Evidence for self-organized criticality in Internet attacks

Article · January 2007

CITATION
1

READS
80

2 authors, including:

Paul Oman
University of Idaho
111 PUBLICATIONS **1,984** CITATIONS

SEE PROFILE

Evidence for self-organized criticality in Internet attacks

James R. Conrad *

Department of Computer Science
University of Idaho
Moscow, Idaho 83844–1010

Paul W. Oman

Department of Computer Science
University of Idaho
Moscow, Idaho 83844–1010

Abstract

We submit quantitative evidence and a qualitative explanation for a Self-Organized Criticality (SOC) paradigm of the complex dynamics of Internet attacks. This discovery suggests the possibility of a new quantitative model explaining observations such as the the size and duration of malware attacks. Our research of 11 years of existing temporal data discovered the duration of historical attacks conforms to a SOC system's signature power-law curve. Likewise our analysis of the spatial data for the number of malware reports also found a power curve. We find this quantitative evidence consistent with an underlying SOC process and provide a comparison of the Internet's behavior with a well-known SOC model.

Key words: Internet; Security; Self-Organized Criticality; SOC; Power law; Malware; Modeling; Forest-Fire Model; FFM

1 Introduction

Models facilitate the analysis of security in real-world computer systems, addressing many questions about the route of attack, the probability of a suc-

* Tel.: +1.208.396.4842
Email addresses: conr2286@uidaho.edu (James R. Conrad), oman@cs.uidaho.edu (Paul W. Oman).

cessful attack, and the effectiveness of a defender's existing or proposed mitigations. Models of Internet-connected computers may be required to consider attacks from malware agents. But the values of many modeling parameters about these agents including the duration of an agent's attack, or the number of Internet hosts compromised by an agent, often appear as subjective expert estimates.

We seek a model explaining behaviors of the complex dynamics of the Internet under attack. Unlike those computer security models addressing questions about access rights, the transfer of authority, confidentiality or integrity (4), we seek a model of system-level behaviors of the Internet: Why do some malware agents "fizzle out" while others bring on widespread infection and service disruption (14)? Why are some malware agents short-lived while others linger for many months? Is there a mechanism to forecast an impending widespread malware attack or to quantify the current risk level? Is there a mechanism to forecast the duration or size of future attacks?

Related work in scale-free networks includes Crucitti, Latora and Marchiori's investigation of cascading failures in scale-free networks including the Internet and an electrical power grid (10), and Chassin and Posse's evaluation of a power grid as a scale-free network (7). Barabsi and Albert offer a qualitative explanation for growth and preferential attachment leading to scale-free organization in the Internet (3). Staniford also notes scale-free organization in the Internet (16). Albert and Barabsi review several scale-free networks including the internet (1). Leland et al. report on the self-similar nature, a form of scale-free behavior, of Ethernet LAN traffic in which every time scale, both short and long, is filled with bursty sub-periods punctuated by less bursty subperiods (13). Siganos et al. relate power-laws to the topology of autonomous subnets (15).

We build upon these well-documented observations of scale-free Internet topologies to consider the possibility of Self-Organized Criticality (SOC) behavior in the Internet's response to malware attacks. An SOC system organizes itself to exhibit scale-free behaviors in both time and space. We begin by reviewing the common characteristics of SOC processes and a well-known SOC model. Then we report quantitative evidence suggesting the empirical observations of 11 years of malware activity are indeed consistent with an underlying SOC process. We continue with a qualitative explanation for why the internet may behave as an SOC process, and we conclude with guidance for the future study of this paradigm.

2 A Quick Review of Self-Organized Criticality

A *self-organized* system autonomously develops a structure or pattern independent of any controlling input (*tuning*) from an external entity (12). All members of a *critical* system influence one another — the effect of perturbing any one member of the system propagates throughout the overall system. *Self-Organized Criticality* refers to the ability of certain systems to organize themselves near a critical state independent of their initial conditions or external guidance. Unlike a critical point at phase transitions of a thermodynamic equilibrium system, an SOC system becomes critical without external *tuning* of a parameter such as its temperature (2). An SOC system dynamically evolves near a critical point.

Research in SOC modeling began with Bak, Tang and Wiesenfeld began investigating mathematical models to explain complex temporal and spatial behaviors observed in certain natural systems (2). SOC models have since been investigated for describing the behavior of landslides, earthquakes, forest fires, solar flares, epidemics, and other natural systems (12) (8) (17). SOC model forecasts agree well with empirical data for many natural and complex man-made systems.

Research into the application of Self-Organized Criticality (SOC) behaviors in man-made complex systems includes Carreras et al. discovery of SOC behaviors in the temporal and spacial characteristics of blackouts in the North American electrical power grid (5) (6).

2.1 Characteristics of SOC Systems

Despite much progress in the physical science community, self-organized criticality remains an emerging area of study. Our research follows the guidance of Jensen and sets the necessary criteria for SOC behavior to require scale-free behavior in both time *and* space (12). Thus we expect to observe power-law distributions of both temporal and spatial behaviors.

Scale-free power function distributions appear frequently in SOC discussions. Scale-free behavior, also known as scale-invariance, refers to a system property that remains the same even when the size of its system is scaled up or down. Fractal coastlines are a classic example: the coastline observed from a distance appears similar to that observed more closely. Interesting properties of an SOC system may include the duration of the system's response to a disturbance, or the number of system elements participating in a disturbance. The duration of fires in the SOC Forest Fire Model, for example, is (within the constraints of the finite model system) independent of the size of the model and their

probability described by a power function. An important characteristic of a scale-free graph is for some nodes to be much more highly connected than are other nodes, and the number of connections is (again, within the constraints of the finite model system) independent of the size of the graph (3) (15).

Avalanches are disturbances within a sandpile system composed of elements of sand. Bak et al. studied the distribution of avalanche lifetimes in a sandpile model and found the noncumulative probability distribution [1], $p(t)$, of the temporal events to be a power function:

$$p(t) \sim t^{-\alpha} \qquad \text{Where } \alpha \approx 1.00 \tag{1}$$

In this paper, the symbol "$\sim$" should be read, "is proportional to." A log-log plot of $p(t)$ in Equation 1 for an SOC system appears as a straight line over long periods of the interval, t. Boundary effects in many SOC models cause $p(t)$ to deviate from a power function above some large interval, t_m. For example, in the SOC Forest Fire Model (FFM) described below, the duration of any model fire will require no more time than is necessary to burn the largest possible model forest. Likewise, the discrete nature of many SOC models also causes $p(t)$ to deviate from a power function as $t \rightarrow t_Q$ for some small quantum interval, t_Q.

In the spatial domain, the non-cumulative probability distribution of an event in an SOC system is a power function of the event size, s:

$$p(s) \sim s^{-\tau} \qquad \text{Where } \tau \approx 1.00 \tag{2}$$

In the example of real-world forest fires located in the Western United States, the noncumulative normalized frequency distribution of fires as a function of their size is a power function having $\tau \approx 1.3$ (17).

Jensen, who documented physical and biological SOC systems, expects self-organized critical behavior in *slowly driven, interaction-dominated threshold* systems (12). He notes the mutual interaction of their many degrees of freedom dominates their behavior. Energy is slowly driven into these systems and is eventually suddenly released (in, for example, an avalanche). In an SOC system, interactions (e.g., friction) between cells in a model must dominate external forces (e.g., gravity).

Power laws in the form of Equations 1 and 2 comprise the principle temporal and spatial evidence for SOC behavior in many systems (12) and will form the foundation of our quantitative argument for SOC behavior in the Internet's response to attacks.

[1] Unlike the physical science norm, this paper adopts the mathematical convention with a lower-case name representing the noncumulative probability density function.

2.2 The SOC Forest-Fire Model

The Forest Fire Model (FFM) is a well-documented SOC model with known applications well outside of its original domain (8) (9). While extremely simplistic in composition (neglecting most real-world parameters such as terrain, insect damage, fuel species, weather, etc.), the FFM's forecasts are nevertheless in reasonable agreement with evidence for SOC behavior in real-world fires (17).

The FFM is a randomly driven cellular automaton on a hyper-cubic lattice with L^d sites (cells) where L is the number of cells along an axis (dimension) and d is the number of dimensions (9). While the model has been studied from one to in excess of six dimensions, we focus on the 2-dimensional case here. Each cell in the lattice exists in one of three states:

(1) *empty*
(2) *tree*
(3) *burning*

The automaton updates the state of every cell in the lattice during the time-interval of each generation (time increment) by applying a set of state transition rules:

(1) *empty* $\rightarrow$ *tree* with probability p
(2) *tree* $\rightarrow$ *burning* with probability f if no neighbor is burning
(3) *tree* $\rightarrow$ *burning* with probability $1 - g$ if at least one neighbor is burning
(4) *burning* $\rightarrow$ *empty*

The first rule represents the slow, random drive of the system by the growth of a new model tree on an empty site as illustrated by the *growth* transition of Figure 1. Over a period of time, this first rule grows clusters of trees in the model forest. The second rule describes the possibility of random ignition arising from the stimulus of the system by lightning as illustrated by the *ignition* transition of Figure 1. This second rule disturbs a system of tree clusters. The third rule governs the random, rapid spread of a fire through a cluster of adjacent sites, the dominant interaction between individual model trees as illustrated by the *spread* transition of Figure 1. The parameter g represents the immunity of a tree to the propagation of the spreading flames from a neighboring site. The fourth rule describes the rapid combustion of a burning site as illustrated by the *combustion* transition of Figure 1.

If the model of Figure 1 is initialized with an empty lattice, clusters of vulnerable trees grow slowly on empty sites. Large clusters of trees appear when $f \ll p$ because fires are relatively rare, eventually creating conditions extremely vulnerable to fire. Lightning has no effect on empty clusters, and may ignite a

small cluster with only limited consequence. But eventually, lightning ignites a large cluster and burns it down as rapidly as the trees' immunity permits. In the most vulnerable case, a system filled with trees in every cell of the lattice, any lightning strike will burn the entire model forest with probability, f. Energy slowly accumulated in the model forest by tree growth rapidly relaxes during a major fire.

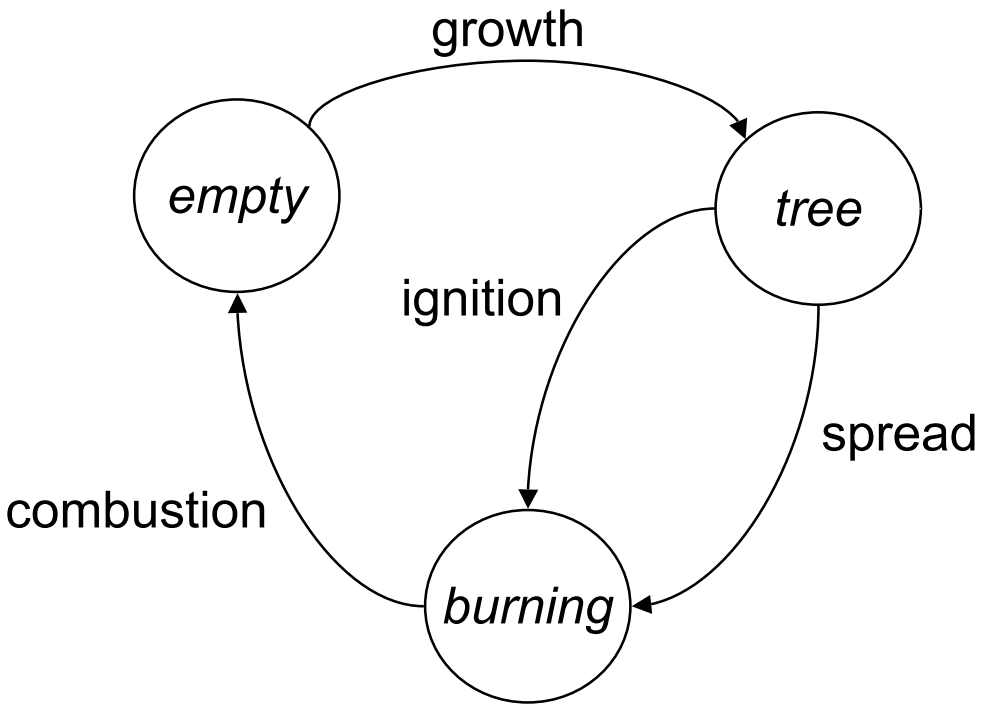

Fig. 1. State Transition Diagram for a Single Site in the Forest Fire Model

Note the similarity between the states of the FFM and those of a susceptible/infective/removed (SIR) epidemiology model: an FFM *empty* site compares with the SIR removed (also known as "recovered") state, a *tree* site compares with susceptible, and a *burning* site compares with infective. The FFM has been used to model the spread of measles in real-world island populations (12).

Several power law relationships appear in the FFM results. The non-cumulative normalized frequency distribution of fires per time step as a function of their area provides spatial evidence of SOC behavior in the FFM:

$$\frac{N_F}{N_S} \sim A_F^{-\tau} \tag{3}$$

where N_F/N_S is the normalized frequency distribution of fires having an area A_F over N_S generations. Turcotte reports results from a simulation of $N_S = 1.638 \times 10^9$ generations for a 128×128 grid in which the exponent, τ, was found

to vary from $\tau = 1.02$ for $1/f = 125$ to $\tau = 1.16$ for $1/f = 2000$ (17). The power law relationship breaks down for large fires whose size approaches that of the finite grid. Similarly, because the area, A_F, of a minimal fire must always be ≥ 1 model tree, the power law relationship does not exist for $A_F < 1$.

Likewise, the duration of fires provides temporal evidence of SOC behavior in the FFM (8):

$$N(T) \sim T^{-\alpha} \tag{4}$$

where $\alpha = 1.27$ and $N(T)$ refers to the number of model fires having duration T.

Strictly speaking, the FFM requires at least limited parameter tuning to arrange $f \ll p$ and is thus arguably not wholly self-organizing. Further study recognized that a double separation of time scales is necessary for SOC behavior in the FFM:

$$(f/p)^{-v'} \ll p^{-1} \ll f^{-1} \tag{5}$$

where $(f/p)^{-v'} \sim T(s_{max})$, the time required to burn down a large cluster (9). Thus, the FFM exhibits SOC behavior when the time required to burn down a large cluster is much less than the time required to grow a new tree which is much less than the time between lightning strikes. If trees grew too fast, fires would have a continuous source of fuel as replacement trees would appear even while a model fire still burned. And if lightning was too frequent, only small forest clusters would burn because large clusters would not have time to emerge.

The partially tuned characteristic of the FFM suggests the system could also be tuned *away* from criticality. Real forests, for example, may be tuned away from criticality with prescribed burns. The possibility of tuning may pose interesting questions for the Internet as described in Section 6.

3 Empirical Data for Internet Attacks

Our investigation into the SOC behaviors of the Internet extracted malware activity data from the Virus Bulletin (VB) archives (18). The VB data contains monthly reports of malware activity from 1995 to the present time. Unfortunately, we can't assume the data was uniformly sampled from the entire population of malware activity during the entire period of study, so we work around this limitation with two assumptions:

(1) We assume all malware agents are uniformly sampled during any single reporting period. Thus, for example, 2000 reports (samples) of malware

agent X activity during a specific month reflect twice as much activity in the overall population as 1000 reports of agent Y activity during that same month (period).

(2) We assume the sampling function changes relatively slowly when compared with the lifetime of a single malware agent. Thus, for example, 2000 reports (samples) of malware agent X activity during the first month of a specific year reflect twice as much activity in the overall population as 1000 reports of agent X during the second month of that same year.

4 Evidence for SOC Behavior in the Internet's Response to Attacks

At a critical point, fluctuations in a real-world system occur at "all" time and length scales. For example, both the duration and the size of real-world forest fires as well as FFM fires exhibit scale-free, power-law behavior. Consequently, Jensen argues the need to investigate both temporal and spatial power-law behaviors before concluding a system is SOC (12). Our analysis complies with Jensen's approach; we studied both temporal and spatial observations of SOC behavior during the Internet's growth from a few to approximately 325 million hosts during an 11 year period from January 1991 through January 2006 (11).

4.1 Temporal Evidence for SOC Behavior During Internet Attacks

Figure 2 illustrates the non-cumulative frequency–duration distribution of the reported VB malware agents. The defined duration (lifetime) of an agent begins when first observed and ends when last observed during the eleven years of analyzed VB data; no special provision was made for malware agents whose activity actually began prior to or ended following the analysis period. This definition is compatible, for example, with that used for the duration (lifetime) of model fires in the SOC FFM.

Our best-fit line for the VB temporal data is:

$$N(T) \sim T^{-\alpha} \qquad \alpha = 1.0917 \tag{6}$$

where $N(T)$ is the number of malware events having duration, T. This result compares with simulation results from the two dimensional FFM for which the exponent of T is 1.27 (8)

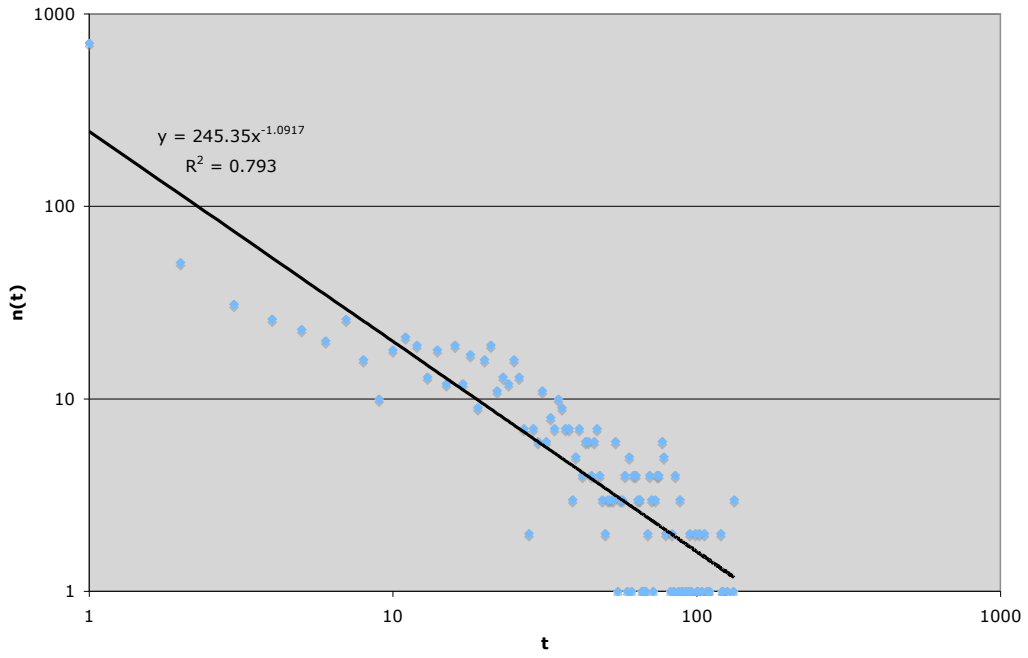

Fig. 2. Frequency Chart of Malware Agents' Durations. The vertical axis represents the number of reported agents whose duration in months is given on the horizontal axis. The solid line represents our best fit.

4.2 Spacial Evidence for SOC Behavior During Internet Attacks

Figure 3 illustrates the normalized non-cumulative frequency–size distribution of reported malware agent activity. We define the *size* of a malware agent's activity as the tally of all sampled reports of its activity in the VB data during the reporting period. This definition is comparable, for example, to that used for the size of model fires (expressed as the count of burned trees in a cluster) in the FFM.

Our best-fit line for the VB spacial data is:

$$\frac{N_F}{N_S} \sim A_F^{-\tau} \qquad \tau = 1.0613 \tag{7}$$

where N_F/N_S is the normalized non-cumulative frequency distribution of malware attacks (ignoring reports of single incidents) resulting in A_F reports to the VB archives over a period of N_S months. This value for τ is in good agreement with the SOC FFM where results for τ range from 1.02 through 1.16, and is comparable with real-world North American forest fires for which $\tau \approx 1.3$ (17).

Power-law behaviors tend to break down near real-world boundary conditions. In the case of the Internet, no successful attack will infect less than a single

host. The power-law behavior of Figure 3 breaks down as the normalized frequency of reports decreases to consider attacks whose size was observed only once, the minimum possible given the discrete nature of the reporting process. These are omitted from the analysis and chart of Figure 3. If included, their extreme uncertainty would cause them to appear as a wide row of points across the bottom of the chart at the minimum frequency.

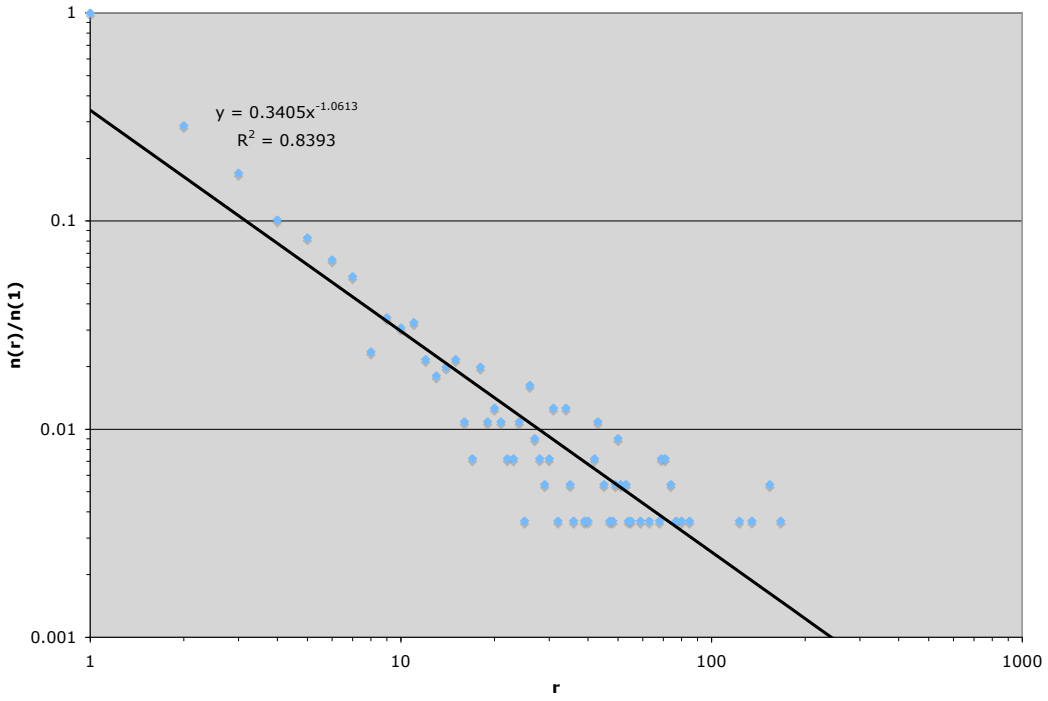

Fig. 3. Normalized Frequency of Malware Agents' Reports. The vertical axis represents the normalized frequency of reports whose size is given on the horizontal axis. The solid line represents our best fit.

The VB data yields good power-law fits for both the temporal and spatial evidence during the years 1995 to 2005 in which the Internet grew from just a few hosts to approximately 325 million hosts (11). This period also experienced changes in network connectivity, computer operating systems, increasingly sophisticated malware, and the rise of defensive technologies. VB's malware sampling function possibly, if not likely, also changed during this period. Yet simple power laws offer a useful forecast of important malware behaviors during this seemingly chaotic period of the Internet's development, an outcome consistent with other observations of real-world SOC behaviors (17).

5 Toward a Qualitative Explanation for SOC Behavior

This section illustrates how the FFM provides a plausible explanation for SOC behaviors in the Internet's response to malware attacks. Section 4 previously offered quantitative evidence of the Internet's behavior and a comparison with the results obtained from the FFM. Here we begin by proposing a FFM-like state machine for an Internet host under attack and examine its operation in detail, noting the many similarities between the Internet model and the FFM state machine.

Figure 4 illustrates a simple FFM-like model for an Internet host under attack by a malware agent. Compare this with the FFM state machine previously illustrated in Figure 1. In Figure 4, the Internet host's *clean* state is analogous to the *empty* state of the FFM, the *vulnerable* state is analogous to the *tree* state, and the *exploited* state is analogous to the *burning* state.

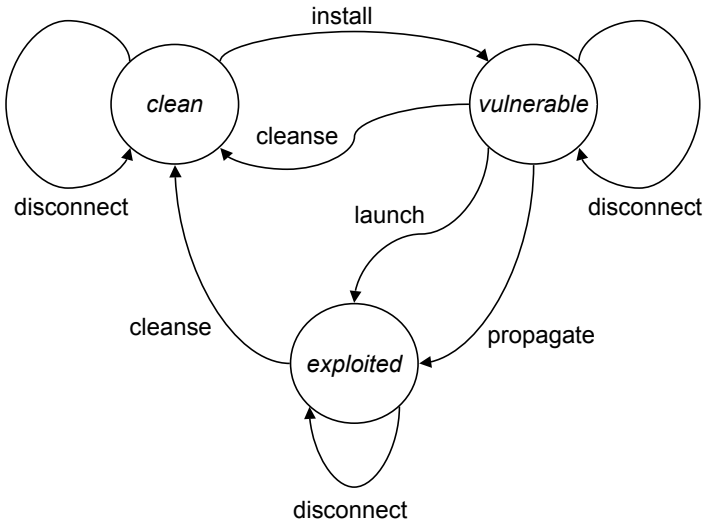

Fig. 4. State Transition Diagram for an Internet Host Under Attack

5.1 Install Transition

Owners of *clean* Internet computers create *vulnerable* hosts by installing defective software packages. The steady-state creation of *vulnerable* hosts behaves

similar to the FFM's growth of vulnerable green model trees, slowly assembling clusters of connected, *vulnerable* hosts.

Like the FFM, the installation of a single vulnerability does not by itself destroy the system; it merely charges the system with vulnerabilities as illustrated in Figure 4. These vulnerabilities establish the potential for exploit, similar to the way the FFM's trees establish the potential for fires.

5.2 Launch Transition

Eventually, someone may develop and launch a malware agent exploiting a vulnerability in a networked host, similar to the role of a lightening strike in the FFM. The launch provides a stimulus that moves its target host to the *exploited* state where the agent is free to propagate to connected hosts in the network. In the terms of an SIR model, the *exploited* state is *infectious*.

Unlike the FFM, a network host may contain multiple vulnerabilities and the malware agent may exploit any one or more of these. This difference may be largely inconsequential in a large system because the FFM model can be applied individually to each vulnerability in a host, assigning a state machine to each vulnerability–host pair. While the number of number of communicating state machines increases, the behavior of each cell remains the same.

5.3 Propagate Transition

Like the spread of fire through a cluster of vulnerable green model trees, the malware agent spreads quickly through the cluster of interconnected hosts sharing a vulnerability susceptible to the agent. The agent is propagated within a cluster of connected hosts from *exploited* hosts to adjacent *vulnerable* hosts. If the cluster is small, the stimulus destroys a relatively small number of hosts; large clusters of vulnerable Internet hosts lead to wide-spread destruction.

Unlike the fixed d-dimensional hypercubic lattice structure of Drossel and Schwabl's original FFM, Internet hosts have variable connectivity; some hosts are far more connected than others (3). The FFM lattice has been explored with simulations in one to six dimensions (8) and, more importantly, has been used to model the spread of measles within a human population with comparable variable connectivity (3) (12). Therefore, although variable connectivity does not appear to prohibit application of the FFM, the quantitative role of variable connectivity on the internet's behavior under attack is not understood.

5.4 Cleanse Transition

As the malware agent propagates, defenders may respond with attempts to cleanse computers from the propagating agent. A cleansing action often begins by removing the exploit from an infected host and subsequently patching (removing) the vulnerability; this approach may be preferred because it preserves the user's data. Alternatively, cleansing can be accomplished by replacing the infected computer with a *clean* machine whose configuration does not include the vulnerability. Cleansing leaves the host invulnerable to the propagating agent, moving it to the *clean* state.

Cleansing is somewhat comparable to fire fighting methods in real-world forests employing retardants to douse burning trees in an attempt to render them inflammable.

But unlike the FFM whose *burning* trees are quickly consumed by the flames and transition to the *empty* state, *exploited* computers remain infective until they are either cleansed or disconnected. Because adjacent (connected) hosts are very quickly infected by a propagating agent, this dwell is of minimal consequence unless the network is growing rapidly with respect to the rate of propagation, confirming one of the prerequisites for SOC behavior in the FFM. Recall that Equation 5 limits SOC behavior in the FFM to when the time required to burn down a large cluster is much less than the time required to grow a new tree.

5.5 Disconnect Transition

When a software patch is unavailable or cannot be applied expediently, computer administrators may choose to disconnect their host[s] from the network. Disconnection leaves a host unchanged in its current state, affecting only its connectivity (not its state) in the network. Disconnection does affect a host's connectivity with its neighbors.

Disconnection has no exact analogy in the fixed lattice structure of Drossel and Schwabl's original FFM. The closest concept may be complete immunity, $g = 1$, in the FFM; but disconnection blocks propagation but only at one local site rather than throughout the system as is the case for complete immunity.

Once ignited, the only possible outcome for a tree in the FFM is an empty site as the tree is wholly consumed by the flames. Once exploited, a host computer will either be cleansed or disconnected. Cleansing always transitions a host to the *clean* state, analogous to the FFM's *empty* state.

Note that disconnection bears some similarity with fire fighting methods that establish fire-lines in real-world forests in attempts to isolate structures of vulnerable trees from an advancing front.

5.6 Other Comparisons Between the Internet and FFM Behaviors

The necessary condition for SOC behavior in the FFM requires the separation of time-scales described by Equation 5: the time required to burn down a large cluster must be much less than the time required to grow a new tree which much be much less than the time between lightening strikes. The internet analogy requires the time for a malware agent to propagate through a large cluster of vulnerable hosts, T_p, to be much less than the time required to add a new vulnerability, T_v, which must be much less than the time between launches of exploits, T_l. We restate the FFM's requirement for SOC behavior, Equation 5, for the Internet as:

$$T_p \ll T_v \ll T_l \tag{8}$$

It is convenient to think of a cluster as a neighborhood of well-connected network hosts having restricted connectivity to the overall internet, an intranet. Staniford, Paxson and Weaver document the onset of the Nimbda agent at the Lawrence Berkeley National Laboratory from essentially no activity to its sustained rate of nearly 100 propagation attempts/second during a 30 minute period on September 18, 2001 (16). They postulated the construction of a *Warhol* worm that would attack most vulnerable targets in less than 15 minutes, a result realized when the Slammer agent achieved its peak scanning rate in only 3 minutes on January 25, 2003, infected 90% of its vulnerable internet hosts in only 10 minutes, and disrupted financial, transportation and government infrastructures (14). Assuming Slammer's peak scanning rate was achieved only after propagating through at least one large cluster of vulnerable hosts:

$$T_p \leq 3 \qquad \text{Malware cluster propagation time (minutes)} \tag{9}$$

The T_v parameter describes the mean time required to add a vulnerability at a site, proportional to p^{-1} where, analogous to the FFM, p is the probability that a *clean* site will become a *vulnerable* site during one interval of the model simulation. While data is not available for real-world hosts, we assume the average time between installations of vulnerable software on a host, T_v, is at least one order of magnitude more than the few minutes required for an agent to propagate through a large cluster.

$$T_v > 30 \qquad \text{Time (minutes) between new vulnerability installations} \tag{10}$$

During the period of January, 1995 through January, 2006 inclusive, the VB database records the reported discovery of 1535 new malware agents, providing a crude estimate, $T_l \approx 3790$ minutes, for the mean time between the reported launch of new agents. Because many new agents are merely minor variations that exploit the same vulnerability used by a predecessor, the actual time between reported exploits of new vulnerabilities is likely even longer. Thus, the inequalities of Equation 8 are satisfied in:

$$(T_p = 3) << (T_v = 30) << (T_l = 3790) \tag{11}$$

by the Internet and provide the necessary double separation of time-scales required for SOC behavior in the FFM.

One notable distinction between the Internet and FFM is the Internet is known to be organized as a scale-free system in which some hosts are very highly connected while others have only limited connectivity (3) (10) (13). The FFM, on the other hand, is a fixed lattice in N dimensions. How, then, can we expect to compare the results of the empirical Internet data with the FFM? Simulations of the FFM disclose that while both the temporal and spacial exponents are indeed somewhat dependent upon d for $d < 6$, the exponents are relatively independent of d for $d >= 6$ (12) (8). This observation suggests the exponents are at least somewhat independent of the number of connections to a site given some minimum connectivity.

6 Conclusions

The Virus Bulletin archives of malware activity during the years 1995 through 2005 include both temporal and spatial evidence for SOC behavior of the Internet under attack. The spatial evidence (Equation 7 and Figure 3) describes the number of reports of malware activity and conforms closely with the FFM forecast (Equation 3). The Internet's temporal evidence describes the reported duration of of malware activity (Equation 6 and Figure 2) and is also comparable to the FFM (Equation 4).

Although the Internet has displayed a complex response to malware attacks, simple power-law distributions model the statistical properties of the size and duration of attacks reported in the analyzed data. More traditional statistical models, such as a normal (Gaussian) distribution, can underestimate the probability of very large or very long malware events on the Internet because the probability of extreme events in the normal distribution is much less than in the power-law distribution.

In addition to the quantitative evidence, qualitative similarities exist between the FFM and the behavior of the Internet before, during and after an exploit

is released. Host computers serve a role similar to trees in the FFM and are vulnerable to various malware attacks. A malware exploit launches an attack on a vulnerable subnet of connected host computers, similar to the way lightning ignites a fire in a cluster of trees. An exploit spreads rapidly through the connected vulnerable hosts, comparable to the progress of a fire through a forest cluster. Infected hosts are eventually disabled, repaired or replaced somewhat similar to the way trees burn to the ground and remove a site from a vulnerable forest cluster.

Self-Organized Criticality offers a possible model for how the Internet can work "well" for very long periods of time and then suddenly suffer a widespread disruption in response to a particular malware agent's attack while dozens of other malware agents may simple "fizzle out." This behavior is typical of SOC systems including the FFM, sandpiles, avalanches, earthquakes, and the electrical power grid.

A definitive conclusion of SOC behavior within the Internet requires further research. A second data source, independent of the VB archives, is needed to substantiate our findings. If substantiated, our research suggests additional questions including:

- As discussed in section 2.2, the FFM can be tuned away from criticality: the time required to burn a forest cluster must be much less than the time required to grow a new vulnerable tree which must in turn be much less than the time between two lightning strikes. Preemptive burns tune real-world forests away from criticality by decreasing the time between two ignition events (lightning strikes) within a forest cluster. Can the Internet be tuned away from criticality?
- Boundary firewalls may help to isolate clusters of vulnerable hosts. Virus Throttles are known to slow the spread of malware; can their behavior be modeled using the FFM's g parameter (immunity) (19). Can we extend the FFM to simulate the role of firewalls? Does the FFM provide insights about the optimal location or tuning for Virus Throttles?
- FFM fires wholly consume (burn to the ground) a tree in a single interval of the model. Internet hosts, somewhat like real-world trees, will remain infective ("burn") for varying lengths of time. Can a modified FFM provide insights about the value of quickly removing infective hosts from the network?
- Can the size, duration or another outcome of relatively common (minor) attacks somehow be used to forecast the probability of an impending widespread disruption of Internet service (perhaps akin to earthquake forecasts)?
- Does Self-Organized Criticality offer new insight on how to manage a vulnerable population?
- Is there any practice in the Internet serving a beneficial role similar to prescribed burns in real-world forest fires?

- Does SOC behavior in the Internet's response to attack arise from a scale-free topology, or do both arise from a more primitive condition (3).

7 Acknowledgments

The authors acknowledge the foresight and contribution of the Virus Bulletin whose malware archives supported this analysis (18). We also gratefully acknowledge the contributions of the referees.

References

[1] R. Albert, A.-L. Barabási, Statistical mechanics of complex networks, Rev. Mod. Phys. 74 (1) (2002) 47–97.

[2] P. Bak, C. Tang, K. Wiesenfeld, Self-organized criticality: An explanation of the 1/f noise, Phys. Rev. Lett. 59 (4) (1987) 381–384.

[3] A.-L. Barabási, R. Albert, Emergence of scaling in random networks, Science 286 (1999) 509 – 512.

[4] M. Bishop, Computer Security: Art and Science, Addison-Wesley, Boston, MA, 2003.

[5] B. A. Carreras, D. E. Newman, I. Dobson, A. B. Poole, Evidence for self-organized criticality in electric power system blackouts, in: Proceedings of the 34th Annual Hawaii International Conference on Systems Sciences, 2001.

[6] B. A. Carreras, D. E. Newman, I. Dobson, A. B. Poole, Evidence for self-organized criticality in a time series of electric power system blackouts, IEEE Circuits and Systems 51 (9) (2004) 1733–1740.

[7] D. P. Chassin, C. Posse, Evaluating north american electric grid reliability using the barabási-albert network model, Physica A: Statistical Mechanics and its Applications 355 (2–4) (2005) 667–677.

[8] S. Clar, B. Drossel, F. Schwabl, Scaling laws and simulation results for the self-organized critical forest-fire model, Physical Review E 50 (2) (1994) 1009–1019.

[9] S. Clar, B. Drossel, F. Schwabl, Forest fires and other examples of self-organized criticality, J. of Phys.: Cond. Mat. 8 (1996) 6803.

[10] P. Crucitti, V. Latora, M. Marchiori, Model for cascading failures in complex networks, Physical Review E 69 (045104).

[11] ISC, ISC domain survey: Number of internet hosts, Tech. rep., Internet Systems Consortium (September 2007).
URL http://www.isc.org/index.pl?/ops/ds/host-count-history.php

[12] H. J. Jensen, Self-Organized Criticality: Emergent Complex Behavior in

Physical and Biological Systems, Cambridge University Press, The Edinburgh Building, Cambridge CB2 2RU, U.K., 1998.

[13] W. E. Leland, M. S. Taqqu, W. Willinger, D. V. Wilson, On the self-similar nature of ethernet traffic (extended version) ; networking, ieee/acm transactions on, Networking, IEEE/ACM Transactions on 2 (1) (1994) 1–15.

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

[15] G. Siganos, M. Faloutsos, P. Faloutsos, C. Faloutsos, Power laws and the as-level internet topology, Networking, IEEE/ACM Transactions on 11 (4) (2003) 514–524.

[16] S. Staniford, V. Paxson, N. Weaver, How to 0wn the internet in your spare time, in: Proceedings of the 11th USENIX Security Symposium, The USENIX Association, San Francisco, CA, USA, 2002.

[17] D. L. Turcotte, Self-organized criticality, Reports on Progress in Physics 62 (10) (1999) 1377–1429.

[18] Unknown, Malware prevalence, Tech. rep., Virus Bulletin (January 2006). URL http://www.virusbtn.com

[19] M. M. Williamson, Design, implementation and test of an email virus throttle, 2003.