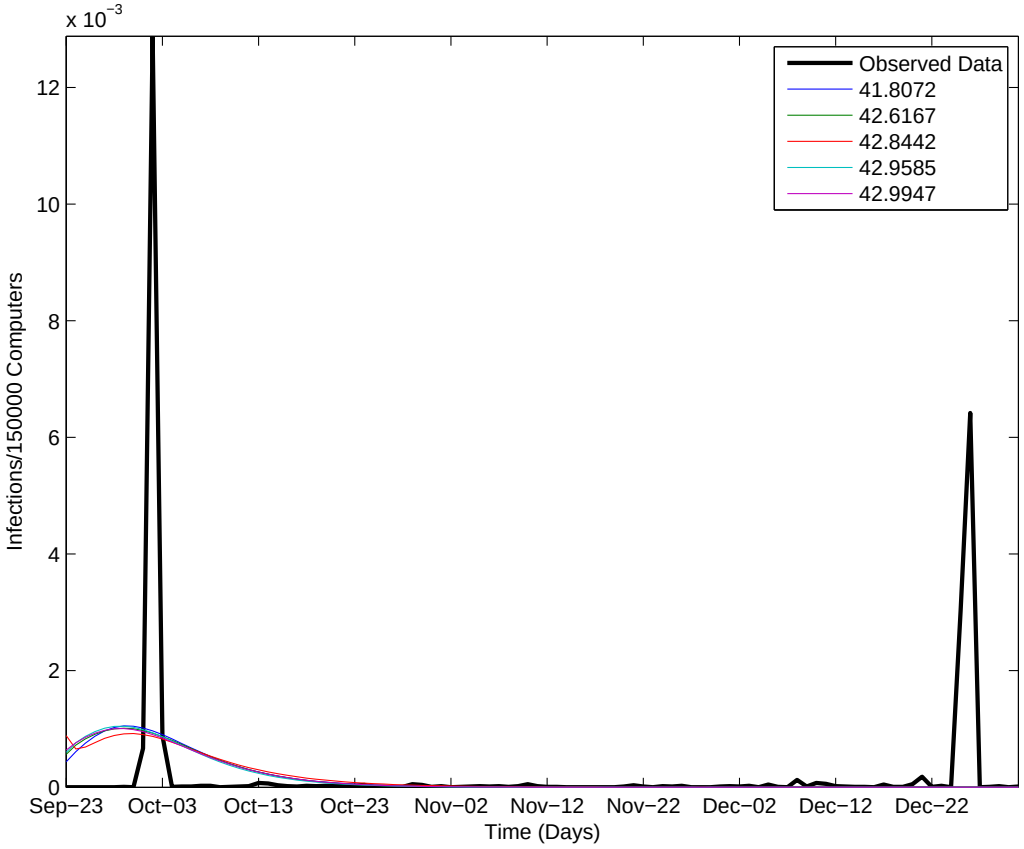

Figure 31: Top five model fits for the 6th ranked targeted online entity

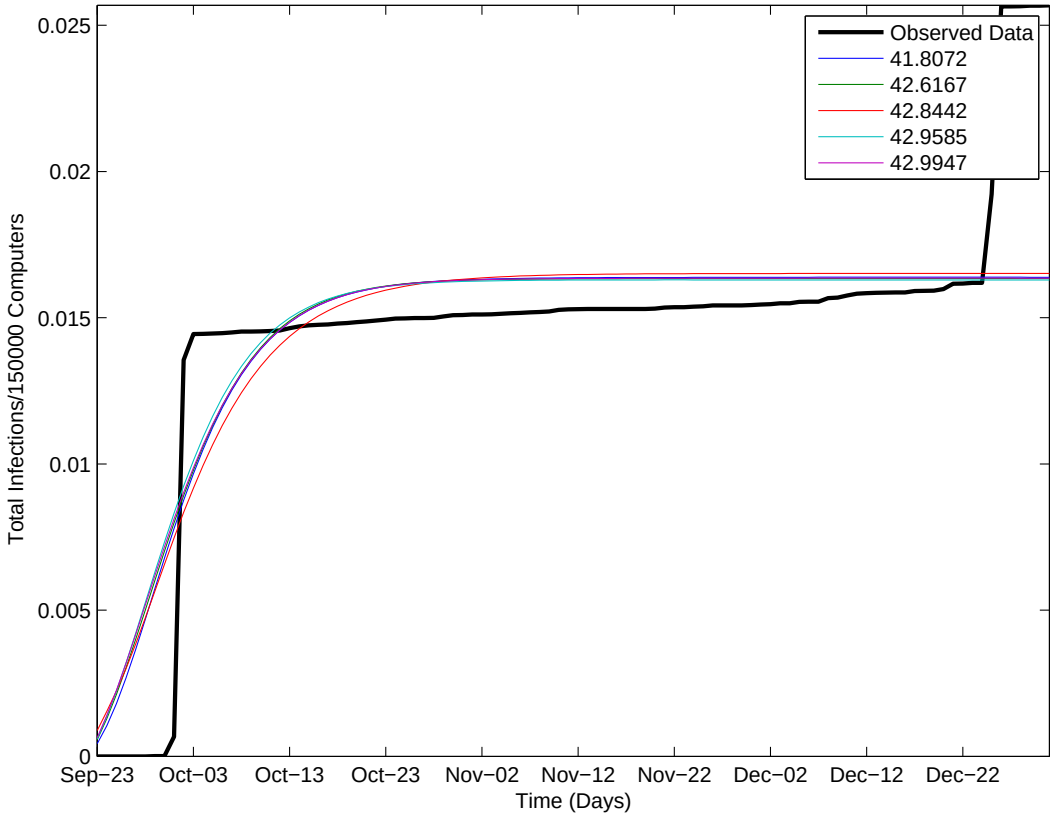

Figure 32: Top five model fits for the cumulative sum of observed attacks on the 6th ranked targeted online entity

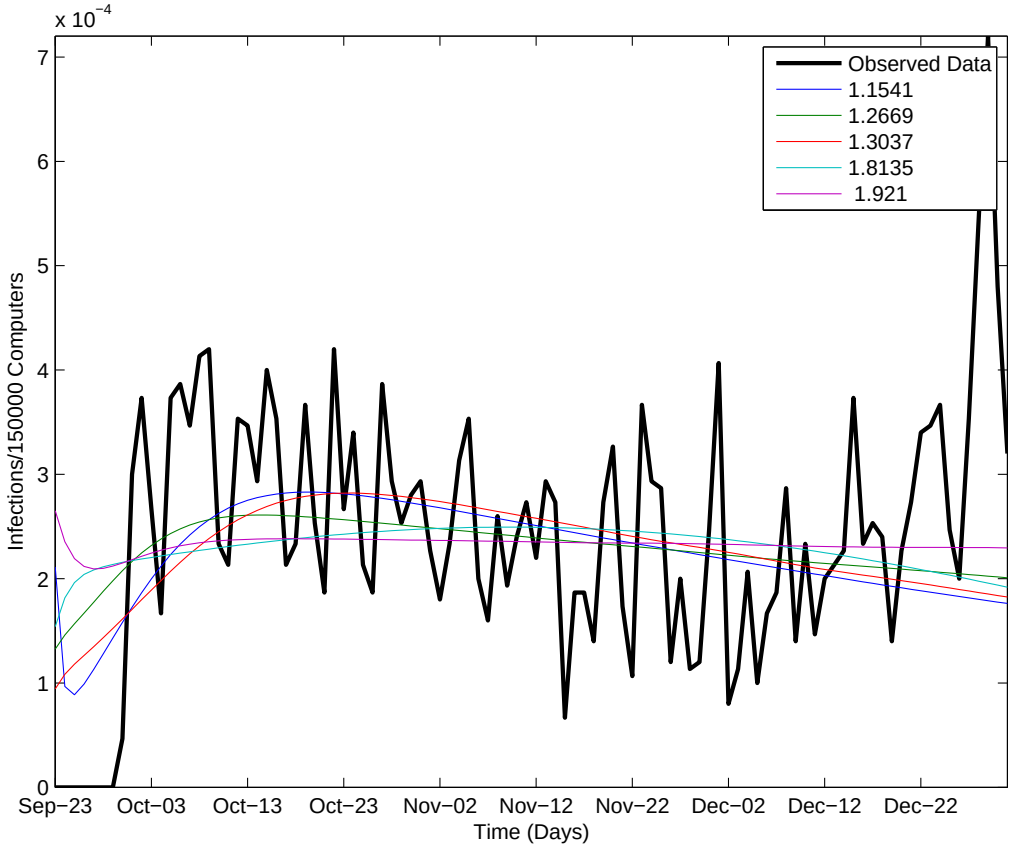

Figure 33: Top five model fits for the 7th ranked targeted online entity

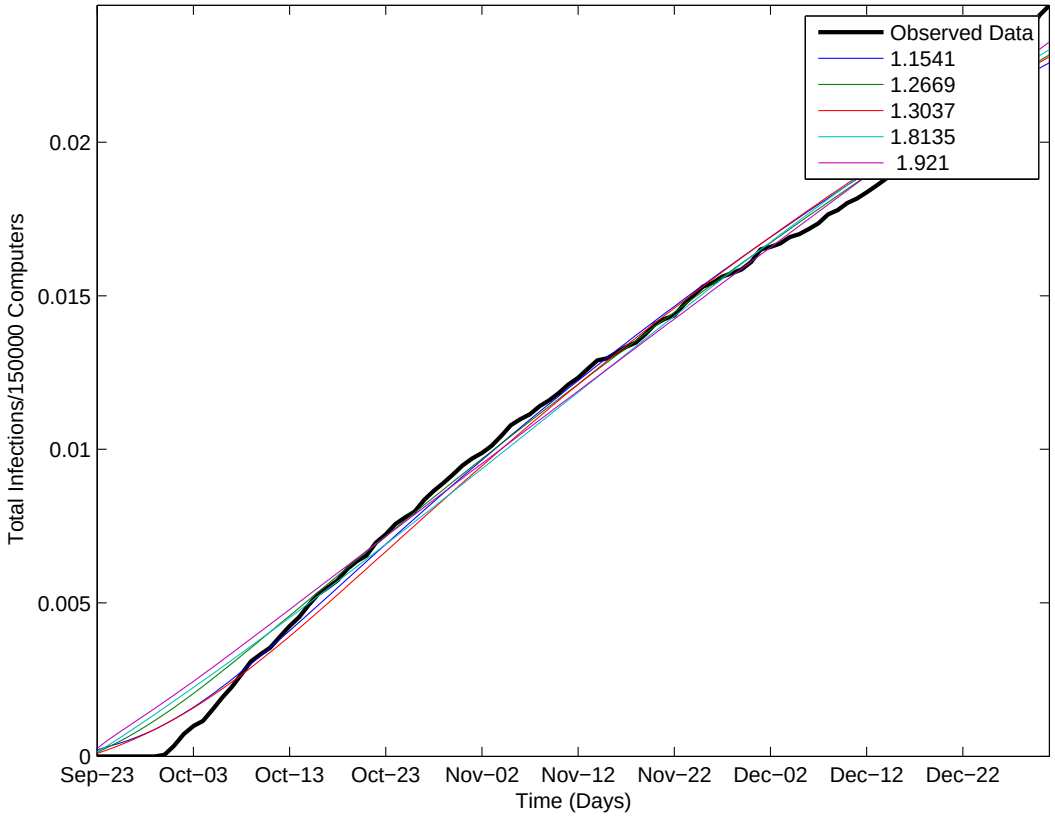

Figure 34: Top five model fits for the cumulative sum of observed attacks on the 7th ranked targeted online entity

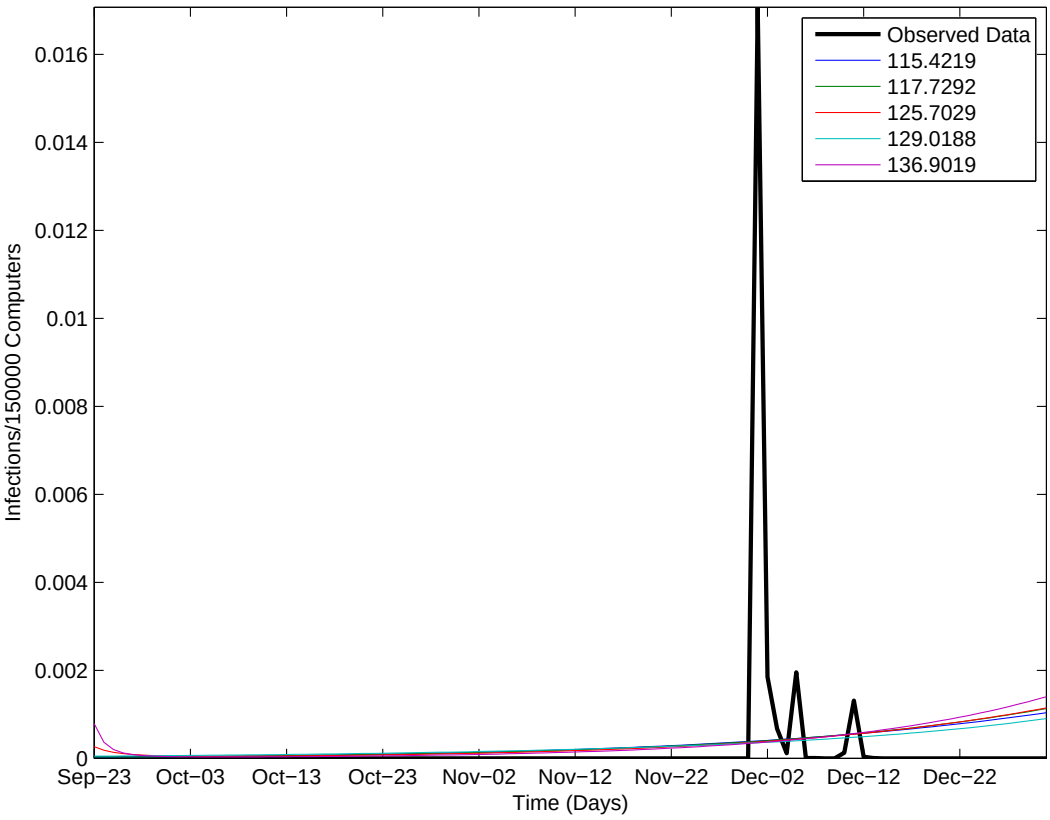

Figure 35: Top five model fits for the 8th ranked targeted online entity

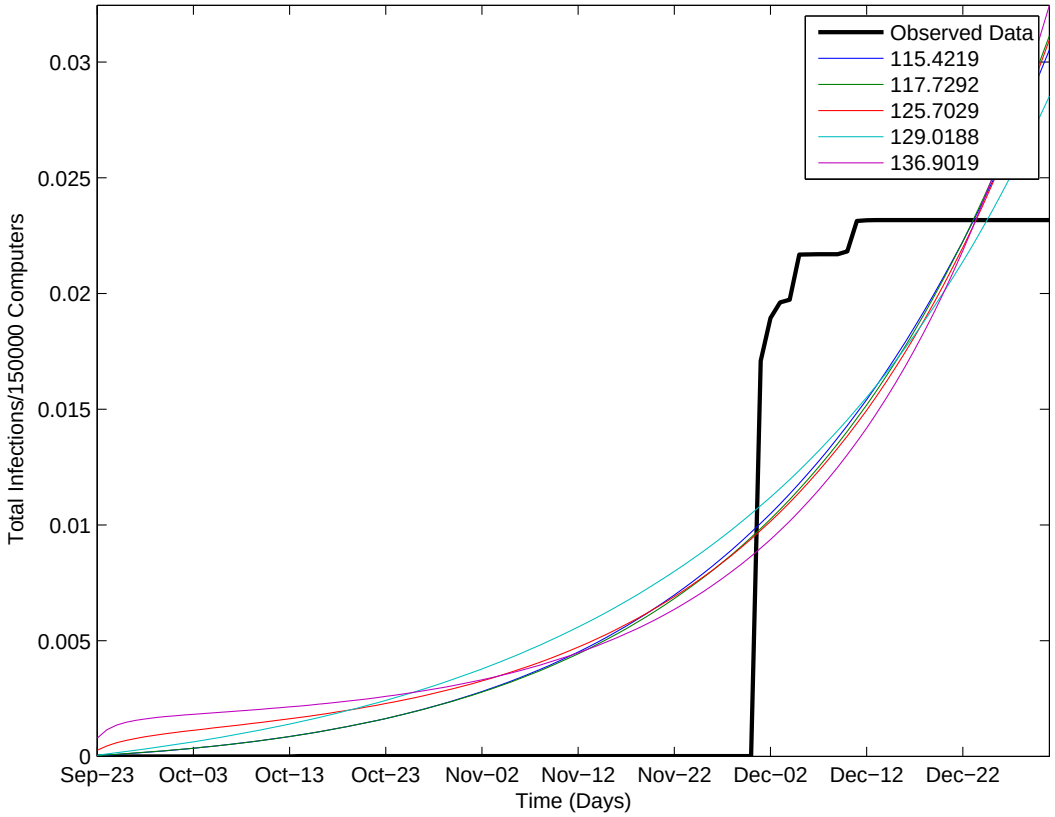

Figure 36: Top five model fits for the cumulative sum of observed attacks on the 8th ranked targeted online entity

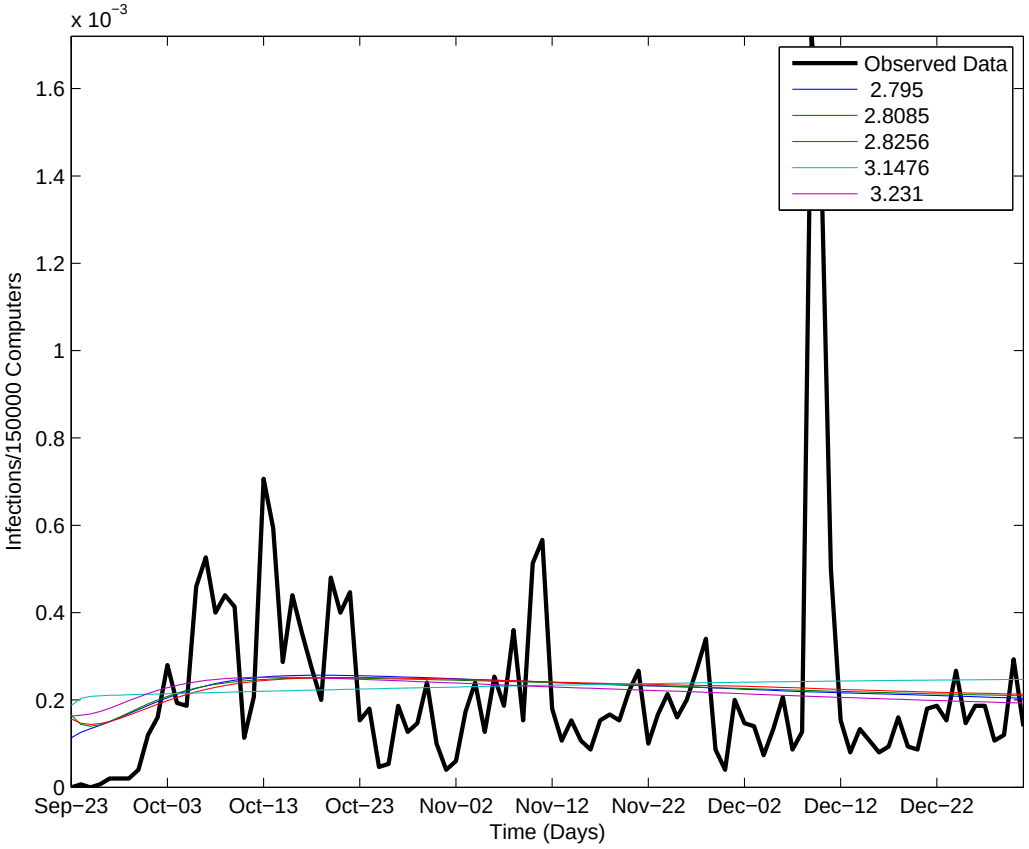

Figure 37: Top five model fits for the 9th ranked targeted online entity

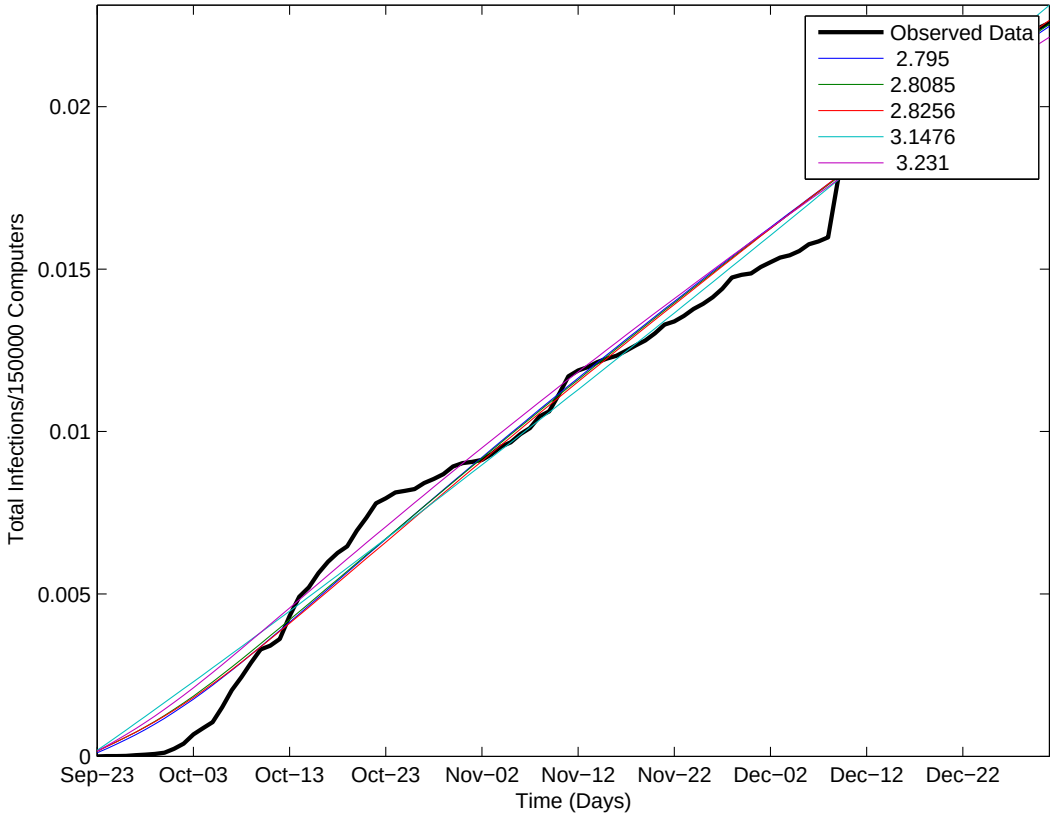

Figure 38: Top five model fits for the cumulative sum of observed attacks on the 9th ranked targeted online entity

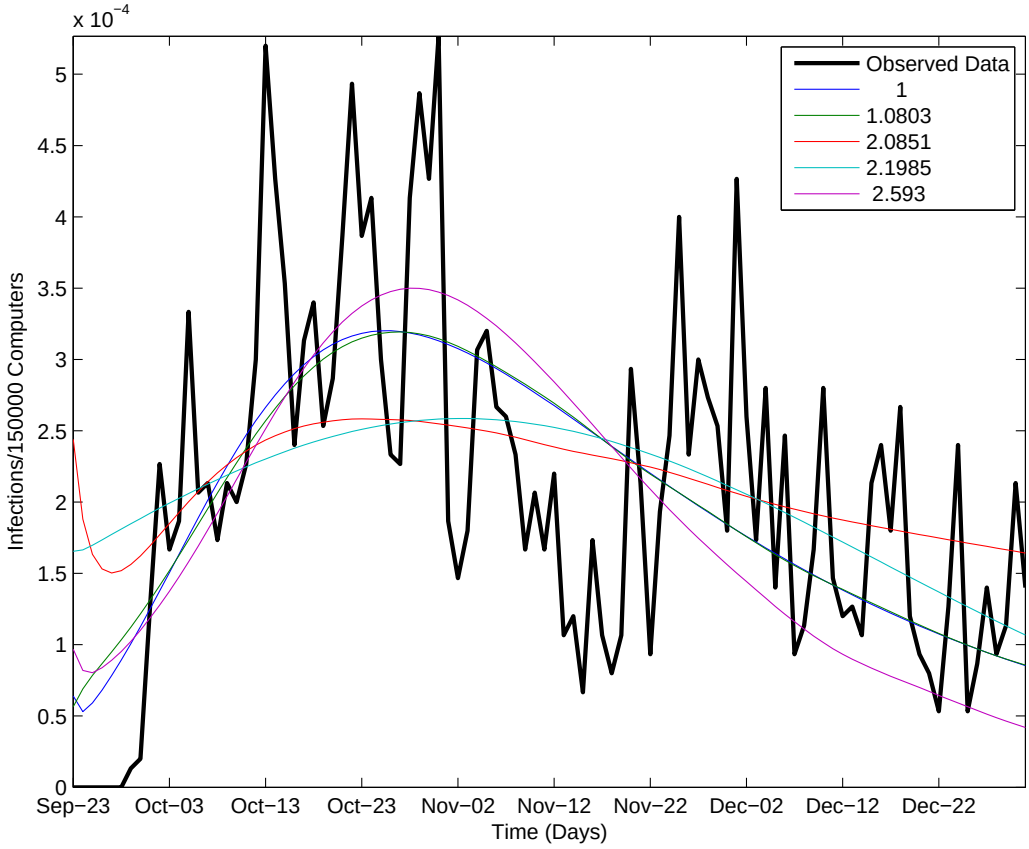

Figure 39: Top five model fits for the 10th ranked targeted online entity

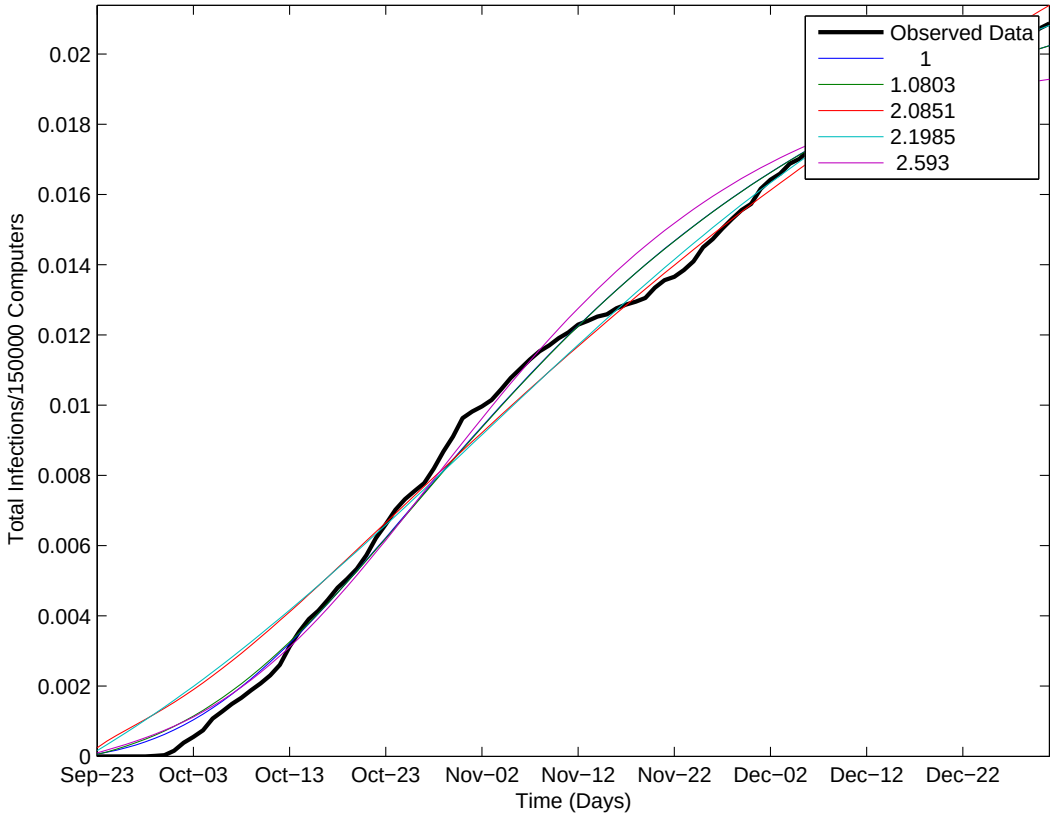

Figure 40: Top five model fits for the cumulative sum of observed attacks on the 10th ranked targeted online entity

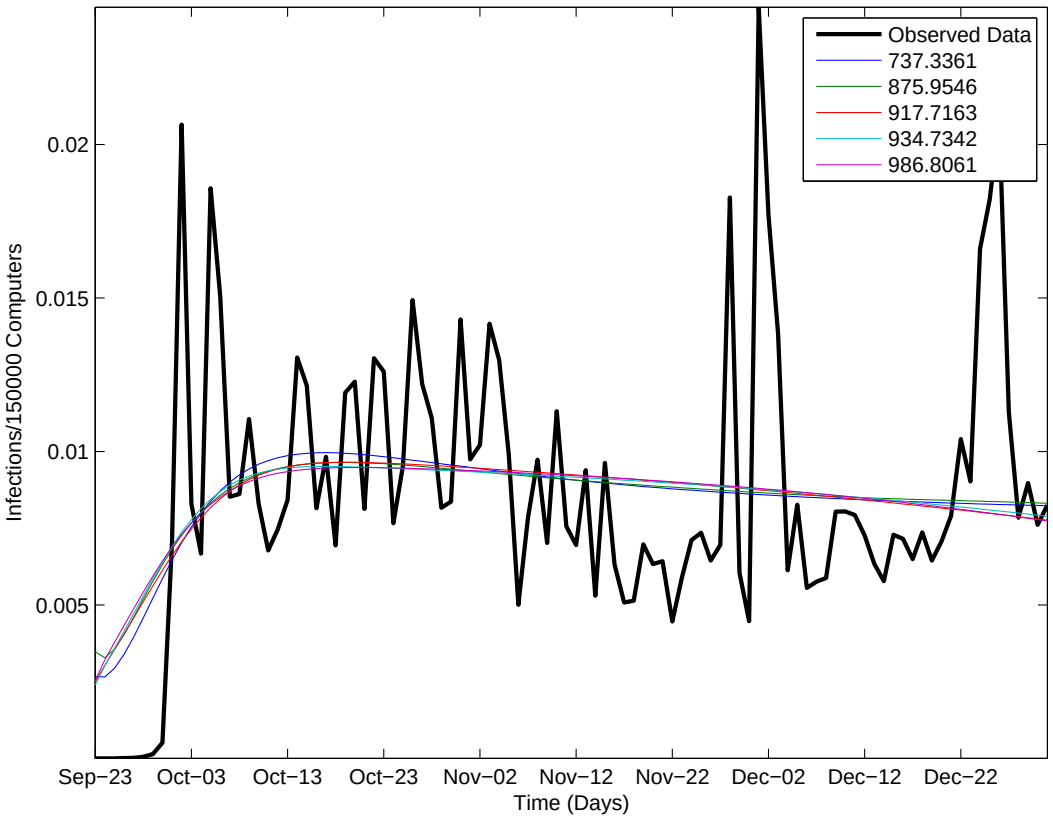

Figure 41: Top five model fits for the total attacks against the top ten targeted online entities.

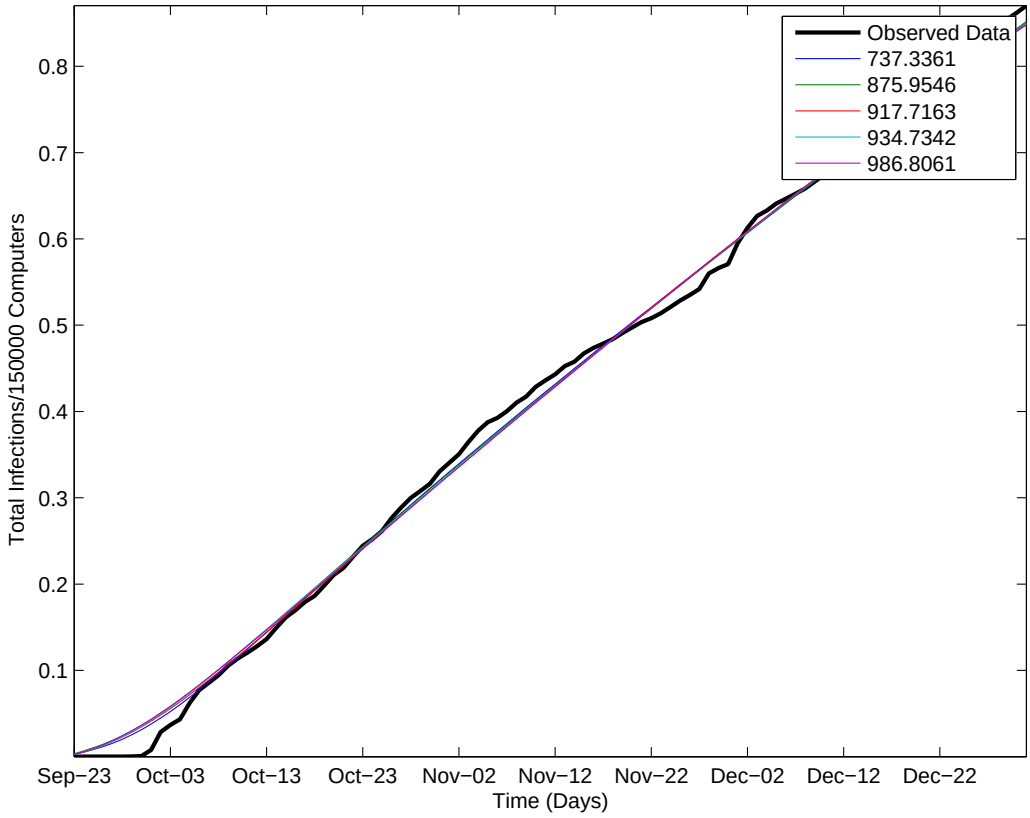

Figure 42: Top five model fits for the total cumulative sum of observed attacks on the top ten targeted online entities