ASIM: An Agent-Based Integrated Model for Complex Networks

Steven Hofmeyr, PI
Lawrence Berkeley National Laboratory
shofmeyr@lbl.gov

Stephanie Forrest, Co-PI
University of New Mexico

Complex networks comprised of large-scale distributed heterogeneous systems are everywhere, for example, in brains, networks of social and economic interactions, and the Internet. Often these systems have surprising common properties, such as power-law degree distributions, which occur in networks as different as metabolic networks and co-authorship patterns in scientific publications. Many mathematical and computational models have been developed over the past decade that attempt to explain emergent properties at a macro-level. However, understanding more detailed structure and behavior of complex networks has required the incorporation of more domain-specific features, for example, incorporating spatial distributions and business relationships into models of the Internet. In this proposal, we focus on the Internet, and Internet-like systems.

The Internet is one of the largest and most complex human artifacts ever created, and operates on many different scales, from the slow expansion of new autonomous systems, to the speed-of-light propagation of data. The Internet involves a multitude of heterogeneous systems and organizations around the world, including many within the purview of the Department of Energy (DOE). Although the DOE directly administers networks such as ESNet, much of the information flow in "open science" depends on the Internet and is outside the direct control of the DOE. The Internet also resembles many systems within the purview of the DOE, and can serve as a model for a wide variety of technological networks. There are many excellent data sets available for the Internet, allowing careful model validation.

We have developed a preliminary version of an agent-based model of Internet-like systems, known as ASIM, and shown that by considering traffic, geography and economics, we can model existing networks more accurately than previous models. We propose studying the effects of potential regulatory policies, more realistic traffic models, geographic country-level boundaries, and preplanned systems (such as ESNet), by extending ASIM. Our goal is to enhance understanding of complex networks, particularly with respect to predicting emergent properties and understanding the impact of different policies. We will also study the impact of malicious behavior on complex networks, particularly the dynamic interplay between security countermeasures and attacks. We propose to extend ASIM to include models of common classes of malicious behavior (such as botnets and worms), and the economic burden of security countermeasures. The ASIM software will be released through open-source licensing, so that it can readily be used by the research community.

Our proposal addresses the area of interest in the LAB 09-23 call related to *modeling and simulation of large-scale complex systems*. In particular, ASIM will enable realistic and large-scale simulations at multiple time scales and levels. The ASIM extensions that incorporate attacks and countermeasures will enable us to explore the robustness of complex networks in adversarial environments. Finally, we will address the issue of rigorous validation by expanding our already comprehensive set of real-world data, and comparing it to the simulated data generated by ASIM.

1 Narrative

Complex networks formed by large-scale distributed interconnected systems appear in a wide variety of biological, social, and technological systems, for example, brains, networks of social and economic interactions, and the Internet [51]. Networks such as these are often characterized by their topology, including measures such as degree distribution, node centrality, and shortest path calculations. The flow of data through such systems, however, is often as important as the topological structure, a feature known as *network dynamics*. A further complication arises when we consider how complex networks grow and evolve through time, by adding or deleting nodes and links (sometimes referred to as *graph dynamics*). This proposal focuses on the modeling and simulation of large Internet-like networks, studying how topology, flows, growth and other emergent properties are affected by different factors, such as economics, geography, traffic patterns, regulatory and security policies, and malicious attacks.

As one of the most complex human artifacts ever created, the Internet provides an excellent example of the kind of network we propose to study. In the Internet, dynamic processes of different time scales operate simultaneously, from the slow expansion of new autonomous systems to the speed-of-light propagation of data. The Internet involves a multitude of heterogeneous systems and organizations around the world, including many within the purview of the Department of Energy (DOE). Although the DOE directly administers and controls networks such as ESNet [50], much of the information flow in "open science" depends on the Internet and is outside the direct control of the DOE. Not only does the Internet have profound effects on systems within the purview of the DOE, but it also bears many similarities to those systems, and can serve as an exemplar for a wide variety of technological networks. Furthermore, there are many data sets available for the Internet [56, 14, 57, 38], allowing careful model validation.

Over the past ten years many mathematical and computational models have been developed to characterize complex networks [52, 26, 51]. Although this work has revealed many interesting features about networks, it has emphasized generic network models to explain topological features (such as degree distribution) at the macro-level. To understand and explain complex networks in more detail, including higher-order characteristics (such as radial structure [35]) and growth dynamics, requires incorporating more domain-specific knowledge in the models, such as spatial distributions in models of the Internet [65]. In previous research, we developed a preliminary *agent-based model* of the Internet at the autonomous systems (AS) level, called ASIM, and showed that by considering traffic, geography and economics, we can model existing networks more accurately than previous models [36].

In agent-based modeling [10], entities in the model are represented explicitly; for example, each individual economic agent is represented rather than each different type or class of agents, as is common in other modeling approaches, notably differential equations. An essential feature of agent-based models is the ability to observe how behavior at different spatial and temporal scales arises from local mechanisms. This requires studying interactions among large numbers of components, and to accomplish this agent-based models exclude much real-world detail by design. The research challenge is to define the model components at the proper level of abstraction, neither including irrelevant or incorrect detail, nor leaving out essential features. Components and interactions of the model are encoded as computer programs, allowing researchers to incorporate experimental findings and hypotheses in the model, even those not easily characterized as mathematical equations.

We propose extending ASIM so that we can study how Internet-like systems are affected by

regulatory policies, geographic country-level boundaries, and preplanned systems (such as ESNet). This will enable us to explore a variety of questions, such as, how do government regulations like censorship affect network growth and traffic flows? What are the best policy approaches to influencing network growth? Our goal is to enhance understanding of complex networks, particularly with respect to predicting emergent properties (such as higher-order structures) and understanding the impact of different policies.

We propose further extensions to ASIM so that we can study how malicious behavior affects complex networks, particularly the dynamic interplay between security countermeasures and attacks. We will include models of common classes of malicious behavior (such as botnets and worms), and a simple economic model of security countermeasures in ASIM. We believe this will be the first model to study the technological and economic impact of cyber-attacks on network growth, topology and dynamics. We will explore different aspects of the model, including simulations of "what if" scenarios, for example: if attacks are infrequent, does the drive towards economic competitiveness force organizations to spend less on security and so leave the system as a whole open to catastrophic failure? Are networks actually more robust when attack frequency is higher? If so, what are the implications for policy-making?

A central aspect of the proposed research is model validation. In previous work, we validated ASIM by comparing the topologies and traffic patterns generated by the model to real-world data, using statistical fitting techniques to evaluate degree distributions and radial structure [35]. We found that the radial structures generated by the model closely fit those of the Internet, even though we did not explicitly model the hierarchical structure of the Internet and used very simple models of traffic patterns. We will explore whether modeling more complex patterns of traffic flow and the incorporation of explicit hierarchies through business-level agreements between ASes will improve the accuracy of the model.

We propose to use a variety of mathematical techniques both to validate the model and to investigate new emergent properties and phenomena in complex networks. We will analyze traffic flows generated by the ASIM traffic model in more detail, exploring the application of techniques from network calculus [11] and stochastic processes [40]. We will comprehensively validate the spatial model, for example, using techniques from fractal geometry [43, 29]. We propose to use mathematical tools from data-mining [33], such as clustering techniques, to help extract higher-order structures from the topology. We will investigate the dynamics of network growth using ideas from metabolic scaling theory [47, 48, 64] and others (for example, dynamical systems theory [61] and percolation theory [20]). We will also investigate the applicability of ideas in predation theory to the interaction between attacker and defender [1].

In addition to the research, our proposal involves developing ASIM into a mature software modeling platform. It will be released through open-source licensing, so that it is available to the scientific community, and where possible, we will also make all our data sets available.

1.1 Background

Considerable effort has been devoted to the modeling and analysis of complex networks, using both simulations and the tools of statistical mechanics and graph theory [52, 2, 26, 51]. Advances have been made in understanding common properties exhibited by broad range of real-world networks, from ecological webs to social networks to the electric power grid [51].

1.1.1 Network Models

One of the most common properties exhibited by real-world networks is power-law correlations in many observables. Of particular interest has been the *degree distribution*, $P(k)$, defined as the probability of randomly choosing a vertex with degree k from a graph G. The power-law degree distribution $P(k) \sim k^{-\alpha}$, has been shown to hold for many different networks, for example citation networks ($\alpha = 3$) [54], the Internet ($\alpha = 2.5$) [19, 30], and metabolic networks ($\alpha = 2$) [39]. These networks are known as *scale-free*, because the functional form $f(x) \sim x^{\alpha}$ does not change when rescaling x, i.e. $f(ax) = bf(x)$. In effect, there are no characteristic length scales in the degree distribution (although there often are in other measures).

An important property of scale-free networks is that the nodes are all close together, in the sense that the shortest path between any pair of nodes is small. The shortest path length (*mean geodesic distance*), d has been shown to scale as $d \sim \ln N / \ln \ln N$ for $\alpha = 3$, as $d \sim \ln \ln N$ for $2 < \alpha < 3$, and $d \sim \ln N$ for $\alpha > 3$ [22]. This relates to the well-known "small world" property [63]: even for very large networks, the shortest distance between arbitrary pairs of nodes is small, for example, the mean mean geodesic distance has been calculated at 2.5 for the Internet [19, 30] and 16.2 for the World Wide Web [12].

In their seminal work, Barabasi and Albert [7] (BA) showed that a simple model of random network growth, called *preferential attachment*, produces topologies that exhibit power-law degree distributions. The BA model grows the network from a small set of initial nodes by iteratively adding new nodes, and connecting to each existing v_i with a probability Π that is dependent on the degree k_i of v_i, $\Pi(k_i) = k_i / \sum_j k_j$. The networks that evolve from the BA model exhibit a time-invariant state for the degree distribution that satisfies a power-law with exponent $\alpha = 3$.

The BA model is appealing for its simplicity and broad applicability to degree distributions, but preferential attachment is only likely to be part of the cause of the kinds of network structure we see. Statistics other than degree distribution can reveal additional structure in networks not accounted for by the BA model. For example, neither the degree correlation or clustering coefficient of the Internet are predicted by the BA model [51, 52]. To model accurately specific networks such as the Internet in more detail requires models that capture more realistic features of the real world.

1.1.2 The Internet

There has been a lot of research aimed at understanding and modeling the Internet [52]. The Internet is intriguing because its complexity and size preclude comprehensive study. It comprises millions of individual end nodes connected to tens of thousands of Internet service providers (ISPs) whose relationships are continually in flux and only partially observable. One way to cope with these complexities is by analyzing a single scale of Internet data, for example, a local office network of computers and their interconnections, or a network of email address book contacts, or the network formed by URL links on the World Wide Web, or the interdomain autonomous system (AS) level. In previous work we have focused on the the graph of the AS-level, which exhibits power-law degree distributions (see figure 1). The vertices in the graph are themselves computer networks; roughly speaking, an AS is an independently operated network or set of networks owned by a single entity. Edges represent pairs of ASes that can directly communicate.

The AS-level Internet has been modeled by extending the simple preferential attachment models to include more realistic details, such as geography. The notion of spatial distribution (geography) was introduced by the BRITE topology generator [44], which models nodes distributed across a grid,

connected via preferential attachment rules, but with the probability of a connection decreasing with distance between nodes on the grid. More realistic spatial distributions have been achieved by distributing nodes to form a fractal set [65], similar to the spatial distribution of Internet routers.

Another approach to Internet modeling uses multi-objective optimization, rather than preferential attachment, to obtain the characteristic power-law distributions. For example, Fabrikant et al [28] (FKP) proposed a version of the Highly Optimized Tolerance (HOT) [16] model where new nodes are randomly placed in the unit square and connections are established in a way that minimizes the conflicting objectives of connecting to the nearest nodes and connecting to the most central nodes, where the "centrality" of a node v_i is defined, for example, as the *closeness centrality*, $C_i = (1/n-1)\sum_i \sum_j d(i,j)$, where $d(i,j)$ is the geodesic distance between vertices v_i and v_j. Although the FKP model gives rise to characteristic power-law degree distributions, like the BA model, it fails to reproduce the higher order structures of the Internet, such as the high degree of clustering of nodes [52]. More accuracy has been achieved by extending the HOT model to incorporate economics and geography [17], modeling agents as spatially extended objects [6], and incorporating peer-to-peer decision processes at the AS-level [18].

Our proposed work is based on the ASIM model [36]. The ASIM model is similar in scope to the more advanced HOT models, but differs in the details, most importantly by adding explicit economics in the form of cost. Other differences include accounting for population density, simplifying the treatment of traffic flow, and not assuming a HOT framework (the ASIM model is described in detail in section 1.2). With the ASIM model, we have shown that we can accurately model higher-order structures (such as radial distributions [35]), as well as geography and traffic dynamics (see section 1.2).

1.1.3 Resilience of Networks

The resilience of random networks has received a great deal of attention, beginning with the work of Albert and Barabasi [3], who studied the impact of node deletion on network connectivity. They found that for scale-free networks, iterative deletion of randomly selected nodes had almost no impact on the mean geodesic distance, whereas targeted deletion (the iterative removal of the highest degree node), resulted in a linear increase in mean geodesic distance (see figure 2). From this they concluded that scale-free networks are highly resistant to random failure, but vulnerable to targeted attack, a property that has been termed the Achilles Heel of the Internet [62].

The deletion of nodes not only impacts the mean geodesic distance, but can also result in the appearance of disconnected subgraphs [3]. This aspect of resilience has been studied [15, 21] using percolation theory, which can be used to determine the phase transition to systemic failure, where a network with a giant component becomes a fragmented network of disconnected subgraphs. It has been shown that with random deletions on a scale-free network, there is no phase transition to systemic failure, whereas targeted deletions result in phase transitions to systemic failure, for example, in the Internet the giant component disappears with the deletion of as few as 4.7% of the nodes [52].

In networks that experience flows, such as the Internet and the power-grid, a single failure can lead to a cascade of failures as load gets redistributed and overwhelms remaining nodes, for example route flap storms [42] and power-blackouts [25, 4]. Cascading failures can be analyzed [37, 49, 46] by assuming that the capacity of a node v_i is proportional to its *betweenness centrality* [31], $B_i = \sum_{s,t} \sigma_i(s,t)/\sigma(s,t)$, where $\sigma_i(s,t)$ is the number of geodesic paths between nodes v_s and v_t that run through v_i and $\sigma(s,t)$ is the total number of geodesic paths between v_s and v_t. When a node's

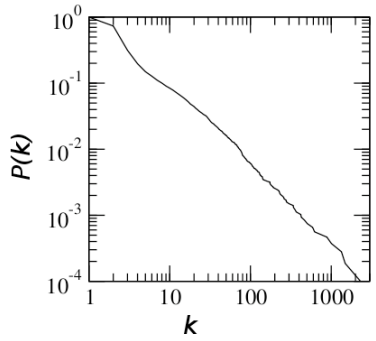

Figure 1: Degree distribution of the Internet at the autonomous systems level. $P(k)$ follows a power-law distribution with exponent $\alpha = 2.5$. Figure reproduced from Chen et al. [19].

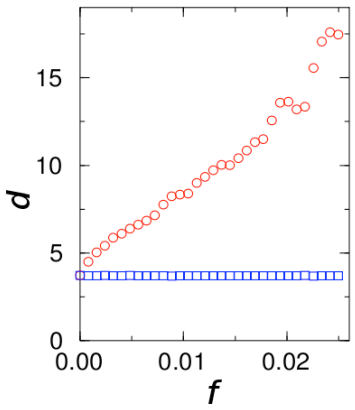

Figure 2: Mean geodesic distance d versus fraction f of nodes removed. The round symbols (upper line) is targeted deletion, and the square symbols (lower line) is random deletion. Figure reproduced from Albert et al [3].

capacity is exceeded, it fails and its load is diverted to the remaining nodes, which could fail in turn. Large-scale cascades can be triggered by the failure of a single key node, namely those with high loads (high betweenness centrality). This result is particularly relevant to the energy security mission of the DOE, because it indicates the vulnerability of power-grids to targeted attack.

1.1.4 Theoretical Models of Network Growth

Insights into the growth of complex networks have been gained through the study of biological scaling laws [13]. One of the models of growth that predicts these scaling laws is the ontogenetic growth model (OGM) [64]. The OGM is a general mechanistic model of organism growth that—in contrast to most previous models—relates model parameters to fundamental biological properties and predicts sigmoidal growth curves that match empirically measured curves for a variety of animal taxa [48]. OGM has been successfully applied to modeling the growth of other complex systems, for example, Bettencourt et al. applied the OGM to the growth of cities [9], by adapting the theory to account for both the diminishing returns of finite resources and network scaling, and by adding a term to account for increasing returns arising from innovation. In contrast to the asymptotic growth curves observed for organisms, their results predict exponential growth curves followed by precipitous crashes.

The OGM is based on conservation of energy: the rate at which energy is devoted to growth (production of new biomass) is equal to the rate at which metabolic energy is assimilated minus the rate at which energy is allocated to maintenance of existing biomass. This can be expressed as $dm/dt = am^{\alpha} - bm^{\beta}$ where m is mass at time t, and a and b are parameters that characterize the amount of energy required to maintain (b) or create (a) a unit of biomass. Depending on the setting, α is usually 3/4, and β is usually 1.0. The 3/4 exponent is observed to hold across a wide variety of organisms, indicating that the model successfully predicts a key emergent property of growing networks.

1.2 AS Simulation Model (ASIM)

ASIM is an agent-based model that attempts to reproduce large-scale features of the AS level of the Internet by modeling localized and well-understood network interactions. The ASes of the Internet lend themselves naturally to discrete agent-based models [10]. Each AS is an economic agent, comprised of a spatially discrete network. Traffic provides income to the ASes, which is then invested in infrastructure, which can then lead to changes in traffic patterns. Over time, ASes create new links to other ASes, upgrade their carrying capacity, and compete for customer traffic. The agents in the ASIM model behave similarly, although they are highly simplified, being designed to be general enough to model any spatially extended communication network built by *economically* driven agents. These agents manage traffic over a geographically extended network (which we refer to as a *subnetwork* to distinguish it from the network of ASes) and profit from the traffic that flows through their network.

We compare the agents to the ASes that comprise the Internet. This is not an exact mapping—some of the Internet Service Providers (ISPs) have many AS numbers (e.g., AT&T), while other ASes are shared by several organizations. We make the common simplifying assumption that once an agent is introduced, it does not merge with another agent or go bankrupt [52, 58, 18]. This is partially justified by the fact that the Internet, from its inception, has grown monotonically, and we seek to capture this dynamic in our model. Most models of the AS graph enforce strict growth [52] as well and are, as ours, justified by their *a posteriori* ability to reproduce measured features.

We assume a network user population distributed over a two-dimensional area. Traffic is simulated by a packet-exchange model, where a packet's source and destination are generated with a probability that is a function of the population profile. The model is initialized with one agent comprised of a subnetwork that spans one grid location (referred to as a *pixel* of the landscape. As time progresses, the agent may extend its subnetwork to other pixels, so that the subnetworks reach a larger fraction of the population. This creates more traffic, which generates profit, which is then reinvested into further network expansion. Through positive feedback, the network grows until it covers the entire population.

An agent i is associated with a set of locations Λ_i (representing sources or end-points of traffic, and peering points), a capacity K_i (limiting the rate of packets that can pass through the agent), a packet-queue Q_i, and a set of neighbor agents Γ_i. A necessary, but not sufficient, condition for two agents to be connected is that their locations overlap in at least one pixel. The locations exist on an $L_x \times L_y$ square grid. A pixel of the grid is characterized by its population $p(x, y)$ and the set of agents with a presence there $\mathcal{A}(x, y)$. The total number of agents in the simulation is denoted by n, and the number of links between agents by m. These quantities, except L_x and L_y, depend on the simulation time. The outer loop of the model then iterates over the following phases:

1. *Network growth.* The number of agents is increased. Existing agents expand geographically, and their capacities are adjusted.
2. *Network traffic.* Packets are created, propagated toward their targets, and delivered. This process is repeated N_{traffic} times before the next network-growth step.

We measure simulation time τ as the number of times phase 1 is executed (the time unit between packet movements is $1/N_{\text{traffic}}$). In the remainder of this section we describe the growth and traffic steps in greater detail.

1.2.1 Network growth

The income of an agent during a time step is proportional to the traffic propagated by the agent during the period. This is a simplification of reality, for example, income could depend both on the amount of traffic and the prices for forwarding the packets set by business agreements. Assume an agent i has a budget B_i that it invests so that it can increase its traffic and thus its profit. Since there is a possibility of congestion in the model, agent i tries first to remove bottlenecks by increasing its capacity K_i (the number of packets that the agent can transit during one time step). When the capacity is sufficient, the agent spends the rest of its budget on increasing its traffic by expanding geographically. There are three prices associated with network growth. The capacity price C_{capacity} is the price of increasing K_i one unit. For simplicity we let C_{capacity} be independent of the size of the agent's subnetwork. The wire price C_{wire} is the price per pixel between a new location and the agent's closest existing location. Finally, C_{connect} is the cost of connecting two agents with locations at the same pixel.

The average degree (number of neighbors of an AS) in the AS graph has been relatively constant over time [52, 24] (increasing about 5% from 2001 to 2007).[1] We take this as a constraint in the model and let the desired average degree k_D be a control parameter. We also assume that each agent tries to spend all of its budget, but not more than that, whenever it is updated.

The network growth phase iterates over the following steps (see figure 3):

1. *Increase of the number of agents.* As long as the network is too dense (i.e. if $2m > k_D n$), new agents are added. New agents are situated in the pixel (x, y) that has the highest available population $p(x, y)/(A(x, y) + 1)$ where $A(x, y)$ is the cardinality of $\mathcal{A}(x, y)$ and $A(x, y) \geq 1$. The budget and capacity of the new agents are initialized to B_{init} and K_{init} respectively.

   If the network is small, $n < k_D + 1$, it is not dense enough for new agents to be added in step 1. Thus, we do not apply this condition when n is less than a threshold n_0 and call the time when $n = n_0$ is reached t_0.

2. *Capacity increase.* Each agent synchronously increases its subnetwork's capacity based upon traffic from the last time step (but not more than the agent can afford). Agent i invests the minimum of $(B_i,\ C_{\text{capacity}}\Delta T_i, 0,\ 0)$ to increase capacity (ΔT_i is the change in traffic propagated by i since the last update).

3. *Link addition.* While $2m \leq nk_D$ (which usually means $k_D - 1$ times), choose two agents randomly that are not already connected and share a common pixel. If the budgets of both agents are larger than C_{connect}, then connect them.

4. *Spatial extension.* Let the agents with remaining budget extend their networks. Iterate through all agents i and add a location at the pixel, not in Λ_i, that has the highest available population $p(x, y)/(L(x, y)+1)$, and is not further than $(B_i - C_{\text{connect}})/C_{\text{wire}}$ from a location in Λ_i (i.e., not further from i than i can afford). (See Figure 3(b)). Alternatively, the algorithm could select the point with the lowest cost per unit of population. However, such an algorithm is computationally prohibitive for studying networks of the Internet's scale.

Each agent's budget is updated immediately after each modification.

[1] This calculation is based on data from Oregon Routeviews, www.routeviews.org. Although more edges of the AS graph can be identified by combining multiple data sources, the Routeviews data set has been compiled in a consistent way over the years, so we believe that the relative degree increase is reliable.

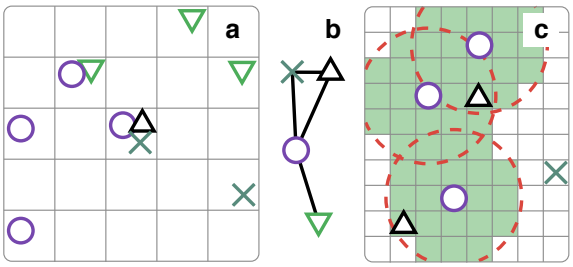

Figure 3: Network growth in ASIM. (a) shows the locations of four spatially distributed agents as different symbols on the geographic grid. These are assumed to be connected by a physical network administrated by the agent, but this is not explicit in the model. (b) is an example graph resulting from (a), and (c) illustrates the area that an agent can afford to expand to (the shaded region). Figure reproduced from Holme et al. [36].

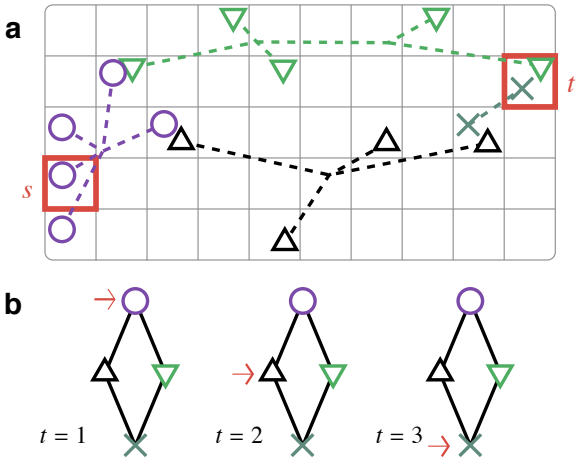

Figure 4: Traffic simulation in ASIM. (a) A packet is propagated from source s to a randomly selected target agent at t. Each agent i queues the packets it receives and can relay K_i packets to neighboring agents. The arrows in (b) symbolize the packet's probabilistic route from source to destination agent. Figure reproduced from Holme et al. [36].

1.2.2 Network traffic

We model traffic with a discrete, packet-exchange model [34, 27]. The packets are generated with specific source and target pixels, but the routing takes place on the network of agents. We neglect intradomain routing among the agent's locations, assuming that the time it takes for a packet to pass through an agent is independent of the specific locations it visits. The dynamics are defined as follows (see figure 4):

1. *Packet generation.* We assume that most traffic originates from direct communication between individuals and does not depend on the distance between them. For each pair of points $[(x, y), (x', y')]$ on the grid, we create a packet with source (x, y) and destination (x', y') with probability $P_{\text{pkg}}\, p(x, y)\, p(x', y')$, where P_{pkg} is a parameter that controls the rate at which new packets are created. Then, an agent is selected at random from those at the source pixel to become the source node. The destination agent is randomly chosen from the agents at the destination pixel. Finally, one unit of credit is added to the sender's budget.
2. *Packet propagation.* Each agent i propagates the first K_i packets from its queue (of length l_i) each time step and receives one unit credit for each propagated packet. A packet can travel only one hop (inter-AS transmission) per time step. A packet at agent i is propagated to a neighbor j with probability $\exp(\lambda(d(i, t) - d(j, t))$ (where t is the recipient AS, $d(\,\cdot\,,\,\cdot\,)$ is the graph distance, and λ is a parameter controlling the deviation from shortest-path routing [59] observed in Ref. [32]).
3. *Packet delivery.* For all agents, delete all packets that have reached their target.

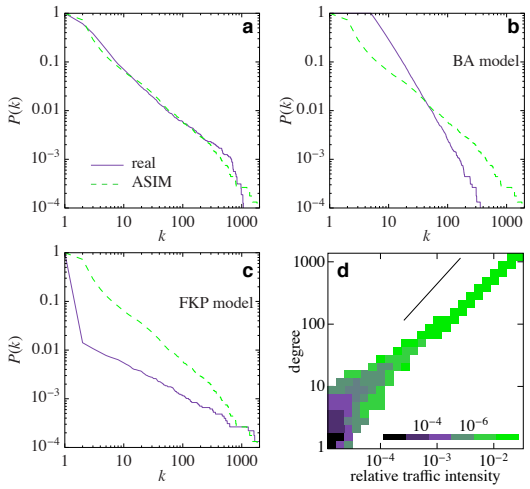

Figure 5: The degree distribution of an AS-graph (AS06) inferred from real data together with degree distribution of a network generated with the ASIM model (a), the BA model (b) and the FKP model (c). Panel (d) is a density plot that illustrates the correlation between traffic and degree in ASIM model runs. Figure reproduced from Holme et al. [36].

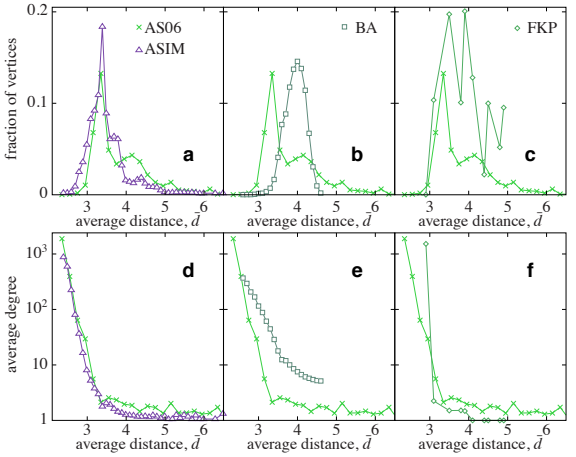

Figure 6: Radial statistics for real and model AS networks. Panels (a)–(c) show the radial densities of nodes for the real AS-graph and the ASIM algorithm (a), the BA (b), and FKP (c) models. Panels (d)–(f) show the average degree vs. average distance $\bar{d}$ for the ASIM algorithm, the BA, and the FKP models, respectively. $\bar{d}$ is the inverse of the closeness centrality. Figure reproduced from Holme et al. [35].

1.2.3 Simulation Results

ASIM provides a simple yet powerful model that generates AS networks that closely match reality and are more accurate than earlier models of Internet growth. Figure 5 shows the degree distribution of networks generated by ASIM compared to those generated by the original BA and FKP models. Looking beyond degree distributions at the higher-order structures of the Internet (figure 6), we can see that ASIM generates more realistic networks than the BA and FKP models. Furthermore, figure 6(a) shows how ASIM-generated networks have peaks and troughs that roughly correspond to the hierarchical levels in the Internet.

1.3 Proposed Research

We propose research to develop models for studying a broad set of questions concerning the impact of policies, regulations, geographic borders, preplanned networks, and security and attacks on networks. To address these questions, we have divided the project into the following emphasis areas: (1) model extensions, (2) simulating cyberattacks and countermeasures, (3) model validation, and (4) software development. In addition to results produced by our own investigations, the project will contribute a mature open-source modeling platform, available to the scientific community.

Our research will build on the ASIM platform. For each proposed extension, we will first study how it affects network topology using the measures described in Section 1.1. Next, we will study the effect on network traffic, using the methods of Ref. [34, 36, 11], and finally the effect if any on network growth using the approaches outlined in Section 1.3.3.

1.3.1 Model Extensions

We propose extensions to ASIM in several areas: traffic modeling, economic models, regulatory policies, and political boundaries.

The current ASIM prototype uses a simplistic traffic model. Although our preliminary work showed that this model yields realistic results, we will implement additional features to see if they have discernible impact. First, we will implement communication rates that depend on the distance between agents, rather than being independent of distance. Second, we will implement service-to-user traffic propagation and compare it to the current method of propagation, which more closely resembles peer-to-peer. Finally, we will extend the routing model to account for business relationships among ASes. In the current prototype, the next hop for routing is selected according to path length. However, commercial incentives dictate route selection based on contractual agreements [55], which govern how ASes exchange traffic on behalf of their customers; most commonly they adhere to the "valley free" rule, in which customers do not transit traffic between providers, and peers do not transit traffic between other peers. The valley free rule is an example of how external factors can impose hierarchies on the structure of the Internet. We will extend ASIM to incorporate a generic form of this constraint (similar to models of business agreements [58, 18]) and study the impact of these hierarchies and other imposed structures on topology, dynamics, and growth.

There are other constraints that we hypothesize will impact network growth. A class of these alter the economics of the model by altering the cost basis. For example, government regulation of packet contents could force providers to implement packet inspection, which could increase the cost of routing, as well as reduce throughput. Regulation may start to directly dictate cost structures. For example, if the Internet were to be regulated as a public utility similar to the power-grid or the telephone network, Internet providers could be forced to provide low-cost connectivity for poorer segments of the populace. The impact on network growth and structure could be even more profound if the constraints were applied asymmetrically, for example, if regulations require only providers of a certain size to inspect traffic. Similar questions arise for quality-of-service guarantees, which could either be mandated or offered as an optional service enhancement. We intend to investigate these issues by extending the model to incorporate alternative, non-uniform cost structures and externally imposed traffic constraints. Such constraints also lay the groundwork for modeling the costs of security countermeasures.

Policies and regulations already vary from one country to the next. Although the current ASIM prototype incorporates geography, it does not model groups of geographically linked ASes that act under a single policy. We will extend ASIM to model countries. A complicating factor is that many ASes extend over multiple countries, and are subject to multiple regulatory regimes, a problem we addressed in [41]. Initially, we will define countries naively by applying boundaries to the geographic grid. Each location on the grid will then be subject to the regulations of the respective country, allowing us to investigate the effects of diverse policies on the network. For example, we can explore the issues of country-level censorship, and what strategies a country or group of countries might use to route around or avoid a country that was known to eavesdrop or censor [41].

ASIM currently models the commercial Internet, where growth is driven by market forces. However, many governments build large, preplanned networks for various purposes, such as military communication, control of information, or for scientific collaboration (e.g. ESNet, Internet2). These networks often have their own cost structure and regulatory policies, and they usually interface with the Internet, which could have unforeseen consequences for both kinds of network. We are unaware of any research that has modeled the effect of preplanned networks on the growth of unstructured

complex systems such as the Internet; we propose to begin exploring this interaction by extending ASIM to model subnets that are planned, operate under their own policies, and interconnect with commercial networks. This aspect of the project is particularly relevant to the Science Mission of the DOE, which promotes open science through large-scale, preplanned collaborative networks such as ESNet.

For each extension, we will study the effect on network topology, for instance, the degree distribution and the radial structure. Some extensions could change not only the exponent of the power-law currently seen in the degree-distribution, but could even change the form of the distribution itself. We will run "what-if" scenarios to determine if the fundamentals of the network topology can be changed by external factors such as government regulation. We will also measure the effect of the model extensions on traffic flow patterns, particularly throughput and latency. Throughput can be measured as the number of packets transferred per time step, and latency is the number of steps it takes for a packet to reach its destination. Throughput and latency will enable us to measure the efficiency of the network in a more realistic manner than simply considering topological metrics. We hypothesize that extensions such as business agreements and government regulations will affect throughput and latency, even if they do not affect the degree distribution or other aspects of the network topology.

1.3.2 Attacks and Countermeasures

Most attacks happen on much shorter timescales than the timescale on which network topology changes, and hence any individual attack is unlikely to have long-term impact. However, the chronic onslaught of multiple attacks could change how the network evolves, similarly to chronic parasitic infections stunting an organism's growth or impairing its long-term fitness. We plan to investigate this interaction by extending ASIM to incorporate models of chronic attacks. We are particularly interested in the economic impact of chronic or large-scale attacks and will emphasize that in our studies. We will implement generic versions of several forms of attack, including node deletion, denial-of-service, botnets, worms, and routing attacks.

In node deletion, nodes are removed from the network, simulating infrastructure attacks or simple node failure. Previous studies [3] have shown that scale-free networks are resilient to random node deletion, but vulnerable to targeted deletion. However, these studies considered only topological and reachability effects at a single instant in time, and ignored the impact of deletions on the growth and evolution of the network. We will study chronic, recurring deletions, and we will consider both targeted and random deletions.

Moving beyond node deletion, we will study other classes of attacks that impact directly on traffic flow and economic costs. In ASIM the income of an agent is proportional to the traffic it propagates. Thus, we can simulate several attack classes as spurious traffic, for which agents receive no income; this enables the model to reflect the economic impact of attacks. The generation and distribution of the spurious traffic will largely depend on the form of attack, for example, a denial-of-service (DoS) attack involves large volumes of traffic flowing from multiple locations, directed at a single target location, whereas spam consists of traffic flowing from many locations to many other locations, and worms generate traffic that begins in a few locations and rapidly spreads to many locations.

We will represent these various forms of attack by overlaying spurious traffic generators on the ASIM agents. These can be regarded as parasitic networks that exist on top of the normal, commercial networks. Of particular interest is modeling the growth of botnets, because botnets

are used extensively to generate spam and launch DoS attacks, and may determine the spatial distribution of spurious traffic. We will model both long-lived generators, such as botnets that evade detection, and short-lived generators, such as worms that spread rapidly and then die out as the worm is detected and eliminated.

Another class of attacks targets the routing infrastructure itself, for example incorrectly rerouting traffic, refusing to route traffic, and "black-hole" routing, where a router takes in traffic but does not pass it on. We will simulate these attacks by applying changes in the traffic model to the compromised routers. A challenge of simulating routing infrastructure attacks is assigning a cost or negative impact. Our initial idea is to regard packets that do not reach their destination because of rerouting or black-holes as spurious traffic, which hence imposes a burden but no income on the ASes.

The interaction between security countermeasures and attacks forms a continual arms-race, which we hypothesize will affect the evolution of the Internet and all other complex networks that are subject to exploitation by malicious elements. We propose a simple economic model of countermeasures, assuming that agents can spend part of their budget on security, which will reduce the number of compromised nodes and reduce the flow of spurious traffic. The effectiveness of the countermeasures can be modeled by relating the amount spent to the probability of success. Because this relationship may not be linear (it could be a step function), we plan to investigate various different cost relationships.

To summarize, we propose to investigate the impact of ongoing attacks on several aspects of the network: its stability when subject to perturbations caused by attacks, the impact of chronic attacks, and the effect of regulations and security measures. Our previous work suggests that several characteristics of the model such as the degree distribution are independent of time; will this still be the case in the presence of attacks? Another important issue is whether preplanned networks are more or less resilient to attacks, and how integrating preplanned subnets into the randomly generated networks changes the resilience of the overall network. At the heart of this investigation is the interplay between attacks and security countermeasures. Cost competitiveness is a driver towards lower security and the impact of attacks is a driver towards higher security. We are interested in how the frequency of attacks affects this trade-off. For example, if the attack frequency is sufficiently low, do agents neglect security sufficiently that the network is vulnerable to catastrophic failure? Is the network actually more resilient when subject to higher attack frequencies?

1.3.3 Model Validation

Although we are interested in the basic principles governing complex networks generally, in previous work we focused on the Internet because it is an obvious and compelling example with a wealth of available data. Using real-world data, we demonstrated that ASIM generates networks that closely match the degree distribution and radial structure of real networks, and is more accurate than other commonly used models (as shown in figures 5 and 6). Furthermore, ASIM generates traffic patterns that closely match those occurring in the real Internet. For example, our probabilistic propagation method (section 1.2.2) has a similar effect on *average* path length—the excess distance of real paths traveled compared to the shortest graph distance—as that observed for real Internet traffic [32].

As we extend the model, we intend to draw from the many available sources of real-world data [56, 14, 57, 38] to ensure that ASIM still provides an accurate model of the Internet. One example concerns hierarchical structure of routing and whether incorporating business agreements

into the model improves model accuracy or not. Another example concerns the round-trip-time (RTT) between nodes, for which there is readily available data, for example, the RTT normalized by the geographic distance shows a distinct power-law distribution, with exponent $\alpha = -2$ [53]. Yet another example is the dynamic evolution of the network structure, which we can compare to the connectivity maps collected over the last ten years by various organizations [14, 56].

In previous work, we validated the spatial distributions produced by the model by seeding it with the population density of the US, and observing that the resulting networks closely matched existing AS networks. This limited form of validation showed that the model produces realistic spatial distributions given real-world empirical starting conditions. We want to ensure that the spatial distributions are valid for generic starting conditions. In this case a direct mapping of the network produced by the model to real-world data cannot be used to validate the spatial distribution. However, there are other methods we can use, for example, it has been shown that the distribution of routers in the Internet is fractal [65], as measured using the box counting dimension [43]. If the spatial distributions generated by ASIM result in similar box counting dimensions, then we have further validation of the spatial aspect of the model.

Investigating the impact of attacks and security countermeasures is an addition that will require careful validation. Although predicting and modeling individual and isolated attacks is difficult in general, we are interested in the large scale, where the rate of attacks, spread of malware, growth of botnets, etc, is more amenable to modeling. Fortunately, there is much data available, for example, on the spread of network worms [45] and email viruses [5], the impact of worms on Internet routing [23], and the extent and spread of botnets [60]. The efficacy of countermeasures can be derived from various data about the cost of security and the failure rate of defenses such as network intrusion detection systems and antivirus software.

There are several theoretical approaches to exploring and validating network growth. Our first approach will be to adapt OGM—the ontogenetic growth model (section 1.1.4)—to describe network growth in ASIM. We observe that population density (consumers) plays the role of energy availability to the system, there are close parallels between network scaling in biological and computational systems [47] (although the geometry is two-dimensional instead of three-dimensional), and that the economic model of ASIM will govern the trade-off between maintenance and expansion. This approach will complement the data-driven approach to validation that we have already described. We will also explore other theoretical approaches such as percolation theory [20], which has proved useful in the analysis of resilience, and could help to identify critical points in the dynamics, particularly with regard to attacks.

1.3.4 Software Development

To further research and collaboration with the scientific community, we will develop a readily-available, mature, open-source modeling platform. The current prototype code is written in C++ and is available at http://www.tp.umu.se/~holme/ãsim/. We will rewrite the prototype using Unified Parallel C (UPC), a Partitioned Global Address Space (PGAS) language that is designed to run efficiently on large-scale distributed-memory computers. This will enable us to run large-scale simulations efficiently, even with the added complexity of the proposed extensions. The Berkeley UPC (BUPC) implementation from Lawrence Berkeley National Laboratory [8] is the most mature and portable UPC implementation and is a logical choice given the close collaboration between PI Hofmeyr and the developers of BUPC.

The software will be made publicly available under the GPL license, and where possible, we will

make all our data publicly available, so that other researchers can use our repositories to validate their own models. Much of the data will obtained from other organizations and hence be subject to other restrictions on distribution.

1.4 Organization

The proposed budget is $393K/year, with $193K spent at LBNL and $200K at the UNM campus. Hofmeyr has 10 years of experience of applied research in the computer industry, managing teams of engineers and directing the research efforts of a commercial cybersecurity company. He will be the lead software architect and primarily responsible for overseeing the software development aspect of the project. This will include decisions about software environments, overall design, coding practices and code reviews, as well as managing software releases.

Forrest has nearly 20 years of experience conducting interdisciplinary research, training students and managing large grants, primarily in biomodeling and computer security. Hofmeyr and Forrest will share responsibility for decisions about model development and experimental design, for example, decisions about which features to add to the model, how to represent attacks such as botnet activity in the model, and designing specific experiments to test hypotheses about the model. The PIs will also share responsibility for written dissemination of the project results, through published papers. The PIs have a long history of productive collaboration, beginning when Hofmeyr was a student at UNM, continuing with the co-founding of a security company, and most recently co-authoring a retrospective paper for a tight deadline.

The students will be physically located in the UNM Computer Science Department and they will be part of the Adaptive Computation Laboratory (ACL) at UNM, directed by Forrest. The students will be assigned offices in the 1000sq. ft. facility and will participate in weekly laboratory meetings. The students will spend the summer at LBNL, further facilitating interaction between UNM and LBNL.

1.5 Project Timetable

The research agenda of the project consists of model design, experimentation, simulation and mathematical analyzes, software development and real-world data gathering (for validation). Information will be disseminated in the form of high-quality research publications and ongoing interactions with the research community. We plan on one major software release per year, for each of the three years (there may also be minor releases at other times).

Year 1: During the first year we will rewrite ASIM in UPC (by Q2), design and implement a set of extensions, including new traffic models and constraints that model business agreements (by Q3). We will devise experiments to validate the extensions and further aspects of the basic model (such as traffic flow, spatial distributions and network growth) against current data sets by Q4. Concurrently, we will begin exploring new mathematical techniques for analysis, such as network calculus for traffic flows, fractal geometry for spatial aspects, and scaling theory for growth, by Q4. We will establish a project website by Q2 and release the first major software version by Q4.

Year 2: During the second year we will extend ASIM to include attack classes and security countermeasures by Q2. We will devise experiments to validate the security countermeasures and attack

extensions by Q3, and we will carry out experiments to investigate the dynamics of the cybersecurity arms race and other aspects of security, by Q4. We will continue the ongoing mathematical analyzes of the traffic and spatial components of the model, and explore the application of techniques from predation theory and percolation theory to the understanding of attacker-defender dynamics. We will release the second major software version by Q4. We will acquire additional data sets, with one of the graduate students assigned full-time to the task of preprocessing the data to remove errors and running algorithms to infer probable graph structure and dynamics from the input data.

Year 3: During the third year we will extend ASIM to include government regulations, country boundaries and preplanned networks, by Q2. We will devise experiments to explore the impact of the new features and we will further extend the existing analyzes; in particular we will investigate new approaches to understanding networks generated by the combination of random and preplanned growth, such as the dynamical systems theories that focus on small-world graph models, by Q4. In Q3-Q4, we will focus on project reports, documentation, dissemination of information, and ensuring that we have a mature software platform. We will release the third and final major software version by Q4.

1.6 Collaboration and Communication

PI Forrest is an External Professor at the Santa Fe Institute (SFI) and serves on its Science board, an affiliation which gives the project access to researchers who study complex networks and scaling. For example, Forrest and Mark Newman, a statistical physicist, have collaborated in the past on mathematical and computational models of technological networks. Newman was mentor to Prof. Petter Holme, currently at the Royal Institute of Technology, Stockholm, and the author of the current ASIM prototype. Holme is interested in collaborating actively on the project. Forrest also collaborates closely with other research groups at LANL and UNM, for example, the West and Brown scaling group, jointly working on projects to reformulate metabolic scaling theory for computational problems. In the computer networking domain, PI Forrest collaborated with Prof. Jennifer Rexford of Princeton University on the development of Pretty Good BGP, a distributed anomaly detection and response system for BGP. Rexford contributes valuable expertise on commercial computer networks. In addition to her own expertise in computer security, Forrest has close contacts in industry, such as Matt Williamson of AVG.

PI Hofmeyr is currently collaborating with researchers, such as Prof. John Kubiatowicz, at the UC Berkeley Parallel Computing Laboratory, on a project to design a new operating system for multicore computers, using ideas from distributed systems. In addition, as a consequence of starting a computer security company, Hofmeyr is still closely tied to the computer security industry, through connections like Elias Levy at Symantec, who was the founder of SecurityFocus.

2 Literature Cited

[1] P. A. Abrams. The evolution of predator-prey interactions: theory and evidence. *Annual Review of Ecology and Systematics*, 31:79–105, 2000.

[2] R. Albert and A.-L. Barabási. Statistical mechanics of complex networks. *Rev. Mod. Phys*, 74:47–98, 2002.

[3] R. Albert, H. Jeong, and A.-L. Barabási. Attack and error tolerance of complex networks. *Nature*, 406:378–382, 2000.

[4] M. Anghel, K. A. Werley, and A. E. Motter. Stochastic model for power grid dynamics. In *Fortieth Hawaii International Conference on System Sciences*, 2007.

[5] J. Balthrop, S. Forrest, M. E. J. Newman, and M. M. Williamson. Technological networks and the spread of computer viruses. *Science*, 527, 2004.

[6] S. Bar, M. Gonena, and A. Wool. A geographic directed preferential Internet topology model. *Computer Networks*, 51:4174–4188, 2007.

[7] A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286:509–512, 1999.

[8] BerkeleyUPC. http://upc.lbl.gov.

[9] L. M. A. Bettencourt, J. Lobo, D. Helbing, C. Khnert, and G. B. West. Growth, innovation, scaling, and the pace of life in cities. *PNAS*, 2007. doi:10.1073/pnas.0610172104.

[10] E. Bonabeau. Agent-based modeling: Methods and techniques for simulating human systems. *Proc Natl Acad Sci*, 99:7280–7287, 2002.

[11] J.-Y. L. Boudec and P. Thiran. *Network calculus: a theory of deterministic queuing systems for the Internet*. Springer, 2001.

[12] A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tomkins, and J. Wiener. Graph structure in the web. *Computer Networks*, 33:309–320, 2000.

[13] J. H. Brown and G. B. West, editors. *Scaling in biology*. Oxford University Press, 2000.

[14] CAIDA. http://www.caida.org.

[15] D. S. Callaway, M. E. J. Newman, S. H. Strogatz, and D. J. Watts. Network robustness and fragility: percolation on random graphs. *Physics Review Letters*, 85:5468–5471, 2000.

[16] J. M. Carlson and J. Doyle. Highly optimized tolerance: a mechanism for power laws in designed systems. *Phys. Rev. E*, 60:1412–1427, August 1999.

[17] H. Chang, S. Jamin, and W. Willinger. Internet connectivity at the AS-level: an optimization-driven modeling approach. In *MoMeTools '03: Proceedings of the ACM SIGCOMM workshop on Models, methods and tools for reproducible network research*, pp. 33–46, New York, NY, USA, 2003. ACM.

[18] H. Chang, S. Jamin, and W. Willinger. To peer or not to peer: Modeling the evolution of the Internet's AS-level topology. In *Proc. IEEE INFOCOM*, 2006.

[19] Q. Chen, H. Chang, R. Govindan, S. Jamin, S. J. Shenker, and W. Willinger. The origin of power laws in internet topologies revisited. In *Proceedings of the 21st Annual Joint Conference of the IEEE Computer and Communications*, 2002.

[20] R. Cohen, K. Erez, D. ben Avraham, and S. Havlin. Resilience of the internet to random breakdowns. *Physics Review Letters*, 85:4626–4628, 2000.

[21] R. Cohen, K. Erez, D. ben Avraham, and S. Havlin. Resilience of the Internet to random breakdowns. *Phys. Rev. Lett.*, 85:4626–4628, 2000.

[22] R. Cohen and S. Havlin. Scale-free networks are ultrasmall. *Physics Review Letters*,

90(058701), 2003.

[23] J. Cowie, A. T. Ogielski, B. J. Premore, and Y. Yuan. Internet worms and global routing instabilities. In *Proceedings of SPIE*, 2002.

[24] I. Daubechies, K. Drakakis, and T. Khovanova. A detailed study of the attachment strategies of new autonomous systems in the AS connectivity graph. *Internet Mathematics*, 2:185–246, 2006.

[25] I. Dobson, B. A. Carreras, V. E. Lynch, and D. E. Newman. Complex systems analysis of series of blackouts: cascading failure, critical points, and self-organization. *Chaos*, 17(2), 2007.

[26] S. N. Dorogovtsev and J. F. F. Mendes. *Evolution of Networks: From Biological Nets to the Internet and WWW*. Oxford University Press, Oxford, 2003.

[27] P. Echenique, J. Gómez-Gardẽnes, and Y. Moreno. Dynamics of jamming transitions in complex networks. *Europhys. Lett.*, 71:325–331, 2005.

[28] A. Fabrikant, E. Koutsoupias, and C. H. Papadimitriou. Heuristically optimized trade-offs: A new paradigm for power laws in the Internet. In *Proceedings of the 29th International Conference on Automata, Languages, and Programming*, Lecture notes in Computer science 2380, pp. 110–122, Heidelberg, 2002. Springer.

[29] K. Falconer. *Fractal Geometry: Mathematical Foundations and Applications*. John Wiley & Sons, 2003.

[30] M. Faloutsos, P. Faloutsos, and C. Faloutsos. On power-law relationships of the Internet topology. *Comput. Commun. Rev.*, 29:251–262, 1999.

[31] L. Freeman. A set of measures of centrality based upon betweenness. *Sociometry*, 40:34–41, 1977.

[32] L. Gao and F. Wang. The extent of AS path inflation by routing policies. In *Proceedings of GLOBECOM 2002*, pp. 2180–2184, 2002.

[33] J. Han and M. Kamber. *Data mining: concepts and techniques*. Morgan Kaufmann, 2006.

[34] P. Holme. Congestion and centrality in traffic flow on complex networks. *Advances in Complex Systems*, 6:163–176, 2003.

[35] P. Holme, J. Karlin, and S. Forrest. Radial structure of the Internet. *Proc. R. Soc. A*, 463:1231–1246, 2007.

[36] P. Holme, J. Karlin, and S. Forrest. An integrated model of traffic, geography and economy in the internet. *ACM SIGCOMM Computer Communication Review*, 38(3):7–15, 2008.

[37] P. Holme and B. J. Kim. Vertex overload breakdown in evolving networks. *Physical Review E*, 65(066109), 2002.

[38] iPlane. http://iplane.cs.washington.edu/.

[39] H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai, and A. L. Barabasi. The large-scale organization of metabolic networks. *Nature*, 407:651–654, 2000.

[40] T. Karagiannis, M. Molle, and M. Faloutsos. A nonstationary poisson view of internet traffic. In *Proceedings IEEE Infocom*. IEEE CS Press, 2004.

[41] J. Karlin, S. Forrest, and J. Rexford. Nation-state routing: Censorship, wiretapping, and bgp. http://arxiv.org/abs/0903.3218, 2009.

[42] C. Labovitz, G. R. Malan, and F. Jahanian. Internet routing instability. *IEEE/ACM Transactions on Networking*, 6(5):515–528, 1998.

[43] B. B. Mandelbrot. *The fractal geometry of nature*. Macmillan, 1982.

[44] A. Medina, I. Matta, and J. Byers. On the origin of power laws in Internet topologies. *ACM Computer Communication Review*, 30(2):18–28, 2000.

[45] D. Moore, C. Shannon, and K. Claffy. Code-red: a case study on the spread and victims of an internet worm. In *IMW '02: Proceedings of the 2nd ACM SIGCOMM Workshop on Internet measurment*, pp. 273–284, New York, NY, USA, 2002. ACM.

[46] Y. Moreno, R. Pastor-Satorras, A. Vazquez, and A. Vespignani. Critical load and congestion instabilities in scale-free networks. *Europhysics Letters*, 62:292–298, 2003.

[47] M. Moses, S. Forrest, A. L. Davis, and J. H. Brown. Scaling theory for information networks. *Journal of the Royal Society Interface*, 2008.

[48] M. E. Moses, C. Hou, W. H. Woodruff, G. B. West, J. C. Nekola, W. Zuo, and J. H. Brown. Revisiting a model of ontogenetic growth: Estimating model parameters from theory and data. *The American Naturalist*, 171(5):632–645, 2008. DOI10.1086/587073.

[49] A. E. Motter and Y.-C. Lai. Cascade-based attacks on complex networks. *Physical Review E*, 66(065102), 2002.

[50] E. S. Network. http://es.net.

[51] M. E. J. Newman. The structure and function of complex networks. *SIAM Review*, 45:167–256, 2003.

[52] R. Pastor-Santorras and A. Vespignani. *Evolution and structure of the Internet: a statistical physics approach*. Cambridge Univeristy Press, Cambridge, 2004.

[53] R. Percacci and A. Vespignani. Scale-free behavior of the internet global performance. *Europhysics Letters B*, 32:411–414, 2003.

[54] S. Redner. How popular is your paper? an empirical study of citation distribution. *European Physics Journal B*, 4:131–134, 1998.

[55] Y. Rekhter and T. Li. A Border Gateway Protocol 4 (BGP–4). RFC 1771 (Draft Standard), Mar. 1995. Obsoleted by RFC 4271.

[56] Routeviews. http://routeviews.org.

[57] R. R. I. Service. http://routeviews.org.

[58] S. Shakkottai, T. Vest, D. Krioukov, and K. C. Claffy. Economic evolution of the Internet AS-level ecosystem. e-print arxiv:cs.NI/0608058, 2006.

[59] V. Sood and P. Grassberger. Localization transition of biased random walks on random networks. *Phys. Rev. Lett.*, 99:098701, 2007.

[60] B. Stone-Gross, M. Cova, L. Cavallaro, B. Gilbert, M. Szydlowski, R. Kemmerer, C. Kruegel, and G. Vigna. Your botnet is my botnet: Analysis of a botnet takeover. Technical report, University of California, Santa Barbara, 2009.

[61] S. H. Strogatz. Exploring complex networks. *Nature*, 410:268–276, 2001.

[62] Y. Tu. How robust is the internet? *Nature*, 406:353–354, 2000.

[63] D. J. Watts and S. H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393:440–442, 1998.

[64] G. B. West, J. H. Brown, and B. J. Enquist. A general model for ontogenetic growth. *Nature*, 413:628–631, 2001.

[65] S.-H. Yook, H. Jeong, and A.-L. Barabási. Modeling the Internet's large-scale topology. *Proc. Natl. Acad. Sci. USA*, 99:13382–13386, 2002.