PLOS | ONE

RESEARCH ARTICLE

Hybrid Epidemics—A Case Study on Computer Worm Conficker

Changwang Zhang[1,2]*, Shi Zhou[1]*, Benjamin M. Chain[3]

1 Department of Computer Science, University College London, London, United Kingdom, **2** Security Science Doctoral Research Training Centre, University College London, London, United Kingdom, **3** Division of Infection and Immunity, University College London, London, United Kingdom

* changwang.zhang.10@ucl.ac.uk (CZ); s.zhou@ucl.ac.uk (SZ)

Abstract

Conficker is a computer worm that erupted on the Internet in 2008. It is unique in combining three different spreading strategies: local probing, neighbourhood probing, and global probing. We propose a mathematical model that combines three modes of spreading: local, neighbourhood, and global, to capture the worm's spreading behaviour. The parameters of the model are inferred directly from network data obtained during the first day of the Conficker epidemic. The model is then used to explore the tradeoff between spreading modes in determining the worm's effectiveness. Our results show that the Conficker epidemic is an example of a critically hybrid epidemic, in which the different modes of spreading in isolation do not lead to successful epidemics. Such hybrid spreading strategies may be used beneficially to provide the most effective strategies for promulgating information across a large population. When used maliciously, however, they can present a dangerous challenge to current internet security protocols.

Introduction

Epidemic spreading phenomena exist in a wide range of domains [1, 2]. Well-known examples include disease spreading [3–5], computer worm proliferation [6–8], and information propagation [9–11]. Modelling and understanding of such phenomena can have important practical values to predict and control real world epidemics [3–5, 12–15].

Some typical spreading mechanisms have been extensively studied, such as the fully-mixed spreading model and the network spreading model. Many epidemics are *hybrid* as they spread via two or more different mechanisms simultaneously. Previous work on hybrid epidemics has focused on what we call the *non-critically* hybrid epidemic, where at least one of the spreading mechanisms alone is able to cause an epidemic outbreak, and a mixture of mechanisms brings no advantage.

We are interested in the *critically* hybrid epidemic, where each spreading mechanism alone is unable to cause any significant spreading whereas the mixture of such mechanisms leads to a huge epidemic outbreak. Recently we proposed a model that explains the behaviour of critically hybrid epidemics, which incorporates two spreading mechanisms in the setting of a

metopopulation [16]. We demonstrated that it is indeed possible to have a highly contagious epidemic by mixing simple, ineffective spreading mechanisms. The properties of such epidemics are critically determined by the ratio at which the different spreading mechanisms are mixed, and usually there is an optimal ratio that leads to a maximal outbreak size.

In this paper we present a detailed analysis of a *real* hybrid epidemic—the Internet worm Conficker, which erupted on the Internet in 2008 and infected millions of computers. The worm is a hybrid epidemic as the code analysis [17] has revealed the worm applied three distinct spreading mechanisms: (1) global random spreading, (2) local network spreading, and (3) neighbourhood spreading. It is a critically hybrid epidemic because the first and second spreading mechanisms are highly ineffective if used alone, and the third mechanism, as we will show later, is most effective when mixed with the other two.

We introduce a mathematical model to describe the spreading behaviour of Conficker. Our study was based on measurement data provided by Center for Applied Internet Data Analysis (CAIDA)'s Network Telescope project [18, 19], which monitors Internet traffic anomalies. We proposed algorithms to extract Conficker–related features from the CAIDA data. Then we infer the values of our model's parameters that characterise the worm.

We evaluated our inference results by comparing theoretical predictions with the actual measurement results. Our predictions closely reproduced the outbreak process of Conficker. We then explored possible spreading scenarios based on simulations using different values of parameters. One of the interesting results was that we showed the worm could spread faster, reach a larger outbreak size or survive for longer time by just revising the ratios at which the worm allocated its time on each of the spreading mechanisms (while keeping everything else the same), which can be easily achieved by changing a few lines in its coding.

This paper's contributions are two fold. Firstly, we present the first study on a real-life critically hybrid epidemic, where the epidemic's parameter values are inferred from measurement data. Secondly, we analyse the complex interactions among Conficker's three spreading mechanisms, and show that the worm can be more contagious if it mixes its three spreading mechanisms in an optimal way.

Background

Epidemic spreading mechanisms

A number of epidemic spreading mechanisms have been extensively studied [20, 21]. For example, in the *fully-mixed* spreading models [20, 22], a node is connected to all other nodes in a population, thus an epidemic can potentially spread between any two nodes according to a probability. Whereas in the *network* spreading models [1, 2, 20, 23], nodes are connected to their neighbours via a network structure, therefore an epidemic can only spread along the connections among nodes. Recent network-based models considered additional physical properties such as location-specific contact patterns [24, 25], human mobility patterns [26–29] and spatial effects [30–33].

Hybrid epidemics

Many epidemics are *hybrid* in the sense that they spread via two or more spreading mechanisms simultaneously. A hybrid epidemic can use fully-mixed spreading and network spreading, or use fully-mixed spreading but at two or more different levels, e.g. at the *global* level covering the whole population or at the *local* level consisting of only a part of the population.

There are many real examples. Mobile phone viruses can spread via Bluetooth communication with any nearby devises (local, fully-mixed spreading) and Multimedia Messaging Service with remote contacts (global, network spreading) [27]. A computer that is infected by the

worm Red Code II spends 1/8 of its time probing any computers on the Internet at random (global, fully-mixed spreading) and the rest of the time probing computers located in local area networks (local, fully-mixed spreading) [34]. Today information is propagated in society via mass media (TV, newspaper, posters) as well as online social media (Facebook, Twitter and emails). Mass media (global, fully-mixed spreading) can potentially deliver the information to a big audience, but the effectiveness of information transmission at an individual level may be small (for example, its ability to alter the target individuals behaviour). In contrast, social media (local, network spreading) may have little or no access to the majority of people who are not connected to the local group, but they provide rapid penetration of a selected target group with higher effectiveness.

It is clear that hybrid epidemics are much more complex than simple epidemics. Their behaviour is affected not only by multiple spreading mechanisms that they use, but also by the population's overlaid structure on which they spread. Studying hybrid epidemics may provide crucial clues for better understanding of many real epidemics.

Previous works on hybrid epidemics

Hybrid epidemics were initially studied as two levels of mixing in a population where nodes are mixed at both local and global levels [35]. Recently hybrid epidemics were studied as two levels of mixing in a network [36–38], in structured populations [39], in structured households [40–42], and in a meta-population which consists of a number of weakly connected sub-populations[43–48]. Studies of epidemics in clustered networks [49–51] are also relevant to the hybrid epidemics.

These previous works focused on analysing how a network's structure affects hybrid spreading. And most of them studied the *non-critically* hybrid epidemics, where at least one of the two spreading mechanisms alone can cause an infection outbreak and therefore the mix of two mechanisms is not a necessary condition for an epidemic outbreak. In this case, a hybrid epidemic using two spreading mechanisms is often less contagious than an epidemic using only one of the mechanisms. [36, 52].

Our recent study on critically hybrid epidemics

We are interested in the *critically* hybrid epidemics, where each of the spreading mechanisms alone is not able to cause any significant infection whereas a combination of the mechanisms can cause an epidemic outbreak. In this case, the mix of different spreading mechanisms is a critically condition for an outbreak (see Fig 1).

Recently we proposed a generic model to study the critically hybrid epidemics [16]. We considered an epidemic which spreads in a *meta-population* (consisting of many weakly connected *sub-populations*) using a mix of the following two typical spreading mechanisms. (1) Fully-mixed spreading on the global level, i.e. infection between any two nodes in the meta-population. (2) Network (or fully-mixed) spreading on the local level, i.e. infection between nodes within a sub-population where the internal topology of a sub-population is a network (or a fully-connected mesh). Each spreading mechanism has its own infection rate and an infected node recovers at a recovery rate. We define a parameter called the hybrid trade-off, α, as the proportion of time that the epidemic devotes to the first spreading mechanism (or the probability of using the first spreading mechanism in a time unit). Thus the proportion of time spent on the second mechanism is $(1 - \alpha)$.

Our mathematical analysis and numerical simulations based on the model highlight the following two results. Firstly, it is possible to mix two ineffective spreading mechanisms to produce a highly contagious epidemic, because the mix of the mechanisms can help to overcome
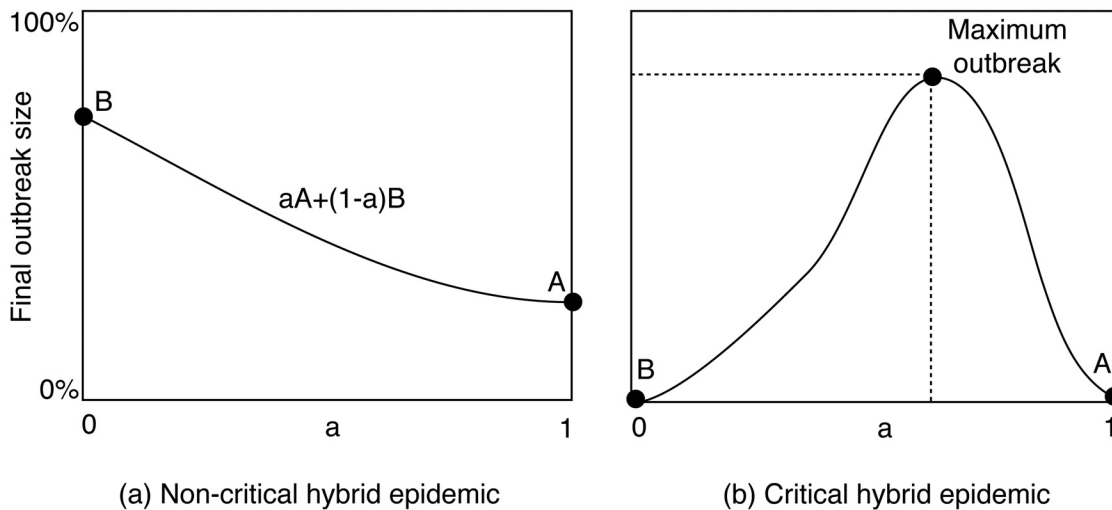
Fig 1. Hybrid epidemics, where two spreading mechanisms A and B are mixed at the ratio of α to $(1 - \alpha)$, where $0 \leq \alpha \leq 1$. (a) Non-critically hybrid epidemic, where at least one of the two mechanisms can cause an outbreak by its own (i.e. when $\alpha = 1$ or $\alpha = 0$). (b) critically hybrid epidemics, where each mechanism alone cannot cause any significant infection whereas a mix of them produces an epidemic outbreak. There exists an optimal α that produces the maximum outbreak.

the weakness of each mechanisms. Secondly, the threshold and the size of outbreak is critically determined by the hybrid trade-off α. We also provided an analytical prediction of the optimal trade-off for the maximum outbreak size.

Computer Worm Conficker

In this paper we will analyse a critically hybrid epidemic, the computer worm Conficker, based on real measurement data. It is one of the most contagious computer worms on record. It erupted on the Internet on 21 November 2008 and infected millions of computers in just a few days [7]. The worm's ability to spread to such a large number of computers in so short a time and the fact [53] that it is still active on the Internet has caused serious concern.

Each computer on the Internet is associated with an Internet Protocol (IP) address. Conficker views the Internet as a meta-population, where computers are located in sub-populations, i.e. Local Area Network (LAN) consisting of computers whose IP addresses share the same prefix. According to the computer security company Symantec [17], Conficker uses three spreading mechanisms (see Fig 2):

- *Global* spreading, where the worm probes computers with random IP addresses on the Internet;

- *Local* spreading, where the worm on an infected computer probes computers in the same Local Area Network (LAN) with the same IP address prefix;

- *Neighbourhood* spreading, where it probes computers in ten neighbouring LANs (with smaller consecutive IP address prefixes).

Previous research on Conficker has studied the geographical distribution of infected IP addresses, the distribution of probing packet size [7, 54, 55], and properties of the worm's global probing [56, 57]. The parameters of Conficker's hybrid spreading and how they affect the
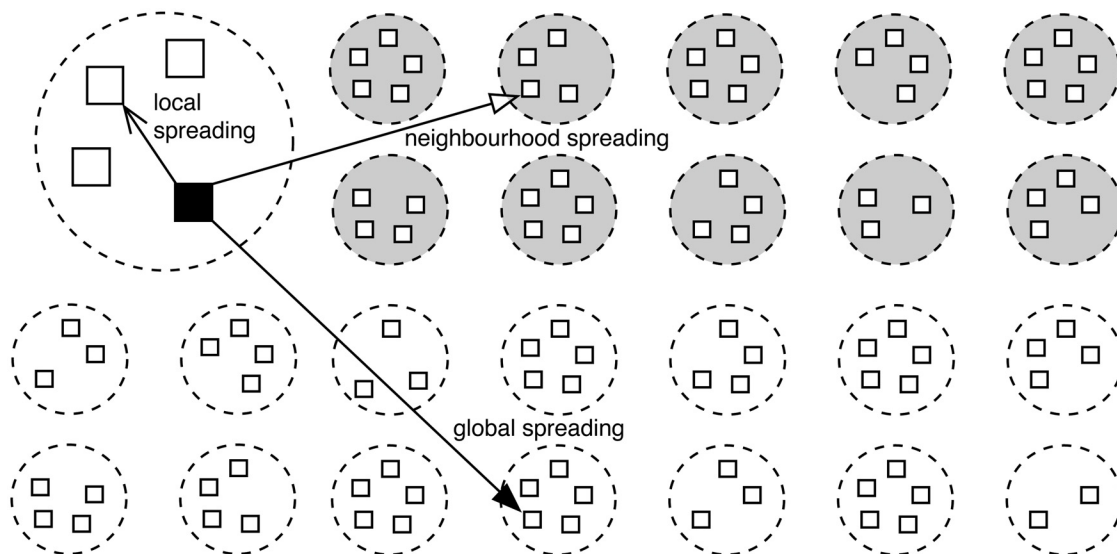
Fig 2. Conficker's three probing strategies. (1) global spreading, where it probes any computer on the Internet at random; (2) local spreading, where it probes computers in the same local network; (3) neighbourhood spreading, where it probes computers in ten neighbouring local networks.

epidemic dynamics of the worm can help explain why the worm is so contagious. But they have been hitherto little studied.

Our Model of Conficker

Here we use the notion of node to represent an IP address at which a computer (or computers) connects to the Internet. For convenience, we call a sub-population (or a LAN) a subnet. A node must be in one of three states: *susceptible*, *infected*, and *recovered*[20]. Initially only a small number of nodes are infected and all others are susceptible. At each time step, an infected node attempts to spread the worm to susceptible nodes using one of the three probing strategies:

- Global spreading with probability α_g, where the worm probes nodes on the Internet at random with the global infection rate $\beta_g \in [0, 1]$.

- Local spreading with probability α_l, where it probes nodes in the local subnet with the local infection rate $\beta_l \in [0, 1]$;

- Neighbourhood spreading with the probability α_n, where it probes nodes in ten neighbouring subnets with the neighbourhood infection rate $\beta_n \in [0, 1]$;

The mixing probabilities satisfy $\alpha_g + \alpha_l + \alpha_n = 1$.

An infected node is recovered with recovery rate $\gamma \in [0, 1]$. A recovered node remains recovered and cannot be infected again. Note that for mathematical analysis, the mixing probabilities could be incorporated into the infection rates. But we have treated them as separate parameters, considering that an infection rate reflects inherent properties of a computer worm in the context of a specific target population, whereas mixing probabilities are settings that can be easily modified in the worm's code. This is also the reason we use the mixing probabilities as controlling parameters in our study below and keep other parameters the same.

Only nodes that can potentially be infected by Conficker are relevant to our study. We call them the *relevant* nodes. A subnet is relevant if it contains at least one relevant node. Irrelevant nodes include unused IP addresses and those computers that do not have the vulnerabilities that the worm can exploit. Note that although the irrelevant nodes and subnets do not participate in the spreading of Conficker, they will be probed by the worm as the worm does not have the priori knowledge about which nodes are vulnerable.

Let n represent the total number of relevant nodes and N the number of relevant subnets. The average number of relevant nodes in a subnet is $n_N = n/N$. Let N^+ represent the average number of relevant subnets in ten neighbouring subnets.

At time t, the total number of susceptible, infected, and recovered nodes are $S(t)$, $I(t)$, and $R(t)$, respectively. Then the average number of infected nodes in a subnet is $I_N(t) = I(t)/N$, and the average number of infected nodes in ten neighbouring subnets is $I^+(t) = I_N(t)N^+$. Hence on average a susceptible node can be infected via (1) global probing by $I(t)$ infected nodes in the Internet; (2) local probing by $I_N(t)$ infected nodes in the local subnet; (3) neighbourhood probing by $I^+(t)$ infected nodes in the neighbouring subnets.

The average probabilities that a susceptible node is *not* infected by the global, local and neighbourhood probing, respectively, are:

$$\begin{aligned}
P_g(t) &= (1 - \alpha_g \beta_g)^{I(t)} \\
P_l(t) &= (1 - \alpha_l \beta_l)^{I_N(t)} \\
P_n(t) &= (1 - \alpha_n \beta_n)^{I^+(t)}.
\end{aligned} \tag{1}$$

The average probability of *not* being infected by any probing is $P(t) = P_g(t)P_l(t)P_n(t)$. Thus the discrete evolution of Conficker spreading can be described as:

$$\begin{aligned}
S(t+1) &= S(t) - S(t)[1 - P(t)] \\
I(t+1) &= I(t) + S(t)[1 - P(t)] - \gamma I(t) \\
R(t+1) &= R(t) + \gamma I(t)
\end{aligned} \tag{2}$$

where $S(t)[1 - P(t)]$ is the number of new infections at t.

Inferring Conficker Parameters From Data

We infer the parameter values of our Conficker model from the Internet measurement data [18, 19] collected by the Center for Applied Internet Data Analysis (CAIDA) in 2008. This is the only publicly available dataset that has captured the *initial* outbreak process of the worm. The CAIDA Network Telescope project [18, 19] monitors Internet traffic sent to a large set of *unusable* IP addresses, which account for around 1/256 of all addresses. No legitimate traffic should be sent to these monitored addresses because they are not allocated for normal usage [58]. Thus the traffic data captured by this project provides a good view on various abnormal behaviours on the Internet.

When Conficker spreads on the Internet, its global spreading mechanism sends out probing packets to randomly generated IP addresses, some of which are unused IP addresses and therefore are monitored by the Network Telescope project. Conficke's probing packets are characterised by the Transmission Control Protocol (TCP) with destination port number 445. This feature can be used to distinguish Conficker packets from other packets in the Network Telescope data.

For each record of Conficker's probing packet, we are interested in two things: (1) the time when the packet is monitored by the Network Telescope project, and (2) the packet's source IP
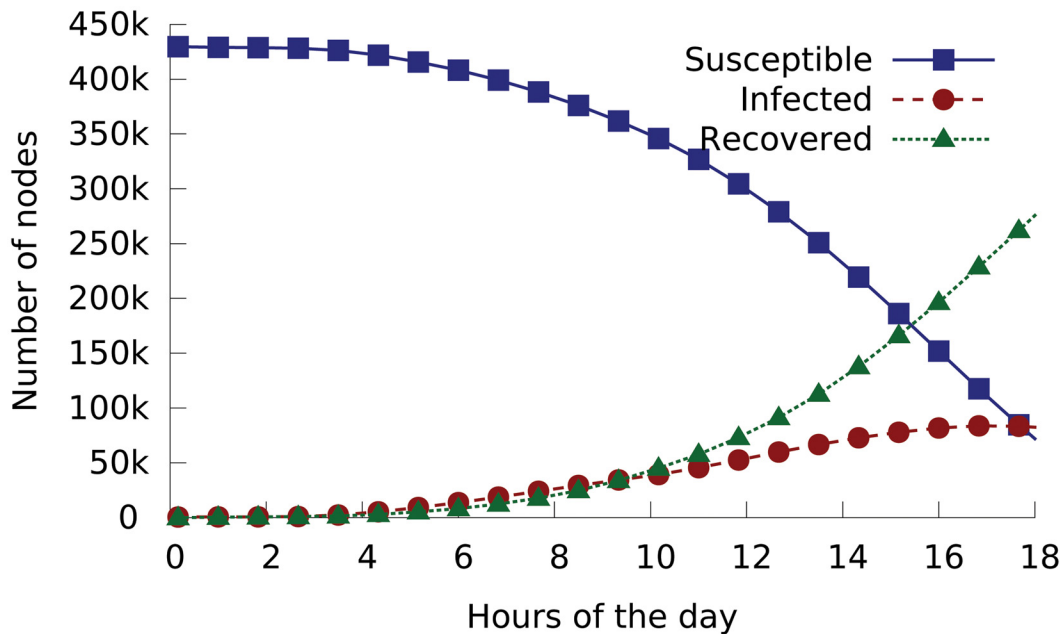
Fig 3. Numbers of susceptible nodes $S(t)$, infected nodes $I(t)$ and recovered nodes $R(t)$ as a function of time t, as inferred from CAIDA's dataset on 21/Nov/2008, the day of Conficker's outbreak.

address, which gives the location of a Conficker-infected node. We ignore the destination address, as it is a randomly-generated, unused IP address.

We study the Network Telescope project's daily dataset collected on November 21, 2008, the day when Conficker broke out on the Internet. We use two earlier datasets collected on November 12 and 19, 2008 to filter out background 'noise' that has been happening before the outbreak. That is, in the outbreak dataset, we discard packets that were sent from any source address that had already sent packets to any of the unusable addresses in the two earlier datasets. We use the prefix of /24 (i.e. IP address mask of 255.255.255.0) to distinguish different subnets [7]. Our analysis uses a 10-minute window.

Step One: Inferring node status at a given time

We first infer the status of each node at time t from the CAIDA data. On the day of Conficker outbreak, all relevant nodes were initially susceptible. In the analysis, we assume a node is just infected by the worm when we observe the first Conficker probing packet coming from it; and the node is recovered when we observe its last probing packet before the end of the day. Fig 3 shows the number of susceptible, infected and recovered nodes as observed in a 10-minute window.

Step Two: Inferring new infections caused by each spreading mechanism

Let $dI_l(t)$, $dI_n(t)$ and $dI_g(t)$ represent the numbers of nodes that are newly infected through local, neighbourhood and global spreading, respectively, at time step t. Our analysis on the data shows that 84% of new infections occurred within already infected subnets or their

neighbourhood subnets, i.e. only 16% of new infections appeared outside the reach of local and neighbourhood probing. This agrees with our understanding that local and neighbourhood probing are significantly more effective than global probing [7]. And 73% of those new infections within the reach of local and neighbourhood probing (i.e. 73%×84% of all new infections) occurred in already infected subnets. This indicates the local probing is more effective than neighbourhood probing. Based on the above analysis we can then approximately identify the probing mechanism that is responsible for a newly infected node by analysing the states of other relevant nodes at the time when the new infection happens.

- IF there is an infected node already in the same subnet, the new infection is caused by that infected node via local spreading.

- ELSE IF there is an infected node in the ten neighbouring subnets, then the new infection is via neighbourhood spreading.

- OTHERWISE, the newly infected node is infected via global spreading.

Fig 4 shows the inferred results, plotting the number of new infections caused by each spreading mechanism as a function of time.

Step Three: Inferring parameters of the Conficker model

From the above result, we can calculate $I_N(t)$ and $I^+(t)$ according to their definitions. Then we calculate the average probabilities that a susceptible node at time t is *not* infected by the local, neighbourhood and global probing, respectively, as:

$$P_l(t) = 1 - \frac{dI_l(t)}{S(t)}, P_n(t) = 1 - \frac{dI_n(t)}{S(t)}, P_g(t) = 1 - \frac{dI_g(t)}{S(t)}. \tag{3}$$
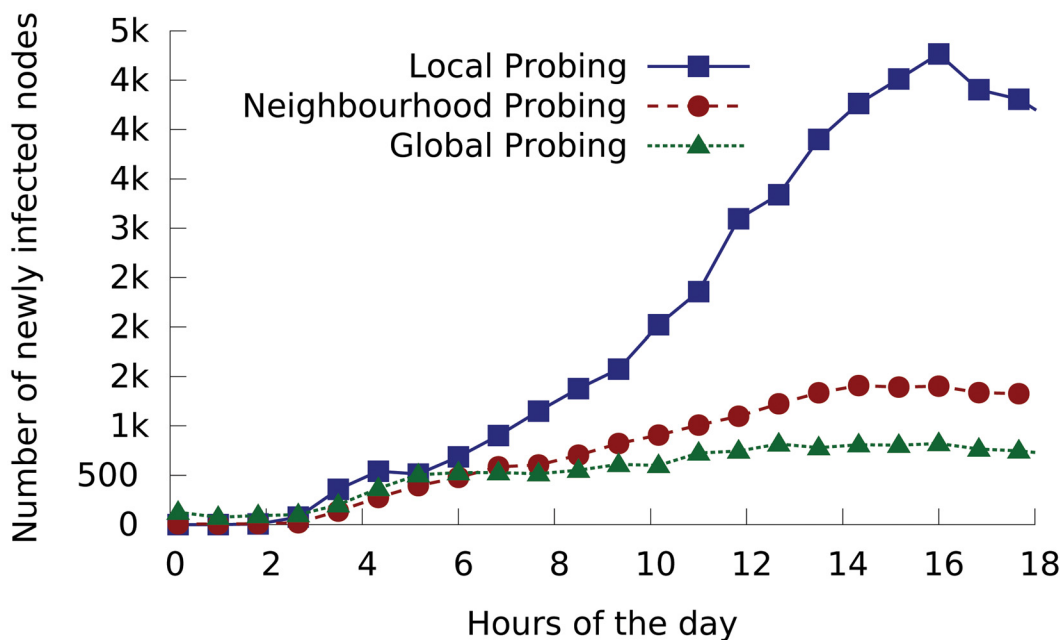

Fig 4. Numbers of nodes newly infected by Conficker via each of the three spreading mechanisms in 10-minute windows on the day of Conficker's outbreak, as inferred from CAIDA's dataset on 21/Nov/2008.

doi:10.1371/journal.pone.0127478.g004

Table 1. Conficker parameters inferred from data[1].

Global spreading	$\alpha_g = 89.1\%$	$\beta_g = 7.7\times10^{-8}$
Local spreading	$\alpha_l = 5.3\%$	$\beta_l = 0.32$
Neighbourhood spreading	$\alpha_n = 5.6\%$	$\beta_n = 0.032$
Recovery rate	$\gamma = 0.064$	
Recovery time	$\tau = 156$ mins	
Probing frequency	$\lambda = 82.5$	

[1] All parameters are measured in a 10-minute window.

doi:10.1371/journal.pone.0127478.t001

We define the effective infection rate as $b_l = \alpha_l \beta_l$, $b_n = \alpha_n \beta_n$ and $b_g = \alpha_g \beta_g$. Then we can use Eq 1 and Eq 3 to calculate b_l b_n and b_g.

Let λ denote the average total number of probes an infected node conducts during each time step. Then the average number of local, neighbourhood, and global probes in a time step are respectively $\alpha_l \lambda$, $\alpha_n \lambda$, and $\alpha_g \lambda$. The number of nodes (relevant and irrelevant) probed by the local, neighbourhood and global probing are 256, 10×256 and 2^{30} (it is not 2^{32} due to a bug in the worm's random number generation algorithm [17]). We can express the effective infection rates as:

$$b_l = \alpha_l\lambda/256, \quad b_n = \alpha_n\lambda/2560, \quad b_g = \alpha_g\lambda/2^{30}. \tag{4}$$

By solving Eq 4 together with $\alpha_g+\alpha_l+\alpha_n = 1$, we can obtain λ, α_l, α_n and α_g. Then we can obtain the infection rates of three spreading mechanisms as $\beta_l = b_l/\alpha_l$, $\beta_n = b_n/\alpha_n$, and $\beta_g = b_g/\alpha_g$. The recovery rate can be calculated as $\gamma = dR(t)/I(t)$, where $dR(t) = R(t + 1) - R(t)$.

Inference results and evaluation

The inferred values of the Conficker model parameters are shown in Table 1, including the mixing probability α and the infection rate β for three spreading mechanisms, the recover rate γ, the recovery time $\tau = 1/\gamma$ which is the average time it takes for an infected node to recover, and the probing frequency λ. The parameter values are averaged over time windows between 4:00 and 16:00 when the spreading behaviour was stable. Computers are online and offline on a daily basis following a diurnal pattern [59]. We find that this factor only has a marginal impact on our results.

We observe in the data that the worm had infected in total $n = 430,135$ nodes, which were located in $N = 92,267$ subnets. On average, each subnet has $n_N = 4.7$ relevant nodes, and $N^+ = 4.3$ of ten neighbouring subnets are relevant.

With these parameter values, we can use our Conficker model (see Eq 2) to theoretically predict the worm's outbreak process. As measured from the data, the number of nodes in the three statuses were $S = 423,899$, $I = 3,945$, and $R = 2,291$ at 4:00am. Our prediction starts from 4.00am and uses these numbers as the initial condition. As shown in Fig 5, our model's predictions closely match the measurement data.

The inferred parameters are in agreement with our expectations. For example the local spreading has a high infection rate because if a computer is already infected, then other computers in the same subnet are likely to have a similar computer system and thus are also likely to be vulnerable to the worm. By comparison, global spreading has an extremely low infection rate. On average, more than 10 million global probings will produce only a single new infection. On average an infected node retains its status for 2.5 hours (156 mins) before it recovers
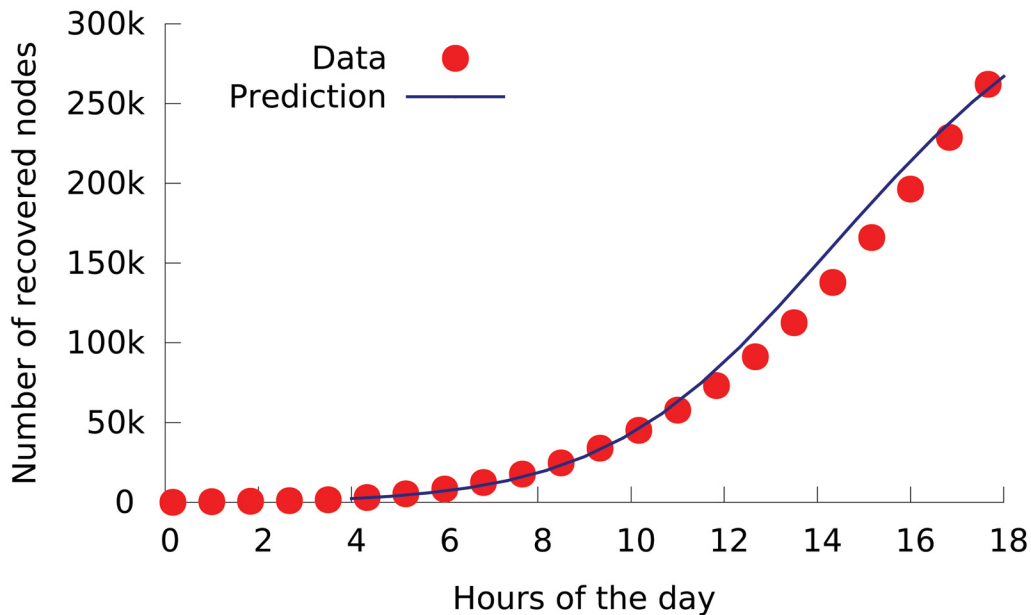
Fig 5. The outbreak of computer worm Conficker. Points are measured from Network Telescope's dataset collected on the outbreak day. Curve is theoretical prediction from our Conficker model using the inferred parameters.

(e.g. switched off or updated with new anti-virus database). The worm only sends out 8 probing packets per minute. Such a deliberately low probing rate helps the worm to evade a computer's or network's security systems.

Analysis on Conficker's Hybrid Spreading

Mix of TWO spreading mechanisms

We run simulations using our Conficker model with the parameter values inferred above. The simulation network has 100k subnets. Each subnet contains 5 relevant nodes and has 4 relevant adjacent subnets. This topology setting resembles Conficker's spreading network observed in the data. Initially two random nodes are infected. The only controlling parameter is the mixing probabilities of the spreading mechanisms. Simulation results on mix of two spreading mechanisms are shown in Fig 6.

Fig 6a shows that as explained above, global spreading or local spreading alone cannot cause an outbreak, whereas a mixture at a ratio of 0.8 to 0.2 produces a large and rapid outbreak. Fig 6b shows that the neighbourhood spreading alone ($\alpha_g = 0$) can cause a large, but very slow outbreak, whereas the mix of neighbourhood spreading with just a small amount of global spreading can dramatically accelerate the spreading process. Fig 6c shows that adding local spreading to neighbourhood spreading slows down the spreading process considerably. When they are mixed at the ratio of 0.8 to 0.2, the spreading reaches the same final outbreak size but the whole process lasts for the longest time.

Mix of THREE spreading mechanisms

Simulation results on mixing three spreading mechanisms are shown in Fig 7. Fig 7a shows it is not difficult to achieve a large final outbreak size when the three mechanisms are all present
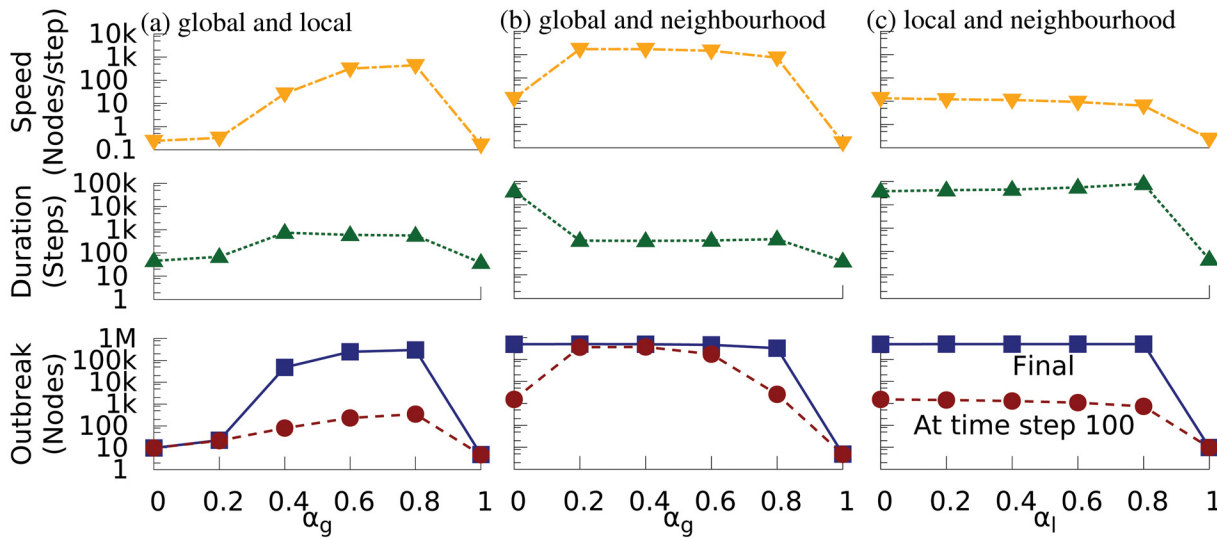
Fig 6. Simulation results for the mix of Conficker's two spreading mechanisms with different mixing probabilities. (a) Mix of global (α_g) and local (1 − α_g) mechanisms; (b) Mix of global (α_g) and neighbourhood (1-α_g) mechanisms; (c) Mix of local (α_l) and neighbourhood (1-α_l) mechanisms. In each case we measure the outbreak size, the total duration of the spreading, and the speed of spreading. The outbreak results include both the final outbreak size (square) and the outbreak size at time step 100 (filled circle). Each data point is averaged over 100 runs of a simulation. Note the y axes are all logarithmic.

doi:10.1371/journal.pone.0127478.g006

and neither local spreading nor global spreading is dominant. Fig 7b shows spreading will last for longer time if there is less global probing. Fig 7c shows that the most contagious variation of the worm is a mix of global, local and neighbourhood spreading at the probabilities of 0.4, 0.2 and 0.4 (see circle on the plot), which causes the largest final outbreak with the highest spreading speed.

Discussion

In this study, we infer the epidemic spreading parameters of the Conficker worm from observed data collected during the first few hours of the epidemic. Simulations of worm spreading, based on these parameters, allow us to reach some important conclusions about the worm's use of hybrid spreading mechanisms.

Advantage of mixing hybrid spreading mechanism

Conficker's global probing is extremely ineffective. The infection rate of global probing is many orders of magnitude smaller than the recovery rate. This means, if Conficker used only the global probing, it would not have caused any significant infection on the Internet at all.

Local probing has a remarkably high infection rate, $\beta_l = 0.32$, which means when an infected node conducts only local spreading, a susceptible node in the same subnet has an 1/3 chance of being infected in a step (10-mins). However, local probing is confined within a subnet. If the worm used only the local probing, it would not have infected any other subnet apart from those initially containing infected nodes.

Neighbourhood probing is constrained to a neighbourhood of ten subnets. It has a high infection rate because computers in adjacent IP address blocks often belong to the same organisation and they use similar computer systems and therefore have similar vulnerabilities that can be exploited by the worm. Since different nodes' neighbourhoods can partially overlap with
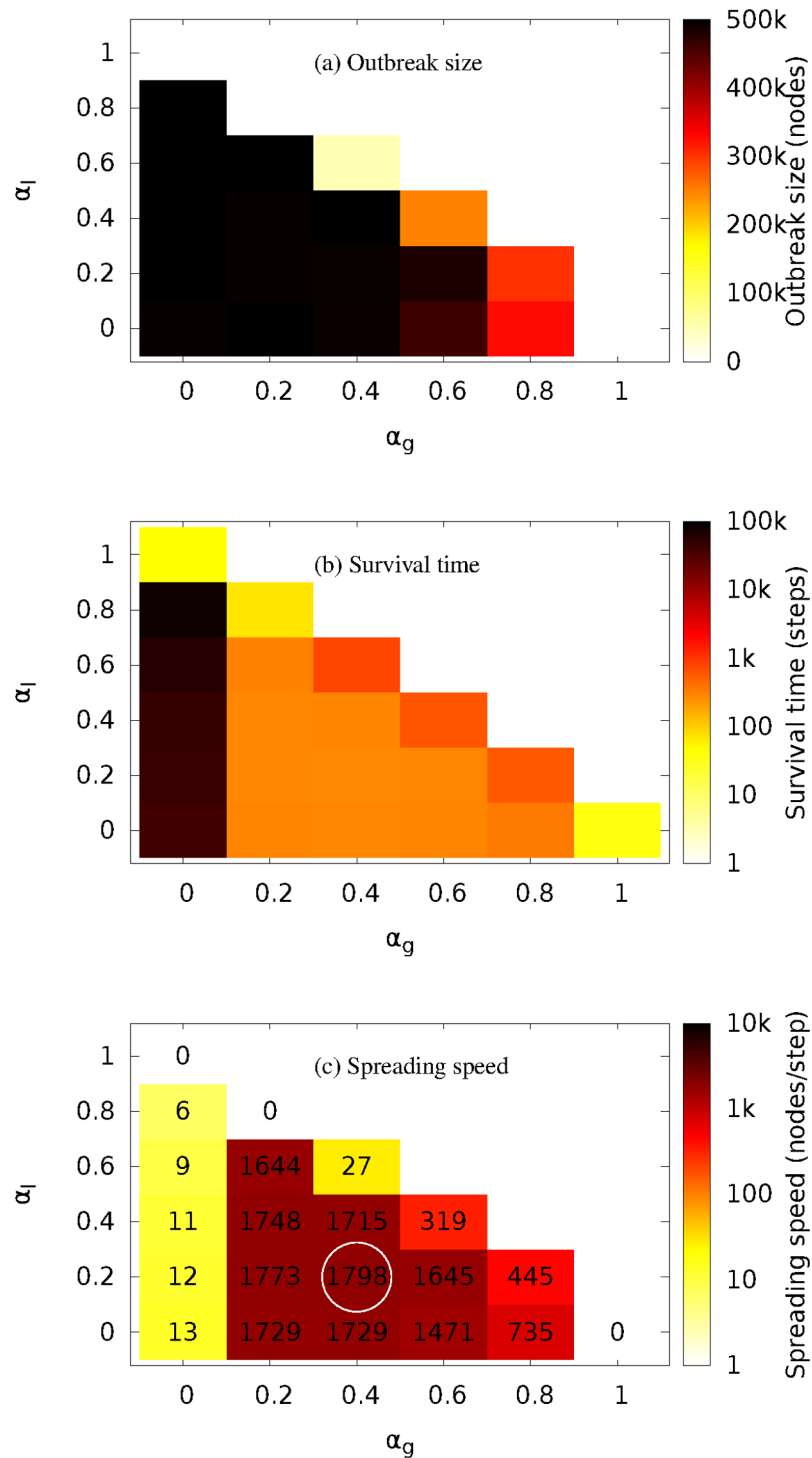
Fig 7. Simulation results when three of Conficker's spreading mechanisms are mixed at different probabilities. Spreading properties shown include the final outbreak size, the survival time and the spreading speed (see colour maps) as functions of the mixing probabilities of global spreading α_g (x axis) and local spreading α_l (y axis), where the mixing probability of neighbourhood spreading is $\alpha_n = 1 - \alpha_g - \alpha_l$.

each other, it is in theory possible for the worm to reach any node in the whole meta-population by using only the neighbourhood probing. Such process, however, would be extraordinarily slow as we have shown in Fig 6b.

In summary, if Conficker used only a single spreading mechanism, it would have vanished on the Internet without causing any significant impact.

Thus the enormous outbreak of the worm lies in its ability to do two things. Firstly it needs to devote great efforts to explore every corner of the Internet to find a new vulnerable computer. Every new victim will open a new colony full of similar vulnerable computers. Secondly it needs to make the most out of each new colony.

This is exactly what Conficker does. It allocates most of its time on global probing with a mixing probability of $\alpha = 89\%$. This in a degree compensates the ineffectiveness of global probing. Although the worm allocates small amounts of time on local and neighbouring probing, their high infection rates allow them to exploit all possible victims in the subnets with efficiency. And all newly infected nodes will join the collective effort to flood the Internet with more global random probes.

In short, the Conficker worm is an example of a critically hybrid epidemic. It can cause an enormous outbreak not because it has an advanced ability to exploit weaknesses of a computer, but because it has remarkable capability to discover all potentially vulnerable computers in the Internet, i.e. it is not the infectivity, but the hybrid spreading that makes Conficker one of the most infectious worms on record.

Implication of critically hybrid epidemics

The analysis of critically hybrid epidemics such as Conficker has important general implications. Firstly, it demonstrates that it is possible to design a high impact epidemic based on mechanisms, each of relatively low efficiency. Indeed our result in Fig 7 suggests that Conficker could have had a larger outbreak with higher speed if it had used a different set of mixing probabilities, which requires change of only a few lines of Conficker's program code. Hybrid mechanisms may therefore be ideal for rapid efficient penetration of a network, for example in the context of an advertising campaign or in order to promulgate important public health or security information. An interesting example might be the use of media campaigns (global spreading) where the reader or viewer is specifically requested to pass on a message via Twitter or Facebook to their "local" group contacts.

Conversely, malicious hybrid epidemics can be extremely difficult to defend against, and many existing defence strategies may not be effective. For example immunising a selected portion of a local population in order to isolate and hence protect the vulnerable nodes will not be effective, because the vulnerable nodes can still be found by the worm through random global spreading.

Another possible measure is to reduce the average time it takes for an infected node to recover, for example to speed up the release of anti-virus software updates or increase the frequency of security scanning on computers. Our theoretical predictions (using Eq 2) in Fig 8 show that the final outbreak size (in terms of total recovered nodes) does not change significantly when the recovery time is reduced from 156 minutes to 140 or 120 minutes. In practice, even achieving such reductions represents a remarkable technical challenge. It is clear from the discussion above that epidemics can spread with extremely low global infection rates (far below individual recovery rates), provided there is efficient local infection. The extremely efficient spreading achieved once a given subnet or set of subnets has been penetrated is therefore obviously a key determinant of the worm's outbreak [7]. Thus, defence strategies that focus on
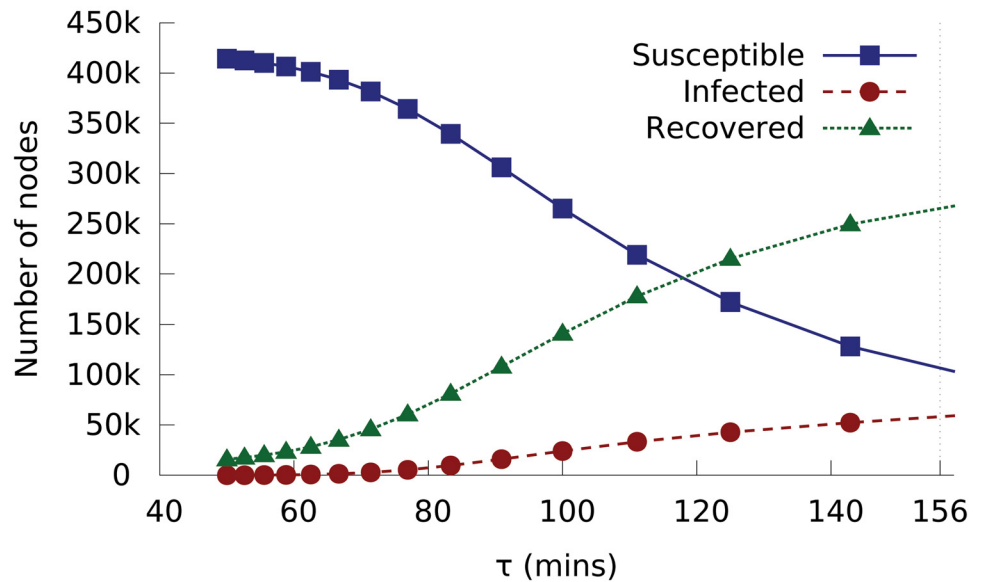
Fig 8. Predicted numbers of susceptible, infected and recovered nodes at 16:00 on the outbreak day as a function of the recovery time τ, which is the average time for an infected node to recover. Conficker's recovery time is 156 minutes.

doi:10.1371/journal.pone.0127478.g008

security co-operation between nodes with a local network neighbourhood (a "neighbourhood watch" strategy [7]) may be the key to future prevention of similar outbreaks.

Our Conficker model

The Conficker worm can be described as a discrete model or a continuous model. The two modelling approaches should give the same prediction results of the spreading dynamics of the worm. In this work we used a discrete approach to model the Conficker worm for three reasons. Firstly the model's parameters can be defined with clear physical meanings. Secondly we can directly calculate the parameters' values from the CAIDA measurement data. Lastly it is more convenient to run simulations with a discrete model. If a continuous model were used, the model parameters would be defined differently with less clear physical meanings, and their values would have to be obtained through iterative data fitting.

In our Conficker model, we set the local and global population as fully mixed, because this is how the Conficker worm perceives the structure of the Internet. We considered more complex network structures in a separate work [16] where we studied hybrid epidemics in general.

Conclusion

Our study uses data collected during the first day of the Conficker epidemic to parametrise a hybrid model to capture the worm's spreading behaviour. The study highlights the importance of mixing different modes of spreading in order to achieve large, rapid and sustained epidemics, and suggests that the trade-off between the different modes of spreading will be critical in determining the epidemic outcome.

Author Contributions

References

1. House T. Modelling epidemics on networks. Contemp Phys. 2012; 53(3):213. doi: 10.1080/00107514.2011.644443

2. Rock K, Brand S, Moir J, Keeling MJ. Dynamics of infectious diseases. Rep Prog Phys. 2014 Feb; 77 (2):026602. doi: 10.1088/0034-4885/77/2/026602 PMID: 24444713

3. Zheng Y, Huang B, Wang Y, Sun X, Jin Z, Zhang J, et al. Transmission dynamics and control for a brucellosis model in Hinggan League of Inner Mongolia, China. Math Biosci Eng. 2014 Jun; 11(5):1115–1137. doi: 10.3934/mbe.2014.11.1115 PMID: 25347802

4. Nie J, Sun GQ, Sun XD, Zhang J, Wang N, Wang YM, et al. Modeling the transmission dynamics of dairy cattle brucellosis in jilin province, china. J Biol Syst. 2014 May; 22(04):533–554. doi: 10.1142/S021833901450020X

5. Li MT, Sun GQ, Wu YF, Zhang J, Jin Z. Transmission dynamics of a multi-group brucellosis model with mixed cross infection in public farm. Appl Math Comput. 2014 Jun; 237:582–594. doi: 10.1016/j.amc.2014.03.094

6. Zou CC, Towsley D, Gong W. On the performance of Internet worm scanning strategies. Perform Eval. 2006 Jul; 63(7):700–723. doi: 10.1016/j.peva.2005.07.032

7. Shin S, Gu G, Reddy N, Lee CP. A Large-Scale Empirical Study of Conficker. IEEE Trans Inf Forensics Secur. 2012 Apr; 7(2):676–690. doi: 10.1109/TIFS.2011.2173486

8. Yu S, Gu G, Barnawi A, Guo S, Stojmenovic I. Malware Propagation in Large-Scale Networks. IEEE Trans Knowl Data Eng. 2015 Jan; 27(1):170–179. doi: 10.1109/TKDE.2014.2320725

9. Ren Z, Liu W, Zhou X, Fang J, Chen Q. Summary-Vector-Based Effective and Fast Immunization for Epidemic-Based Routing in Opportunistic Networks. IEEE Commun Lett. 2014 Jul; 18(7):1183–1186. doi: 10.1109/LCOMM.2014.2327038

10. Chen PY, Cheng SM, Chen KC. Optimal Control of Epidemic Information Dissemination Over Networks. IEEE T Cybern. 2014 Dec; 44(12):2316–2328. doi: 10.1109/TCYB.2014.2306781

11. Sahneh FD, Chowdhury FN, Brase G, Scoglio CM. Individual-based Information Dissemination in Multi-layer Epidemic Modeling. Math Model Nat Phenom. 2014 Jan; 9(02):136–152. doi: 10.1051/mmnp/20149209

12. Zhang J, Sun GQ, Sun XD, Hou Q, Li M, Huang B, et al. Prediction and Control of Brucellosis Transmission of Dairy Cattle in Zhejiang Province, China. PLoS ONE. 2014 Nov; 9(11):e108592. doi: 10.1371/journal.pone.0108592 PMID: 25386963

13. Zhang J, Jin Z, Sun GQ, Sun XD, Wang YM, Huang B. Determination of Original Infection Source of H7N9 Avian Influenza by Dynamical Model. Sci Rep. 2014 May; 4:4846. PMID: 24786135

14. Moore D, Paxson V, Savage S, Shannon C, Staniford S, Weaver N. Inside the Slammer worm. IEEE Secur Priv. 2003 Aug; 1(4):33–39. doi: 10.1109/MSECP.2003.1219056

15. Shannon C, Moore D. The spread of the Witty worm. IEEE Secur Priv. 2004 Aug; 2(4):46–50. doi: 10.1109/MSP.2004.59

16. Zhang C, Zhou S, Cox IJ, Chain BM. Optimizing Hybrid Spreading in Metapopulations; 2014. Preprint. Available: arXiv:1409.7291. Accessed 10 Feb 2015.

17. Chien E. Downadup: Attempts at Smart Network Scanning; 2010. Available: http://www.symantec.com/connect/blogs/downadup-attempts-smart-network-scanning. Accessed Dec 2014.

18. Center for Applied Internet Data Analysis. The CAIDA UCSD Network Telescope "Three Days Of Conficker"; 2008. Available: http://www.caida.org/data/passive/telescope-3days-conficker_dataset.xml. Accessed Dec 2014.

19. Center for Applied Internet Data Analysis. The CAIDA UCSD Network Telescope "Two Days in November 2008" Dataset; 2008. Available: http://www.caida.org/data/passive/telescope-2days-2008_dataset.xml. Accessed Dec 2014.

20. Newman M. Networks: An Introduction. Oxford University Press, USA; 2010.

21. Keeling M, Eames K. Networks and epidemic models. J R Soc Interface. 2005 Sep; 2(4):295–307. doi: 10.1098/rsif.2005.0051 PMID: 16849187

22. Anderson RM. Discussion: The Kermack-McKendrick epidemic threshold theorem. Bull Math Biol. 1991 Mar; 53(1–2):1–32. doi: 10.1007/BF02464422

23. Pastor-Satorras R, Vespignani A. Epidemic Spreading in Scale-Free Networks. Phys Rev Lett. 2001 Apr; 86(14):3200. doi: 10.1103/PhysRevLett.86.3200 PMID: 11290142

24. Wang L, Wang Z, Zhang Y, Li X. How human location-specific contact patterns impact spatial transmission between populations? Sci Rep. 2013 Mar; 3:1468. doi: 10.1038/srep01468 PMID: 23511929

25. Wang L, Zhang Y, Wang Z, Li X. The impact of human location-specific contact pattern on the sir epidemic transmission between populations. Int J Bifurcation Chaos. 2013 May; 23(05):1350095. doi: 10.1142/S0218127413500958

26. Balcan D, Colizza V, Gonçalves B, Hu H, Ramasco JJ, Vespignani A. Multiscale mobility networks and the spatial spreading of infectious diseases. Proc Natl Acad Sci USA. 2009 Dec;. doi: 10.1073/pnas.0906910106 PMID: 20018697

27. Wang P, González MC, Hidalgo CA, Barabási AL. Understanding the Spreading Patterns of Mobile Phone Viruses. Science. 2009 May; 324(5930):1071–1076. doi: 10.1126/science.1167053 PMID: 19342553

28. Balcan D, Vespignani A. Phase transitions in contagion processes mediated by recurrent mobility patterns. Nat Phys. 2011 Jul; 7(7):581–586. doi: 10.1038/nphys1944 PMID: 21799702

29. Meloni S, Perra N, Arenas A, Gómez S, Moreno Y, Vespignani A. Modeling human mobility responses to the large-scale spreading of infectious diseases. Sci Rep. 2011 Aug; 1. doi: 10.1038/srep00062 PMID: 22355581

30. Sun G, Jin Z, Liu QX, Li L. Pattern formation in a spatial S-I model with non-linear incidence rates. J Stat Mech. 2007 Nov; 2007(11):P11011. doi: 10.1088/1742-5468/2007/11/P11011

31. Sun GQ, Jin Z, Liu QX, Li L. Chaos induced by breakup of waves in a spatial epidemic model with non-linear incidence rate. J Stat Mech. 2008 Aug; 2008(08):P08011.

32. Sun GQ, Liu QX, Jin Z, Chakraborty A, Li BL. Influence of infection rate and migration on extinction of disease in spatial epidemics. J Theor Biol. 2010 May; 264(1):95–103. doi: 10.1016/j.jtbi.2010.01.006 PMID: 20085769

33. Sun GQ. Pattern formation of an epidemic model with diffusion. Nonlinear Dyn. 2012 Jan; 69(3):1097–1104. doi: 10.1007/s11071-012-0330-5

34. Moore D, Shannon C, Claffy KC. 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. IMW. ACM; 2002. pp. 273–284.

35. Ball F, Mollison D, Scalia-Tomba G. Epidemics with two levels of mixing. Ann Appl Probab. 1997 Feb; 7(1):46–89. doi: 10.1214/aoap/1034625252

36. Kiss IZ, Green DM, Kao RR. The effect of contact heterogeneity and multiple routes of transmission on final epidemic size. Math Biosci. 2006 Sep; 203(1):124. doi: 10.1016/j.mbs.2006.03.002 PMID: 16620875

37. Ball F, Neal P. Network epidemic models with two levels of mixing. Math Biosci. 2008 Mar; 212(1):69. doi: 10.1016/j.mbs.2008.01.001 PMID: 18280521

38. Estrada E, Kalala-Mutombo F, Valverde-Colmeiro A. Epidemic spreading in networks with nonrandom long-range interactions. Phys Rev E. 2011; 84(3):036110. doi: 10.1103/PhysRevE.84.036110

39. Vazquez A. Epidemic outbreaks on structured populations. J Theor Biol. 2007 Mar; 245(1):125–129. doi: 10.1016/j.jtbi.2006.09.018 PMID: 17097683

40. House T, Keeling MJ. Deterministic epidemic models with explicit household structure. Math Biosci. 2008 May; 213(1):29–39. doi: 10.1016/j.mbs.2008.01.011 PMID: 18374370

41. Ball F. An SIR epidemic model on a population with random network and household structure, and several types of individuals. Adv Appl Probab. 2012 Mar; 44(1):63–86. doi: 10.1239/aap/1331216645

42. Ma J, van den Driessche P, Willeboordse FH. Effective degree household network disease model. J Math Biol. 2013 Jan; 66(1–2):75–94. doi: 10.1007/s00285-011-0502-9 PMID: 22252505

43. Watts DJ, Muhamad R, Medina DC, Dodds PS. Multiscale, resurgent epidemics in a hierarchical metapopulation model. Proc Natl Acad Sci USA. 2005; 102(32):11157–11162. doi: 10.1073/pnas.0501226102 PMID: 16055564

44. Colizza V, Vespignani A. Invasion Threshold in Heterogeneous Metapopulation Networks. Phys Rev Lett. 2007 Oct; 99(14):148701. doi: 10.1103/PhysRevLett.99.148701 PMID: 17930732

45. Mata AS, Ferreira SC, Pastor-Satorras R. Effects of local population structure in a reaction-diffusion model of a contact process on metapopulation networks. Phys Rev E. 2013 Oct; 88(4):042820. doi: 10.1103/PhysRevE.88.042820

46. Min Y, Jin X, Ge Y, Chang J. The Role of Community Mixing Styles in Shaping Epidemic Behaviors in Weighted Networks. PLoS ONE. 2013 Feb; 8(2):e57100. doi: 10.1371/journal.pone.0057100 PMID: 23437321

47. Keeling MJ, Danon L, Vernon MC, House TA. Individual identity and movement networks for disease metapopulations. Proc Natl Acad Sci USA. 2010 May; 107(19):8866–8870. doi: 10.1073/pnas. 1000416107 PMID: 20421468

48. Apolloni A, Poletto C, Ramasco JJ, Jensen P, Colizza V. Metapopulation epidemic models with heterogeneous mixing and travel behaviour. Theor Biol Med Model. 2014 Jan; 11(1):3. doi: 10.1186/1742-4682-11-3 PMID: 24418011

49. Miller JC. Spread of infectious disease through clustered populations. J R Soc Interface. 2009 Dec; 6 (41):1121–1134. doi: 10.1098/rsif.2008.0524 PMID: 19324673

50. Tildesley MJ, House TA, Bruhn MC, Curry RJ, O'Neil M, Allpress JLE, et al. Impact of spatial clustering on disease transmission and optimal control. Proc Natl Acad Sci USA. 2010 Jan; 107(3):1041–1046. doi: 10.1073/pnas.0909047107 PMID: 19955428

51. Volz EM, Miller JC, Galvani A, Ancel Meyers L. Effects of Heterogeneous and Clustered Contact Patterns on Infectious Disease Dynamics. PLoS Comput Biol. 2011 Jun; 7(6):e1002042. doi: 10.1371/journal.pcbi.1002042 PMID: 21673864

52. Wang Y, Jin Z. Global analysis of multiple routes of disease transmission on heterogeneous networks. Physica A. 2013 Sep; 392(18):3869–3880. doi: 10.1016/j.physa.2013.03.042

53. ESET Virusradar. Win32/Conficker Charts; 2014. Available: http://www.virusradar.com/en/Win32_Conficker/chart/week. Accessed Dec 2014.

54. Irwin B. A network telescope perspective of the Conficker outbreak. In: Information Security for South Africa; 2012. pp. 1–8.

55. Chiba D, Tobe K, Mori T, Goto S. Analyzing Spatial Structure of IP Addresses for Detecting Malicious Websites. Journal of Information Processing. 2013; 21(3):539–550. doi: 10.2197/ipsjjip.21.539

56. Li R, Gan L, Jia Y. Propagation Model for Botnet Based on Conficker Monitoring. In: International Symposium on Information Science and Engineering; 2009. pp. 185–190.

57. Yao Y, Xiang WI, Guo H, Yu G, Gao FX. Diurnal Forced Models for Worm Propagation Based on Conficker Dataset. In: International Conference on Multimedia Information Networking and Security; 2011. pp. 431–435.

58. Aben E. Conficker/Conflicker/Downadup as seen from the UCSD Network Telescope; 2009. Available: http://www.caida.org/research/security/ms08-067/conficker.xml. Accessed Dec 2014.

59. Dagon D, Zou C, Lee W. Modeling botnet propagation using time zones. In: Annual Network & Distributed System Security Symposium; 2006.