

Running head: MODELING OF ATTACKS

Modeling of Attacks over Scale-Free
Computer Networks using Colored Petri Nets

Dissertation Proposal

Submitted to the College of Technology

Eastern Michigan University

By Lawrence M. Healy

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

Dissertation Committee:

Ann Remp, Ph.D., Dissertation Chair

Peter Stephenson, Ph.D., EMU Graduate School Representative

Mary Brake, Ph.D.

Jae Park, Ph.D.

William Sverdlik, Ph.D.

December 12, 2007

Ypsilanti, Michigan

Table of Contents

List of Table/s	iii
List of Figure/s	iv
Introduction.....	1
Significance of the Problem.....	6
Nature of the Problem.....	9
Scale-free Network Theory.....	9
Preferential Attachment	11
Cyber Attack Markers and Preferential Attachment Theory	13
Research Contributions.....	15
Research Objective and Question	16
Research Limitations	16
Methodology.....	18
Baseline Samples of Ordered Node-pairs.....	19
General Scheme for the Simulation.....	22
Selection of CPN Modeling and Simulation Language.....	28
Development of CPN Modules.....	32
Initial CPN Attack Simulation.....	35
Data Collection and Analysis.....	46
Timeline	48
Conclusion	49
References.....	51

LIST OF TABLE(S)

Table	Page
1. Proposed Data Structures	34
2. Research Timeline	48

LIST OF FIGURE(S)

Figure		Page
1.	Live Internet Data, Power Law Degree Distribution	12
2.	Live Internet Data, Preferential Attachment and Linking Probability.....	13
3.	Node-pairs and Network Connectivity Model.....	20
4.	Attack-Flow Overview - Baseline	22
5.	Process Overview.....	24
6.	Colored Petri Net Example - Key	30
7.	Colored Petri Net Example – Before Transition	31
8.	Colored Petri Net Example – After Transition	31
9.	Attack-Flow Overview - Key	38
10.	Attack-Flow Overview – Attack Implementation	39
11.	Attack-Flow Overview – Attack Implementation Detail.....	40
12.	Attack-Flow Overview – Cascade Affect	41
13.	Attack-Flow Overview – Potential Link Discovery	42
14.	Attack Simulation Example	45

Modeling of Attacks over Scale-Free Computer Networks using Colored Petri Nets

The purpose of the research presented in this proposal is to determine the feasibility of inferring the existence of a cyber attack over the Internet's router infrastructure through Colored Petri Net modeling and simulation. We surmise that two problems must be solved to enable a successful cyber attack investigation. We must be able to infer (within an acceptable degree of certainty) an attack against the Internet's infrastructure has occurred and once we have relative confidence of the attack's existence, we must then determine its origin. Before or in concert with potential solutions for determining an attack's origin, we must lay the foundation to develop tools to determine the existence of an attack over the Internet's vast and complex infrastructure. Therefore, the objective of the research proposed in this paper is to provide the foundational feasibility of this first step.

For a contextual understanding, this section we will present discussion relating to cyber attack investigation mechanisms as well as issues associated with the determination of an attack's origin. This section seeks to present the relevancy and the associated difficulties in the determination of a cyber attack existence over the Internet's connectivity topology. This section will present background discussion with regards to inherent difficulties associated with tracking and tracing internet communication packets and how these difficulties relate to cyber attack investigations. The nature of the problem relating to a relevant theory of scale-free computer networks, preferential attachment and a potentially useful cyber attack mechanism theory will then be discussed. We will then present the proposed contributions to the field of cyber attack theory, our primary research objectives and scope. The second section of this research proposal will discuss the proposed methods for a Colored Petri Net model and simulation.

Over the last six years, computer vulnerabilities and exploits have grown significantly and do not show any sign of slowing down. A 2006 FBI computer crime survey (Gordon, Loeb, Lucyshyn, & Richardson, 2006) reported that from 1999 to 2005, 50% to 60% of the organizations surveyed had experienced unauthorized use of their computer systems. This research proposal will focus on potentially catastrophic cyber attacks targeted against the message communication mechanisms of the Internet's router infrastructure (as opposed to enterprise level attacks).

Similar to the current wave of denial-of-service attacks at enterprise level networks (end-user), a targeted attack against the Internet's router infrastructure might be implemented to halt or degrade inter-router communication. While denial-of-service attacks against a corporate network might result in a single Internet Service Provider's (ISP) service degradation, a similar type of cyber attack targeted at the Internet's router infrastructure backbone might lead to a catastrophic degradation of the internet's ability to function normally. Recently, these types of attacks against the end-user systems over the Internet have become quite common and show no signs of slowing down. For instance, an FBI computer crime study indicated that 25% of the unauthorized computer access exploits were related to denial-of-service attacks over the Internet (Gordon et al., 2006). Between 2001 to 2004, there were over 68,000 denial-of-service attacks directed at over 34,000 distinct victim IP addresses (Moore, Shannon, Brown, Voelker, & Savage, 2006).

The Internet's communication and its ability to provide efficient service is dependent upon 13 Domain Name System (DNS) root servers located around the world. These 13 DNS root servers are the Internet's backbone infrastructure. They are responsible for the coordination of the Internet communication information (such as Internet Protocol (IP) address translation and

assignment) that is essential for the Internet's normal operation. As a result of the varied Internet infrastructure survivability mechanisms, the Internet can sustain limited damage concurrently to a few of these root servers without experiencing major service degradation (Peng, Leckie, & Ramamohanarao, 2007). However, an attack that cripples a sufficient number of these servers might be ample stimulus to severely limit Internet communication (Cheung, 2006).

The root servers are administered in cooperation with the International Corporation for Assigned Names and Numbers (ICANN). They published a report pertaining to the occurrence of two targeted denial-of-service attacks against the DNS root servers (critical Internet infrastructure) over the past 5 years (International Corporation for Assigned Names and Numbers, 2007). In this report, ICANN noted that there were targeted denial-of-service attacks against the root servers in October, 2002 and February 2007.

In October 2002, over a one hour period, nine root servers for experienced a denial-of-service attack that degraded Internet service (Peng et al., 2007). ICANN also notes that over a two-hour period in February 2007, six root servers were attacked with two of the servers being severely damaged. While both of these infrastructure attacks resulted in minimal damage, the degree of the attack complexity and its nature may indicate future attack strategies. Due to the survivability mechanisms of the Internet and the short attack span, the extent of the damage was limited. However, the instance may provide insight into the potential for larger and more damaging attacks on the Internet's infrastructure and "had the attacker increased the attack traffic rate or extended the attack time, more catastrophic damage would have been done to the overall Internet" (Peng et al., 2007, p. 15).

One characteristic of the aforementioned large-scale infrastructure attacks that amplifies its consequences is the probable likelihood of an attacker will maintain anonymity. Identity

concealment techniques such as the use of surrogate attack sources have made identification of a perpetrator's system of origin extremely difficult (Daniels, 2002; Martins, 2005). Daniels argues that "network origin concealment systems" are commonly used by attackers to conceal the attacker's network origin and identity. He describes "network origin concealment systems" as computer network mechanisms that provide an attacker with the ability to create or modify network data elements (such as IP packets) as they are forwarded through multiple network nodes to their ultimate destination.

The Internet's design for processing inter-node communication presents a challenging environment for the understanding of cyber attacks and the mechanisms used for exploiting computer system vulnerabilities. As argued by Lipson (2002) the Internet's inherent design severely inhibits the relevant tracking and tracing of IP packets partially due to the following Internet communication processing design assumptions:

- The user community would be trustworthy and wouldn't seek to obfuscate identity or manipulate the Internet's communication mechanisms for malicious purposes.
- Due to the cooperative and collaborative research driven environment, significant and performance prohibitive security audit functionalities were not essential.
- High speed traffic and performance requirements were essential and therefore any attempts at significant tracking would be too costly in terms of system performance criteria.
- Due to the large volume of packets in a relatively short timeframe, storage of packet information was not viable.

“Network origin concealment systems” are significantly enabled by the Internet’s inherent lack of tracking and tracing mechanisms and subsequently inhibit investigations of anomalous computer network behavior and the reconstruction of network events and attack path reconstruction leading to uncertain investigatory conclusions (Casey, 2002; Daniels & Spafford, 2000a; Wang et al., 2006). This is evident in the literature of cyber conflict: “In the cyberrealm, distinguishing potential terrorist activity from normal system failures, exploratory hacking, and other threats such as espionage is very difficult ” (Rattray, 2001).

Currently, Internet traffic measurement techniques are, for the most part, incompatible with existing network protocols resulting in investigative complexities (Casey, 2004). As network traffic between routers travel in IP packets, it is relatively simple to maliciously modify the IP packet to conceal the true source of the network message. This attack technique is known as IP spoofing. Usually IP spoofing is employed along with another common attack technique known as packet redirection. The result of packet redirection is that the attacker maliciously violates a surrogate machine to launch its intrusion. This makes it appear as if the attack is coming from another machine or set of machines. However, the machine that appears to be implementing the attack is actually a victim being used as a surrogate. These attack origin concealment techniques significantly constrict an investigation’s ability to reconstruct the attack path and backtrack the attack to its source (Daniels & Spafford, 2000b). Unfortunately, determining the origin of attack is a difficult and time-consuming task that may lead to inconclusive results (Daniels, 2002).

A properly conducted end-to-end digital investigation of a malicious exploit consists of a “set of steps the investigator must perform in order to preserve, collect, examine and analyze digital evidence” (Stephenson, 2003a). Stephenson argues that a comprehensive investigative

process consists of the following: collection of evidence, analysis of events, preliminary correlation of events, event conflict resolution, second level correlation of events, timeline analysis, and chain of evidence construction and corroboration of events. Unfortunately, there are very few appropriately adequate investigative tools for a reliable and consistent digital investigation leading to the determination of an attack's existence and origin (Carrier, 2003; Casey, 2002; Institute for Security Technology Studies, 2002, 2004a, 2004b).

Determination of an attack's existence is a requisite task prior to the conduct of an end-to-end- investigation to ascertain a cyber attack's origin. Due to the inherent complexities of very large network infrastructures, such as the Internet, the determination of an attack's existence is significantly more problematic than for a small enterprise network infrastructure. Consequently, the rigors for determining an attack's origin over the Internet are magnified.

This research is primarily a feasibility study that seeks to provide the first step for determining a cyber attack's origin through determination of an attack's existence. Subsequently, if the premise is feasible, it may be possible to develop a relevant end-to-end investigative process to detect an attack's existence and the determination of an attack's origin.

Significance of the Problem

The difficult nature of determining the source and existence of a cyber attack may result in insufficient options for an appropriate and timely response. The importance of a timely and accurate determination of an attack's origin is presented by the Center for Strategic and International Studies (Borchgrave, Cilluffo, Cardash, & Ledgerwood, 2001):

The anonymity of the Internet lends itself well to the covert and invisible launching of coordinated attacks, using the low-level noise mentioned previously as cover. Planning a three month window to address hostile activities that may last for only seconds or

minutes, and for which varying degrees of immediate response must be available both for deterrence of further attacks and retribution against existing ones, is not a viable policy by any standard (p. 48).

To be effective, appropriate attack response by the victims requires accurate and appropriately early determination of an attack's existence and its origin. Inaccurate and misleading evidence might result in unintended negative consequences.

To ensure an appropriately early and accurate response to a cyber attack requires a comprehensive end-to-end digital investigation supported by relevant tools. As indicated earlier, Stephenson (2002b) argues that a comprehensive investigative process consists of a number of steps beginning with collecting evidence. Collection of evidence in a "computer security incident is very time sensitive" (Stephenson, 2002b). Events that do not appear to have major significance may be of considerable importance when taken in the proper context. Therefore, evidence collection must not be judgmental until it is shown that these events resulted in negative consequences.

There are several phases in an end-to-end digital investigation process. During the analysis phase, duplicate events should be eliminated from the analysis. The nature of all incident events will be investigated without prejudice to any pre-constructed theories. It is at this point in the process that an understanding of how individual events come together from a macro level must be determined. This is done during the preliminary correlation phase when the examination of individual events and their interactions are analyzed. At this point in the process it can be observed that one event may be reported through multiple sources. During the event normalizing phase the investigator seeks to normalize (combine) events attributed to multiple

sources. If some events are reported multiple times by the same source, then event deconfliction is performed to eliminate the duplicated reporting of the same event by multiple sources.

Once the redundancy in event reporting is reduced, the next step in the process is to create an incident timeline with the resultant events. The timeline will help guide the construction of the digital chain of evidence. The chain of evidence is the flow of events and their artifacts as a function of time. “Ideally each link of the chain, supported by one or more pieces of evidence, will lead to the next link” (Stephenson, 2002b). Finally all of the events in the constructed chain of evidence will be corroborated with other independent evidence and events. Regardless of whether or not the incident is prosecuted in a court of law, the end-to-end digital investigation process provides the incident victim with a comprehensive and systemic means to ascertain an incident’s events. In the case of a cyber attack, the results of an end-to-end investigation may provide the foundational nature of the incident so that appropriate and timely retribution options may be considered. Stephenson has written two comprehensive articles explaining the end-to-end digital investigation process (Stephenson, 2002a, 2003b).

Excessive loss of time to respond to an attack due to an inability for the investigation to determine the existence of an attack may result in delayed response by the attack victim. Cascading router failures potentially magnify the damage inflicted by an Internet attack and further justifies the significance of early identification of the attack’s existence to enable proactive quarantining of the infected servers and limit the infection’s spread due to the residual consequences of a cyber attack. As a consequence of these abnormal traffic patterns there may be a rapid increase in traffic load on the surviving critical nodes (Ash & Newth, ; Crucitti, Latora, & Marchiori, 2004; Motter & Lai, 2002). It may be that the cascading service degradation observed during rapidly spreading infections, such as SQL-Slammer, primarily

occurs due to the malicious removal of critical routers leading to an unnatural re-routing of Internet traffic.

The evidence collection and analysis phase of a potential attack investigation may be severely restricted due to the extreme complexity and large quantity of network event information collected by a plethora of incompatible network monitoring tools. These difficulties are compounded by the inherent anonymous nature of the Internet and the many techniques that can be employed by the attacker to obfuscate identity such as IP spoofing (Daniels, 2002; Daniels & Spafford, 2000a; Tang & Daniels, 2005).

Investigative tools that assist in the appropriately early determination of a cyber attack's existence in an accurate manner will enhance the investigation by reducing the volume of relevant evidence collected during the evidence collection phase. As stated earlier, the collection of evidence in a "computer security incident is very time sensitive" (Stephenson, 2002b). Therefore, an investigative tool that decreases the time to collect evidence may serve to enhance the quality of the corroborated chain of evidence. This proposal will now present the nature of the environment in which a potential investigative tool to infer an attack's existence will need to operate.

Nature of the Problem

Scale-free network theory.

Complex networks can encompass biological, social or computer networks that exhibit similar characteristics such as uneven node distribution, degree distribution with heavy tail characteristics, a hierarchical and communal organization and a high degree of node clustering. The current literature indicates that many complex networks exhibit similar node connectivity and infrastructure characteristics (Barabasi & Albert, 2002).

Complex networks can be described by power laws. Power laws are expressed in the form: $y \propto x^a$; where x and y are the variables of interest, 'a' is constant and \propto indicates that the two variables have a proportional relationship. The power law polynomial relationship exhibits scale invariance, i.e. the scaling coefficient (a) is constant. For example, $y = x^2$; where x is raised to a constant power of 2 ($a = 2$). Another generally accepted property of power law relationships is that a plot of $\log y$ versus $\log x$ (log-log plot) will result in a linear fit within acceptable error limits. The slope of the resulting line gives a constant 'a'. For scale-free networks the distribution of link degree over all nodes is a power law degree distribution. Therefore, the probability of any node (new and existing nodes) in the network having k number of links is proportional to k links raised to the power of a constant scaling coefficient (γ). The notation is as follows: $P(k) \propto k^{-\gamma}$. It has been empirically determined that the scaling coefficient is between 2 and 3 for most "real world" networks (Barabasi & Albert, 2002). The power law degree distribution states that a new node is more likely to connect to an existing node with a higher degree. Link establishment and the distribution of links between nodes is a probabilistic non-linear distribution as can be observed in figures 1 and 2.

The uneven distribution of link density among nodes in a scale-free computer network exhibit power law degree distribution and is based on the link degree of the node. This leads to heterogeneous nodes. For purposes of the research proposed in this paper we have classified the nodes into 2 categories, critical (important) and non-critical (non-important) nodes. Critical nodes will be considered nodes with a large number of links and all other nodes will be considered non-critical. The criteria for determination of the critical number of links is yet to be determined but the literature supports the distinction between critical and non-critical nodes as relevant (Albert, Jeong, & Barabasi, 2000; Crucitti, Latora, Marchiori, & Rapisarda, 2004; Sun,

Liu, Chen, & Yuan, 2007). Empirical studies have shown that that scale-free computer networks (such as the Internet) are extremely robust yet are vulnerable to attack (Albert, Jeong, & Barabasi, 2000; Crucitti, Latora, Marchiori, & Rapisarda, 2004; Sun, Liu, Chen, & Yuan, 2007). These findings indicate that important (critical) nodes are vulnerable points on the network due to their large volume of links and network traffic. Figure 1 presents live Internet data that supports the existence of power law degree distribution. The data was extracted from Internet activity during April, 1998 for 3,530 nodes and 6,432 links (Michalis, Petros, & Christos, 1999). As depicted in the figure, we see that there are a large number of nodes with only a few links and a very small number of nodes with many links.

The connectivity characteristics of scale-free networks are confirmed through empirical studies of the Internet (Barabasi & Albert, 1999; Guillaume, Latapy, & Magnien, 2005; Jeong, Neda, & Barabasi, 2003; Michalis et al., 1999; Saffre, Jovanovic, Hoile, & Nicolas, 2004), as well as the World Wide Web (Albert, Jeong, & Barabasi, 1999; Ravi et al., 2000), metabolic pathways (Jeong, Tombor, Albert, Oltvai, & Barabasi, 2000), and many social networks (Aiello, Chung, & Lu, 2000; Albert & Barabasi, 2000; Barabasi & Albert, 1999; Redner, 1998).

Preferential attachment.

The research proposed in this proposal will investigate scale-free computer network preferential attachment behaviors as a function of the Internet's router message path relationships. While there are many computer network connectivity models presented in the literature (Barabasi, Ravasz, & Vicsek, 2001), the current prevailing model of computer network

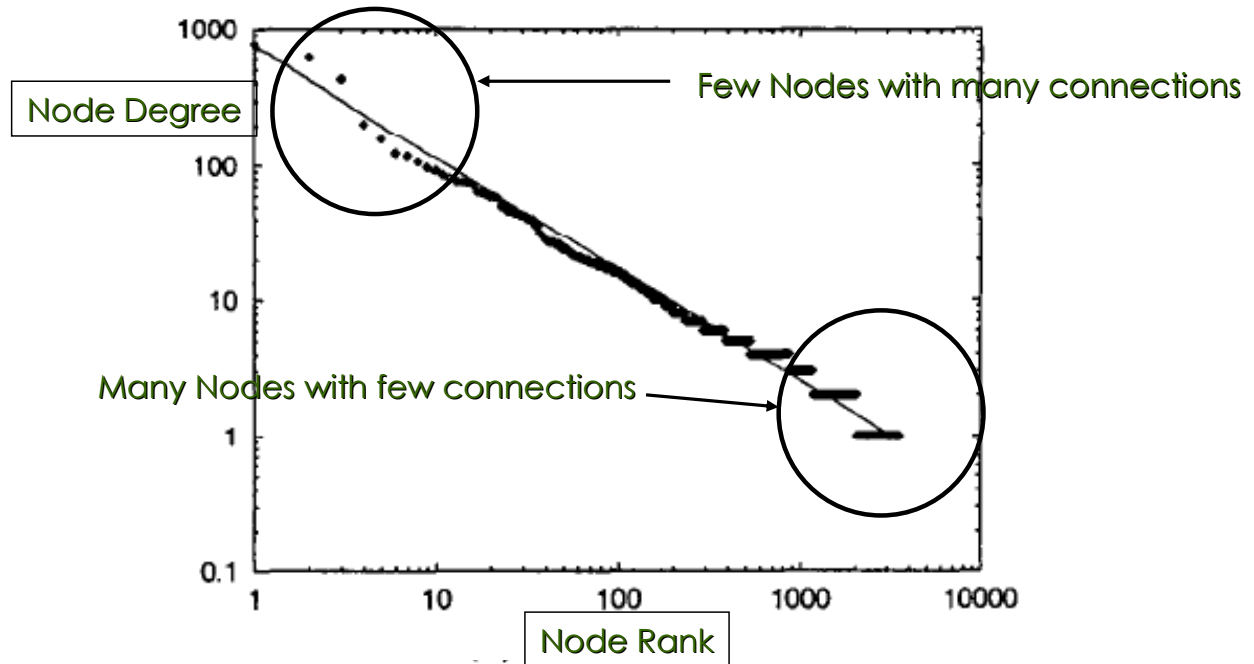


Figure 1. Live Internet data, power law degree distribution (Michalis et al., 1999).

connectivity predominant in the literature is the “Theory of Evolving Networks” (Barabasi & Albert, 1999, 2002; Barabasi, Albert, & Jeong, 2000). They postulate that scale-free computer network connectivity between nodes (specifically, routers) is not random; rather it follows a probabilistic linking behavior known as preferential attachment. They also posit that computer network connectivity is scale-free and new nodes are incrementally added to the network as required to complete a message path. Preferential attachment dictates that the probability that an existing node establishes new links with other nodes and is influenced by its number of existing links. The linking behavior follows the power law distribution and is not relational to network size (scale-free) (Barabasi & Albert, 2002; Jeong et al., 2003; Michalis et al., 1999; Yook, Jeong, & Barabasi, 2002).

Specifically, scale-free computer networks, such as the Internet, have been shown through empirical study to behave by preferential attachment rules (Barabasi et al., 2001; Saffre

et al., 2004). Figure 2 depicts evidence for the existence of preferential attachment in live Internet data extracted in 2000. The sample taken consisted of 12,400 nodes with 13,445 links (Jeong et al., 2003). As depicted in the figure, we can see that there is a greater likelihood of communication link establishment for nodes with the greatest number of existing links.

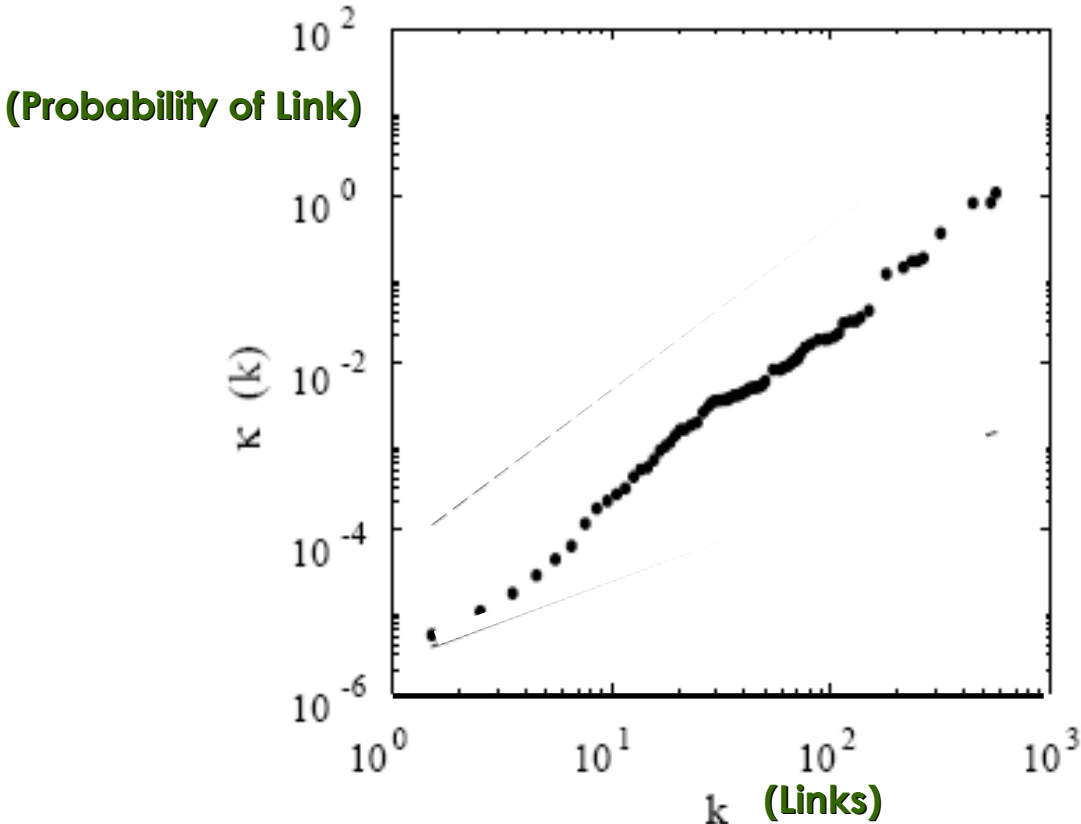


Figure 2. Live Internet data, preferential attachment (Jeong et al., 2003).

Cyber attack markers and preferential attachment theory.

The literature also supports the proposition that the Internet's connectivity topology exhibits fractal (self-similar) properties relative to geographic population centers (Caldarelli, Marchetti, & Pietronero, 2000; Chakraborty et al., 2004; Lakhina, Byers, Crovella, & Matta, 2002; Yook et al., 2002). It may be plausible that the scale-free connectivity behaviors observed

over the Internet are in some manner a manifestation of the Internet's fractal connectivity topology (Stephenson, 2006). One novel approach that may provide a foundational premise for anomaly detection over the Internet is known as the "Theory of Cyber Attack Mechanics" (Stephenson & Prueitt, 2005). Using a unique set of foundational concepts and formalisms they theorize that it may be possible to identify a cyber attack's origin by observing traffic disruptions in the Internet's fractal connectivity. As related to network communication connectivity and very large computer networks (such as the Internet), Stephenson and Prueitt (2005) hypothesize that:

The interaction between cyber attack space and fractal network space results in event markers that may anticipate the existence of a cyber attack. Because attack markers are complex, they may result in halting conditions within the network. These halting conditions can be represented formally and the source of the cyber attack may be deduced (p. 9).

These halting conditions may be observable through disruption in the Internet's traffic mechanisms as identifiable events (attack markers) (Stephenson & Prueitt, 2005). The "Theory of Cyber Attack Mechanics" leads Stephenson to postulate that the halting conditions brought on by the aforementioned disruptions in the Internet's fractal connectivity may be a result of an underlying violation of the Internet's preferential attachment linking rules and these violations may possibly be observable as event markers (attack markers) (Stephenson, 2006). There are empirical studies present in the literature that may support their hypothesis, suggesting that disruptions in normal Internet traffic patterns can be observed (Liljenstam, Yuan, Premore, & Nicol, 2002; Yegneswaran, Barford, & Ullrich, 2003).

The model and simulation proposed in this proposal may lead to an improved understanding of scale-free computer network behaviors (preferential attachment and halting conditions) so that future researchers may be able to develop the appropriate investigative tool for the confirmation of a cyber attack and the possible determination of its origin.

Research Contributions

The resulting model and simulation of this dissertation will be referred to as the Attack-Flow model to distinguish it from other models mentioned in this dissertation proposal. We seek to enhance the body of knowledge for cyber attack modeling through the following contributions:

- Provide evidence that the Attack-Flow is a plausible and scientifically sound premise for the future development of investigative tools to infer cyber attacks.
- Enhance the scope of applications for Colored Petri Net (CPN) modeling and simulation of concurrent and complex network communications.
- Provide a novel approach for the modeling and simulations of cyber attacks using Colored Petri Nets.
- Offer a unique approach for simulating cyber attacks based on attack interactions over scale-free computer networks through violation of simulated preferential attachment mechanisms.
- Develop the preliminary feasibility for future experimental research and formal proof of the “Theory of Cyber Attack Mechanics”.
- Identify potential anomalistic Internet traffic event markers as indicated by our CPN simulation that may potentially enhance current Intrusion detection systems abilities to proactively detect unusual behavior over the Internet.

Research Objective and Question

This research is primarily a feasibility study that seeks to provide the first step for determining a cyber attack's origin through determination of an attack's existence. We postulate that a cyber attack can be simulated through violations of the Internet's preferential attachment rules. It may be possible to infer the existence of the attack through observations of changes in its connectivity state of our simulated connectivity model. If we can infer the attack's existence for a specific network connectivity state, we will then attempt to assign an appropriate probability that the model correctly inferred a cyber attack's existence.

If it is a valid assumption that disruptions of the Internet's fractal scale-free connectivity are observable through cyber attack markers and these disruptions can be simulated through violations of the network's preferential attachment mechanisms, we ask: Using Colored Petri Nets to model and simulate the Internet's preferential attachment mechanisms, can we effectively represent the relevant cyber attack interactions? Specifically, will the modeling and simulation experiments of this research dissertation provide a feasible and relevant technique to potentially infer a cyber attack's existence and provide a useful first step towards the development of investigative tools to determine a cyber attack's origin?

Research Limitations

The limitations of this proposed research are:

1. This research will not introduce any attack motivation theories or discussion.
2. We will limit our simulations to scale-free computer network connectivity mechanisms as related to preferential attachment and network traffic flow characteristic patterns.
3. We will simulate the cyber attack interactions in a virtual environment, there will be no live internet experiments performed in this research.

4. Actual tools and prototypes based on our model will be considered future work.
5. We will limit our simulations and experiments to one specific scale-free network, the Internet.
6. Message path simulations, attack mechanisms and the resulting attack inferences will be made from the Internet's router infrastructure perspective only, the organizational and enterprise perspective will not be considered.

Methodology

A robust, reusable, automated model will be developed that simulates complex scale-free computer network communication connectivity as represented by the set of intermediate node-pairs in a complete message path. We seek to simulate the consequences of a malicious attack on message path route relationships. Simulations of a baseline network connectivity state profile (absent attack) using a pre-determined set of ordered node-pairs representing an entire message path will be performed. We surmise that each node-pair from the initial source node, through the intermediate nodes and to its ultimate destination were formed as a result of preferential attachment mechanisms. We will then simulate an attack on the preferential attachment link connectivity between the node-pairs through critical node removal (by destroying the affected node-pair relationships). We hypothesize that the node degree distribution may be affected by the attack (critical node removal and residual affects) and this change will be observable. To test the sensitivity of our simulation, we may then simulate node removal without cascading affects. Finally, we may then also look at partial node removal through removal of a nodes links as a potential method with acceptable inference sensitivity. Our basic approach is very similar to classic experimental sensitivity analysis techniques. (Kleijnen, 1992, 2005; Saltelli, Ratto, Tarantola, & Campolongo, 2006; Saltelli, Tarantola, Campolongo, & Ratto, 2004).

To accomplish this design, the methods of the study must provide a source for the ordered node-pairs (baseline samples); describe a general scheme for simulation, select the modeling and simulation language; develop data collection, process monitoring and Attack-Flow modules and propose at least an initial simulation description. This section now addresses each of these requirements.

Baseline Samples of Ordered Node-pairs

Traceroute is a well-known computer network research tool that traces network packets from their original source, through their intermediate paths and to then to its ultimate destination. It has been found to be an appropriate technique for collecting network path data for the study of the Internet's topology characteristics such as node degree distribution (Leguay, Latapy, Friedman, & Salamatian, 2007; Mahadevan et al., 2006; Mahadevan et al., 2005). The Cooperative Association for Internet Data Analysis (CAIDA) is a well-known network research institute that coordinates the Active Measurement Project (AMP). From 1998 to 2006, AMP collected large volumes of specific site-to-site traceroute path data between 150 servers around the world (Cooperative Association for Internet Data Analysis, CAIDA - Active Measurement Project, n.d.).

We propose using real internet traceroute path data as observed through the Active Measurement Project (AMP) of CAIDA as the foundational data to develop the baseline network connectivity model proposed for this research. A complete message path, from its initial source to its ultimate destination, is made up of multiple hops between intermediate router pairs. As depicted in figure 3, we surmise that the complete path may be represented by the sum of each of its individual intermediate router-to-router ordered pairs (ordered because the path is one-way). We will attempt to simulate the connectivity characteristics of each complete source-destination path through representation of the individual intermediate ordered router pair in the complete path. These ordered pairs may reflect router relationships and their preferential attachment mechanisms behaviors.

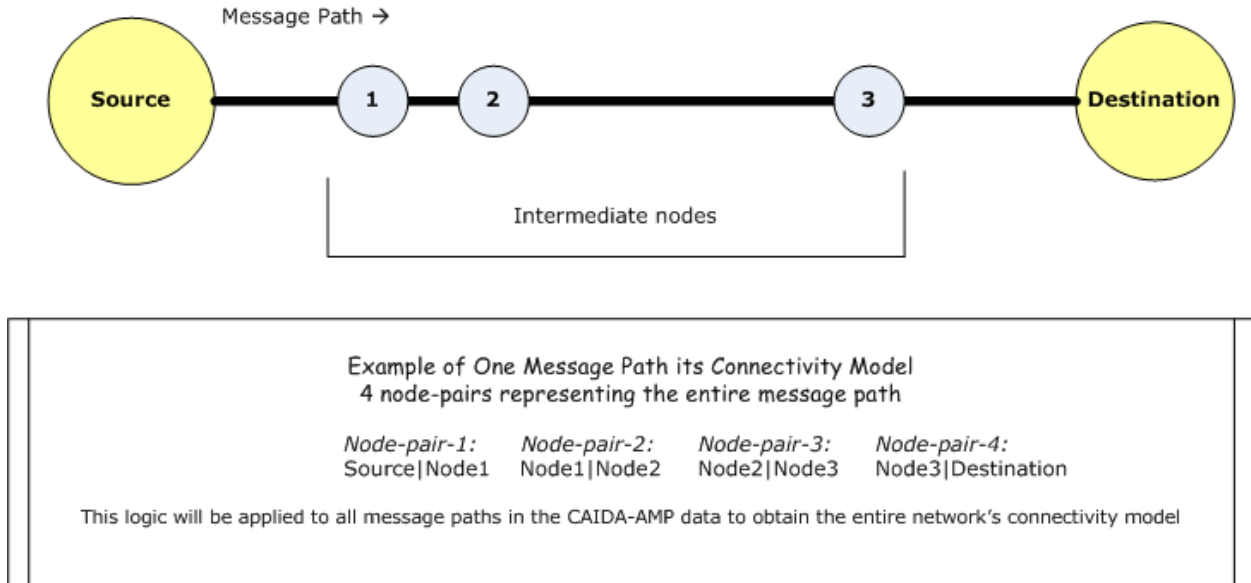


Figure 3. Node-pairs and network connectivity model.

Figure 4 presents our approach for the creation of the node-pairs which will represent the baseline network connectivity model. As shown in figure 9, we propose to split all traceroute paths in our data into a set of ordered node-pairs based on their adjacencies. This will essentially divide each path into a set of node-pairs that represent adjacent nodes. The total path will be represented as the sum of all its individual node-pair adjacencies. We believe that the formation of these intermediate node-pairs during entire message path creation process reflects router relationships. These relationships are of interest in our study of cyber attacks and scale-free computer networks. We may be able to infer an attack's existence through observation of anomalies in these relationships initiated by the attack. We may utilize queuing data structure in the CPN simulation to store the set of ordered node-pairs representing the baseline connectivity relationships. Each ordered node-pair will be released into the simulation through the queue utilizing a First-in, First-out (FIFO) queue algorithm.

Initial data extracts indicate that there will be approximately 430,000 node pairs.

Therefore, for purposes of the simulation in this research proposal, we anticipate the need to

reduce this data into significant router (node) clusters. As discussed earlier in this proposal, one central notion of preferential attachment theory is that there are a small number of nodes with a large number of links; that is scale-free computer network nodes tend to cluster by link degree. Therefore, we intend to utilize the notion of clustering to reduce the universe of node-pairs to simulate to a few significant and critical node clusters.

Link analysis (Stephenson & Prueitt, 2005) is a powerful analytical tool that can graphically present relationships between elements in a given universe. These graphical representations provide assistance in the reduction of large datasets into manageable clusters for further analysis. We intend to utilize link analysis to reduce the number of node-pairs for the CPN simulation to a small number of the most significant node clusters based on their node degree. Previously in this research proposal we discussed the fractal nature of scale-free computer network connectivity. We hypothesize that the proposed simulation of an attack's disruption of connectivity affects over a small cluster of significant nodes will be relevant due the self-similar (fractal) connectivity patterns. We then surmise that the simulation observations may be appropriate to infer an attack over the larger Internet.

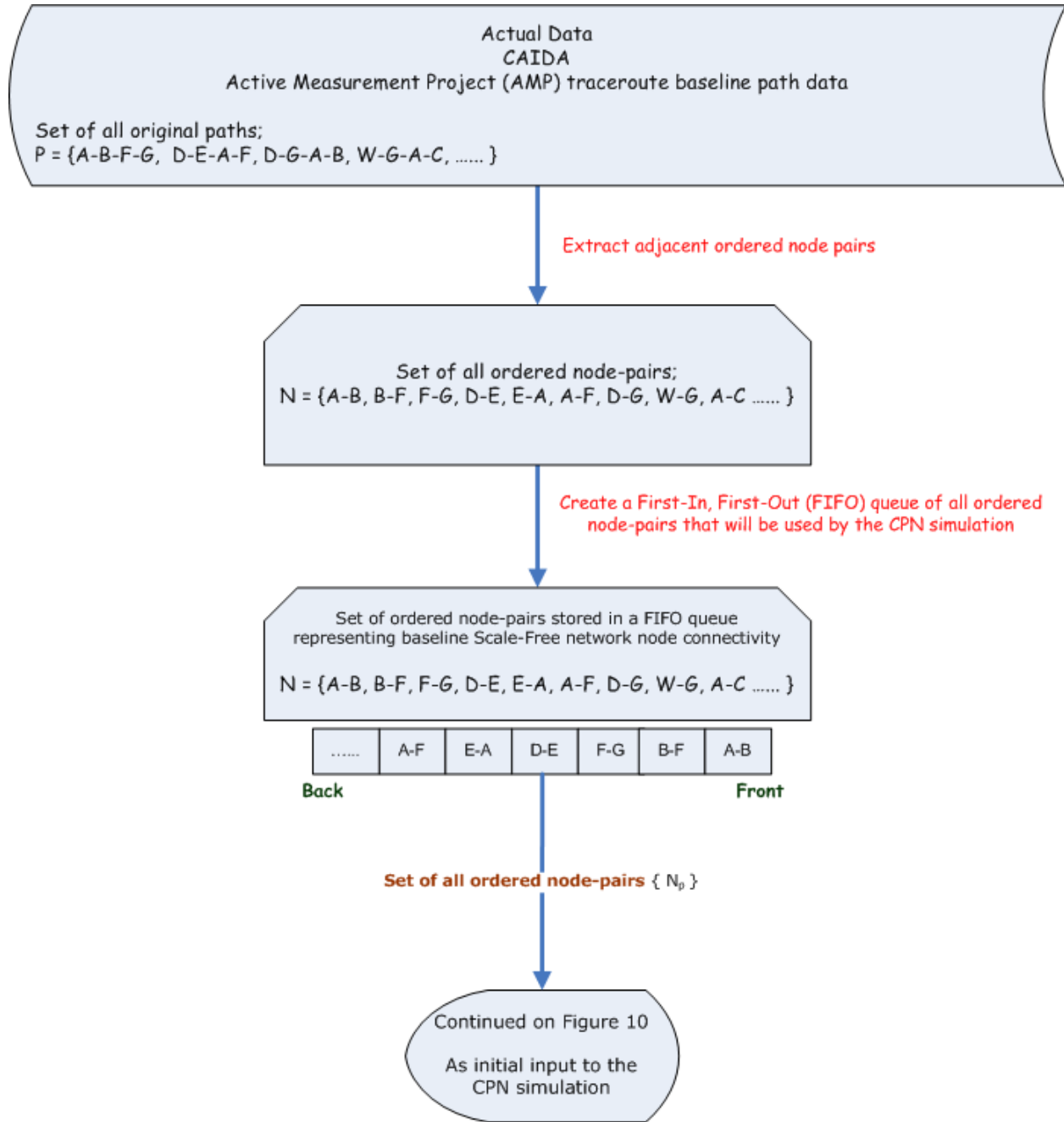


Figure 4. Attack-Flow, create network connectivity baseline.

General Scheme for the Simulation

Figure 5 depicts the general scheme of the research design used in our simulations. The ordered set of node-pairs extracted from actual CAIDA traceroute path data will represent the baseline network connectivity initial state. Attack scenarios will be simulated through the

removal of specific sets of nodes. The algorithm to select the critical nodes and the actual attack scenarios has not yet been determined. For each run of the simulation, an attack scenario represented by a set of removed nodes will be applied against the baseline connectivity set of ordered node-pairs. Next, a node-pair will be evaluated. Either the node-pair will not be affected by the attack (ie neither node in the pair is a member of the removed node set) or if one of the nodes in the pair is a node selected for removal (ie is a member of the removed node set) then the pair will be destroyed. Upon removal of a node from the node-pair, a temporary orphaned node (the node in the node-pair that was not in the removed set of nodes) is created. If possible, a new pair will form with the orphaned node through preferential attachment probabilities. If it is not possible to form a new node-pair then this orphaned node will be permanently orphaned and added to the set of removed nodes. During the formation of new node-pairs, nodes may exceed their link capacity, when this occurs, the node will also be added to the set of removed nodes. The dynamic nature of the set of removed nodes through the addition of orphaned nodes and nodes with exceeded link capacities simulates a cascaded node removal simulation (cascaded attack).

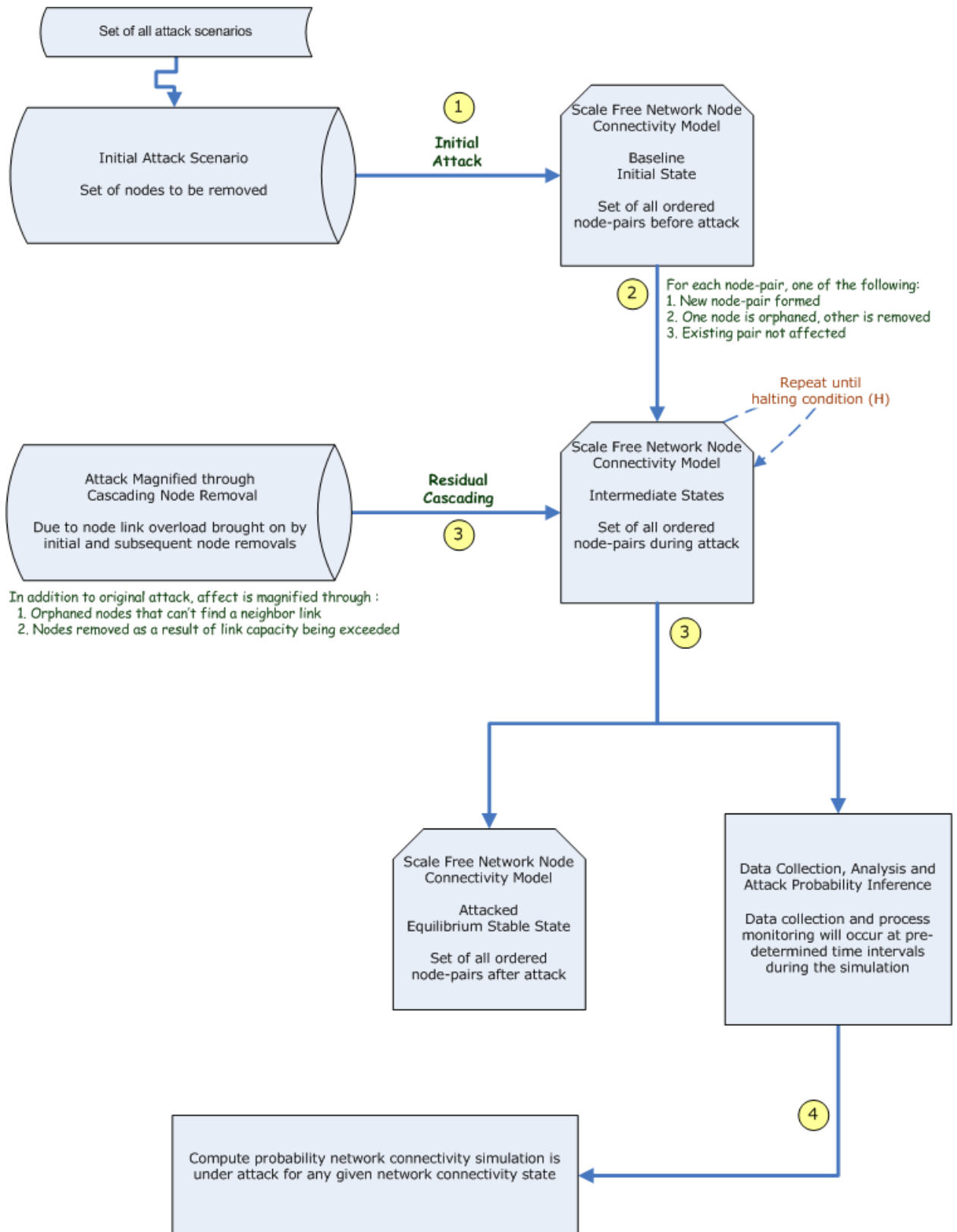


Figure 5. Process overview.

Node pairs are newly formed or destroyed in a parallel process that will be explained in detail later in this proposal. The set of ordered node-pairs is dynamic since as the simulation proceeds, node-pairs are formed and destroyed. The dynamic set of ordered node-pairs represents a series of intermediate network connectivity states that are changing during the attack simulation. These intermediate connectivity states occur as a result of dynamic attacked node removal and subsequent formation of new node-pairs. This simulates the affect of an attack on node relationships. While the simulation is running, we will collect global network connectivity data at specific pre-determined time intervals and network connectivity states. This data will then be analyzed by the attack inference algorithms.

The following premises are assumed for the simulation proposed in this paper and are based on earlier discussion of theory of scale-free networks under targeted infrastructure attack, specifically attempts to remove or degrade infrastructure connectivity:

Premise 1: As discussed earlier in this research proposal, preferential attachment theory, network connectivity relationship changes, attack identification through critical node removal and its surmised cascading affects are relevant hypothesis for early attack inference and these affects may be simulated.

Premise 2: Ordered node-pairs represent network connectivity relationships

Premise 3: The initial attack scenario (a pre-determined set of removed critical nodes) serves as a primer for subsequent network connectivity relationship residual affects observed during the simulation.

Premise 4: Cascading attacks are simulated through node-pair destruction and potential permanently orphaned nodes on the ordered node-pairs resulting in a dynamic set of removed nodes that are unavailable for new node-pair formation.

Premise 5: As the cascading affect proceeds, the network connectivity relationships be altered and these intermediate states will be observable.

Premise 6: At some point during the simulation a halting condition will occur and is defined as an intermediate state with one of the following conditions:

- a. The network connectivity relationships (ie. intermediate set of ordered node-pairs) will achieve an equilibrium state stability where no new node-pairs are formed or destroyed.
- b. Sometime during the simulation an intermediate state will reach a breaking point (critical mass) where the network will become permanently unstable, either dead-lock or live-lock occurs.

Premise 7: A stable network is one without changes to the set of ordered node-pairs yet the node degree distribution of the final state does not violate preferential attachment theory.

Premise 8: An unstable network is one without changes to the set of ordered node-pairs yet the node degree distribution of the final state which violates preferential attachment theory.

Premise 9: The final state (stable or unstable) of the network connectivity relationships significantly different than the original baseline and through inferences based on node degree distribution metrics may be anticipated.

Finally, we state our hypothesis based on these 9 premises as follows: The change in network connectivity relationships may violate preferential attachment theory and may be significant enough to observe early in the attack and therefore permit an avenue for attack inference during the implementation of an attack and the subsequent intermediate states.

This research will first look at complete node removal and the cascading affect through the simulation of a catastrophic attack on the infrastructure. As discussed earlier in this research

proposal, early attack inference is essential. It may be that the simulation proposed above will result in intermediate states during the simulation that present attack evidence sufficiently early enough evidence for an attack inference (before significant node removal cascading has occurred). While it may be optimal to infer an attack before the cascading affect has commenced and the network has become unstable (catastrophic attack), it may not be possible using our hypothesis. We seek to determine that “critical mass” of an attack, the point at which we can infer an attack’s existence (within a determined confidence level) as early as possible and with minimal network instability. To determine that “critical mass” the attack simulation affects will also be observed at varying degrees of cascaded node removal. We may also simulate the attack complete (removal of a few selected critical nodes) without cascading so as to determine if we can observe more subtle changes in the network before cascading occurs and therefore provide an earlier inference of attack at a more sensitive technique.

Another possible approach that we may endeavor is simulation of partial node removal through link removal, that is the node is still functioning but at a reduced level of service due to the removal of some of its links to neighbor nodes. By observing partial node service degradation as exhibited through link (but not the entire node) removal we may be able to improve the simulation’s attack inference sensitivity. This may provide an earlier recognition of a potential attack before a catastrophic event is observable. We might define partial node removal as eliminating a pre-determined percentage of a links to a random selection of neighbor nodes.

In summary, our research design will focus on complete node removal and the residual cascading affects. We surmise that the basic premises of our CPN model and simulation should

be flexible enough to refinement such the variations on node removal strategy should be appropriate.

Selection of CPN Modeling and Simulation Language

The modeling and simulation language for this dissertation will be Colored Petri Nets (CPN). The relevance of CPN models is discussed below. Colored Petri Nets (CPN) are used to model and simulate a wide variety of industrial strength applications through virtual representations (Kristensen, Christensen, & Jensen, 1998). These include communication protocols, audio/visual systems, operating systems, hardware designs, embedded systems, software system designs and business process re-engineering.

The foundational building blocks of Colored Petri Nets programming syntax are places, tokens, arcs, colors, transitions, markings and guards. A place represents an environment (such as a network router) that is assigned markings (token values) to represent the system state (configuration) of the place. Tokens are computer bit strings (such as variable values) that are transmitted across arcs (communication lines) to other places (routers). Communication between 2 places is initiated by transitions so that a set of tokens may be passed between 2 places in the simulation. It is possible to transmit multiple tokens between 2 places sequentially or concurrently. Colors (programming declaration statements) and guards (conditional statements) provide the essential flow control of tokens as they are passed between places. Colored Petri Net development may occur in both a graphical (visual) and non-graphical (text only) traditional programming environment.

Figure 6 depicts a mapping of the foundational syntax and graphical elements of the modeling and simulation language. A simple representation of a CPN simulation of preferential attachment relevant to the research dissertation proposed in this proposal is presented in Figures

7 and 8. Figure 7 shows the CPN simulation state initially before transition 1 is enabled, and figure 8 presents the simulation state after transition 1 is enabled. A database will be created to store node information, such as the number of existing links for each node (linksP2 and linksP3). Node 1 represents the source node of the message and nodes 2 and 3 depict intermediate nodes between the source and destination. The message is simulated by passing the tokens (one token per message per path id) between the nodes (routers) until the destination is reached. As the token is passed between nodes during the simulation, relevant data will be collected for analysis. The token value is set to the path id to provide a convenient way to identify the particular simulation as it proceeds. Two well written explanations of this modeling and simulation language by the international authority on Colored Petri Nets (Jensen, 1997, 1998) are available at University of Aarhus' website.

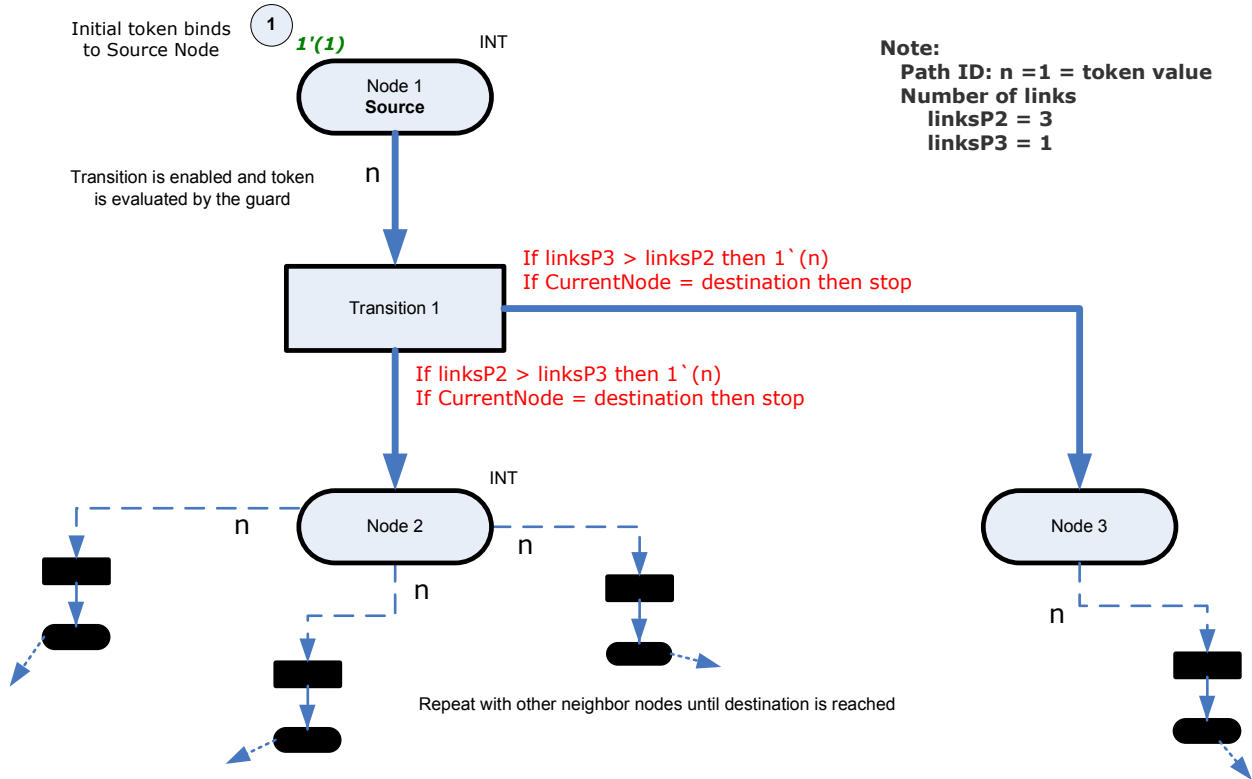


Figure 7. Colored Petri Net example, initial simulation state before transition 1 is enabled.

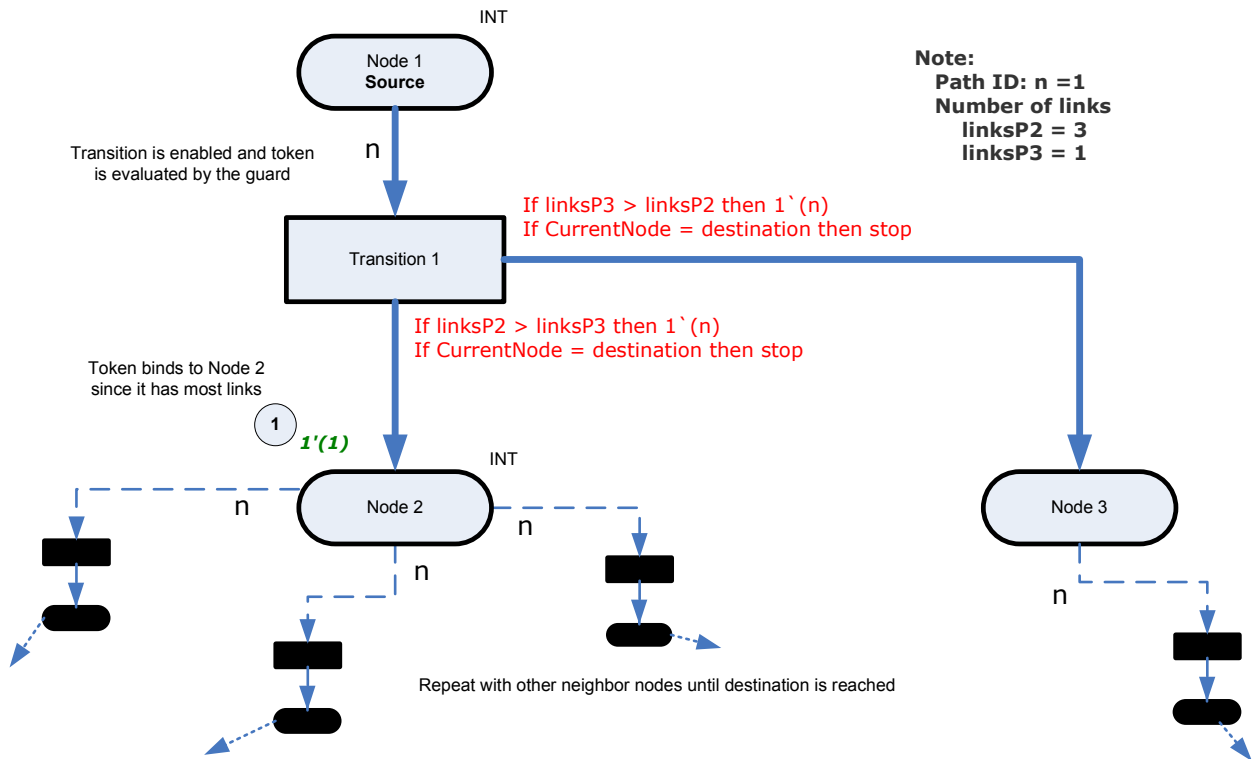


Figure 8. Colored Petri Net example, simulation state after transition 1 is enabled.

Kurt Jensen and his team at the University of Aarhus in Denmark are major contributors to the development and practical use of CPN. They define the essence of the language as follows: “Colored Petri Nets provide a framework for the construction and analysis of distributed and concurrent systems” (Kristensen et al., 1998). Cyber attack interactions and router message transmissions over a scale-free computer network are complex, concurrent and distributed. Therefore, CPN modeling is an ideally suited, mathematically proven virtual representation that may present these complex systems in a controllable and analytical context. Currently, there are many tools for design, development and simulation of Color Petri Nets (Petri Nets World, n.d.). Jensen’s team has developed an intuitive tool for CPN modeling and simulation (the CPN Tool) that is used by more than 700 research organizations in over 70 different countries (Coloured Petri Nets at the University of Aarhus, n.d.). Their CPN Tool will be the foundational instrument used for model development and simulation proposed in this proposal.

Advantages of CPN modeling and simulation as stated by the CPN group include: (1) an intuitive modeling language that allows for the flexibility and power of modern programming languages as well as graphical representations of the model; (2) well-defined semantics leading to unambiguous models of system behavior; (3) a flexible modeling environment that can be used in a wide variety of complex industrial-strength applications; (4) a modeling language consisting of a few powerful programming primitives; (5) a model that exhibits true concurrency, not interleaving; (6) timed and probabilistic simulation functionalities; and (7) a universally accepted formal analysis and verification of any derived model (Coloured Petri Nets at the University of Aarhus, n.d.).

Development of CPN Modules

Each robust CPN module developed for this research will be designed for a specific Attack-Flow function and will be compliant with generally accepted software engineering principles (Pressman, 2004). Figure 9 through 14 depict a generalized representation of the proposed methodology and the Colored Petri Net implementation scheme. Figure 9 shows an explanatory key for the graphical elements presented in figures 10 through 14. The CPN simulation begins with a baseline network connectivity model. Figure 4 depicts the creation of the network connectivity baseline and it will be discussed in the data sampling section of this proposal.

Data will be collected during pre-determined time intervals during the simulation. After the node-pairs and cascade affect has stabilized, the simulation will be halted. We will then repeat the attack implementation with a different attack scenario. The data utilized for the link capacity and preferential attachment computations will be stored in a relational database and node-pair comparisons will be implemented using CPN tokens. The database is updated in real time as the simulation proceeds to add and remove nodes and links. The proposed data structures are presented in table 1. CPN tokens will be utilized to store node identification numbers and multiple tokens (node id) will be stored in a token array.

Data structure	Description
Database (node = key)	<p>Current linking capacity</p> <p>Based on literature computation (P. Crucitti, Latora, & Marchiori, 2004; Motter & Lai, 2002), constant during simulation, represents the maximum number of links for a node, computed once ($t = 0$).</p> <p>Current number of links</p> <p>Initialized at $t = 0$, updated with the addition of each new link.</p> <p>List of neighbor nodes</p> <p>Initialized at $t = 0$, Dynamic, changes as a neighbor nodes (Distance = 1 hop) are removed.</p>
CPN tokens	<p>Set of node-pairs</p> <p>Array of current set of node-pairs, Initialized at $t = 0$, updated as node pairs are destroyed and created</p> <p>Individual and orphaned nodes</p> <p>When nodes are removed or orphaned, updated throughout the simulation individual nodes will be passed as tokens.</p> <p>Set of removed nodes</p> <p>Arrays of current set of removed nodes, original attack (nodes removed) initialized at $t = 0$, updated as node pairs are destroyed or nodes are overloaded.</p>

Table 1. Proposed data structures.

Initial CPN Attack Simulation

The general flow of the attack implementation is shown in Figure 10. The set of critical nodes will be based on the node degree. Based on earlier discussion of cyber attack concepts, nodes with the highest degree will most likely be considered critical nodes. The number of links that a node should have to be classified as a critical node has not yet been determined. Attack scenarios will be created by extracting subsets of critical nodes from the set of all critical nodes. Each subset of initial attacks will represent the initial attack scenario (set of removed nodes). The initial attack scenario nodes will be applied to the connectivity baseline as the node-pairs are released for analysis one at a time from the baseline model. If one of the node-pair elements is a member of the set of removed nodes (the initial attack scenario) then that node-pair will be destroyed and re-formed through preferential attachment. These new node-pairs will be placed on the back of the node-pair queue to be reevaluated later in the simulation against the dynamic set of removed nodes. Newly established links may exceed a threshold for one of the nodes in the new node-pair, if this occurs then the node is considered to be overloaded resulting in a cascaded node removal affect and it will be added to the set of removed nodes for dynamic evaluation against upstream node-pairs. The baseline network connectivity model (set of ordered node-pairs) will be altered as a result of the attack.

Figure 11 represents a more detailed view of the initial attack implementation simulation. Node-pairs are released from the stack in a classic first-in first-out algorithm. At each clock tick of the simulation, a new node-pair will be released from the stack for processing. Each node-pair is evaluated against the current set of removed nodes. Initially the set of removed nodes is the original attack scenario represented as a pre-selected set of critical nodes. As the node-pair is processed, if either node of current node-pair being evaluated is found in the list of removed

nodes then the other node (the node not in removed list) will be orphaned. Subsequently, this newly orphaned node or nodes (if both are orphaned) must locate a new linking partner (presented in figure 13). . If neither node of the current node-pair is a member of the set of removed nodes then the process continues to figure 12. New link formation (with a new node partner as shown in figure 13) will be based on the theories of the Barabasi-Albert model (Barabasi et al., 2001) of preferential attachment. The linking algorithm will implemented in a probabilistic manner such that the orphaned node's neighbor with the greatest link degree will be have a proportionately greater likelihood of being selected as the new linking partner

The surmised cascading affect of a cyber attack through node removal by exceeding a baseline link capacity is a relevant means for simulation of cascaded attacks over complex computer networks (Ash & Newth, ; Crucitti, Latora, & Marchiori, 2004; Motter & Lai, 2002). This research will utilize the criteria as discussed in these papers. As mentioned above, if neither node of the current node-pair is a member of the set of removed nodes then each node is evaluated for potential link overload (exceeds its link capacity). In figure 12, the process will determine whether the link capacity for either node is exceed by comparing the each node's link capacity against its current number of links (as stored in the database). If the link capacity is exceeded for either node then the node that has exceeded its link capacity will be added to the dynamic set of removed nodes. Subsequently, the newly overloaded node will be considered in all upstream node-pair evaluations.

The node orphaned (due to the overloaded node shown in figure 12) as a result of the overload will be processed as presented in figure 13. If the link capacity is not exceeded for either node then the node-pair is pushed to the back of the node-pair queue and will be re-evaluated later in the simulation through the FIFO algorithm upstream as re-released (figure 11).

As mentioned above, the newly orphaned node (as a result of the overloaded node in figure 12) will proceed as shown in Figure 13 (link discovery process for orphaned nodes). The dynamic nature of link formation and destruction in the simulation is captured and processed in pseudo-real time that is at each clock tick. Therefore information such as the current node degree and available neighbors will be dynamic as the simulation proceeds. At this step in the simulation (figure 13), all neighbor nodes of the orphan are evaluated and current preferential attachment probabilities for its neighbors are re-computed. If there are available neighbor nodes for the orphan then in a probabilistic manner the new link will be formed with the selected neighbor creating a new node-pair. This node-pair will then be pushed on the back of the node-pair queue and will be re-evaluated through the FIFO algorithm upstream as re-released (figure 11). If the orphaned node has no available neighbor nodes, it is permanently orphaned and will be added to the dynamic set of removed nodes.

Once a halting condition occurs such as the node-pairs stabilizing and forming equilibrium (no new node-pairs) and the cascading affect has halted (no more link capacities exceeded) then the attack scenario has ended.

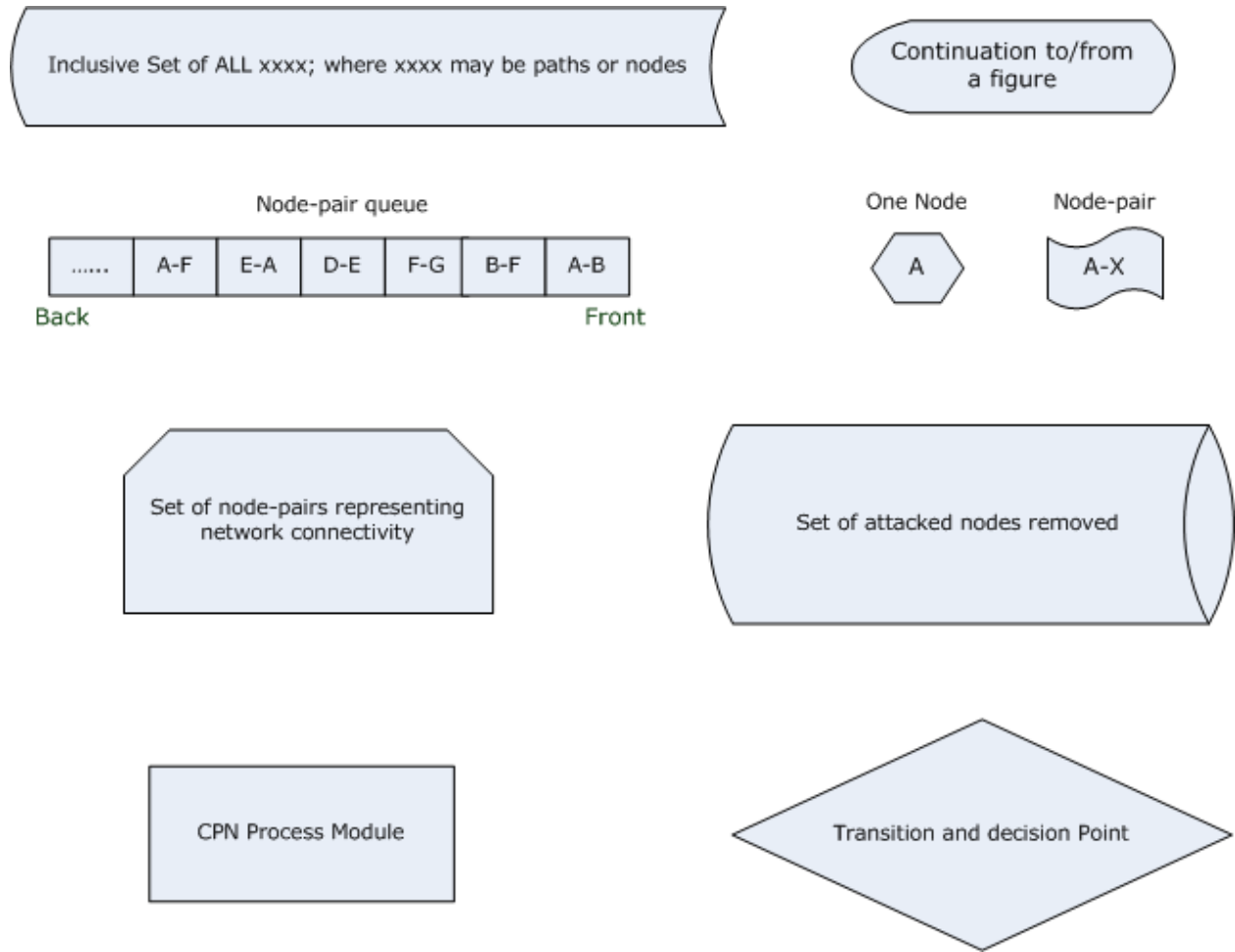


Figure 9. Attack-Flow, overview graphics key.

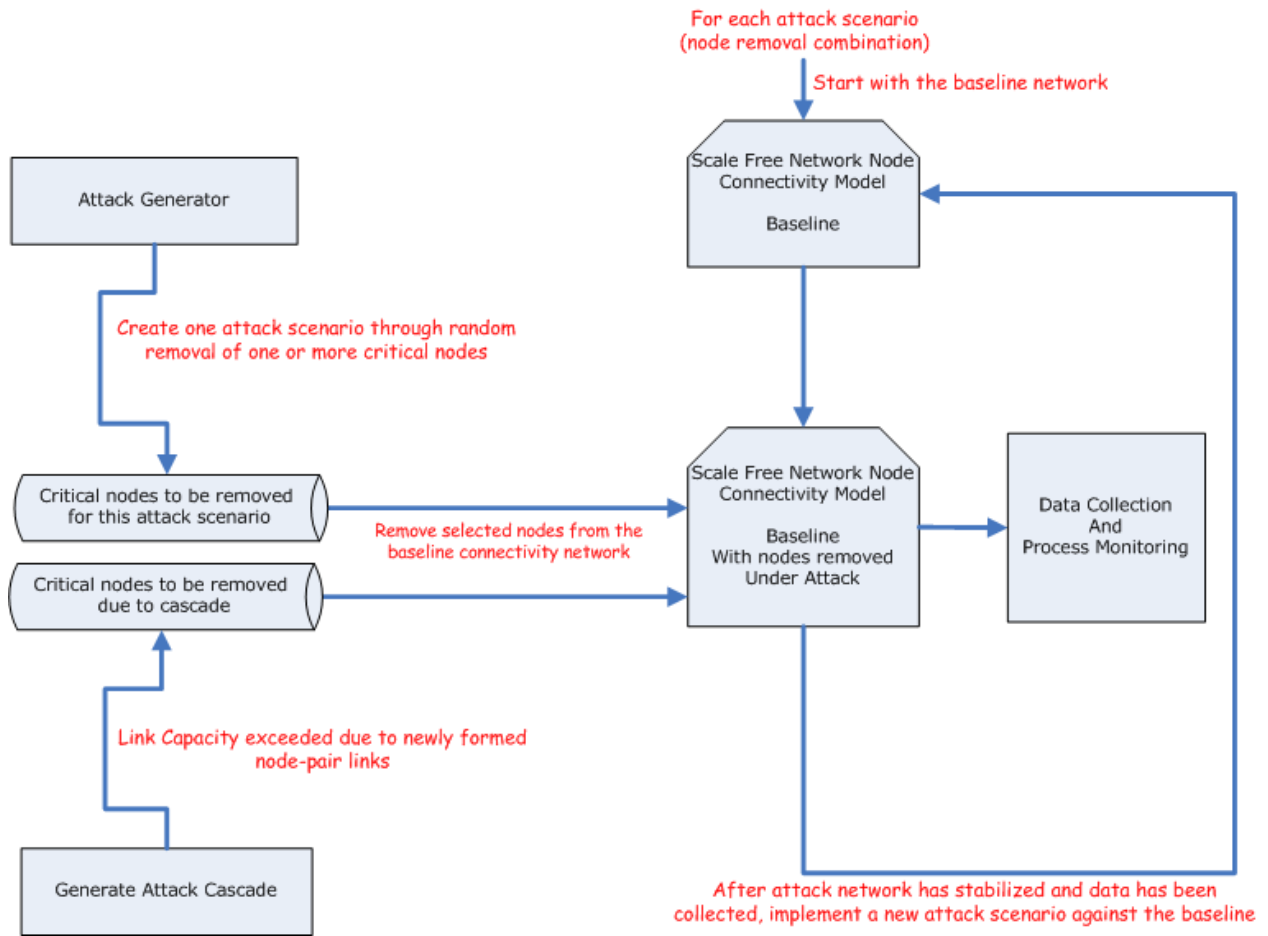
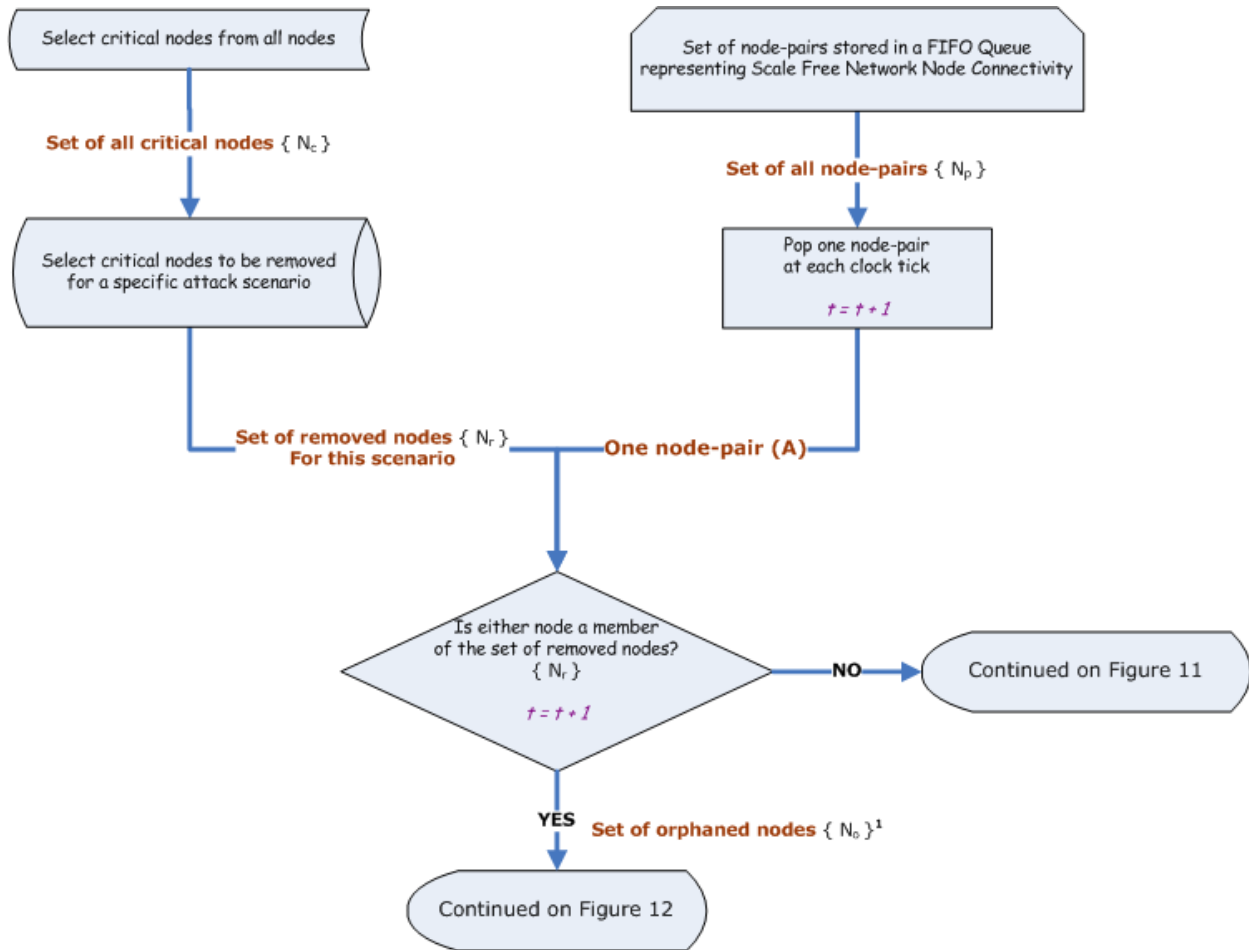
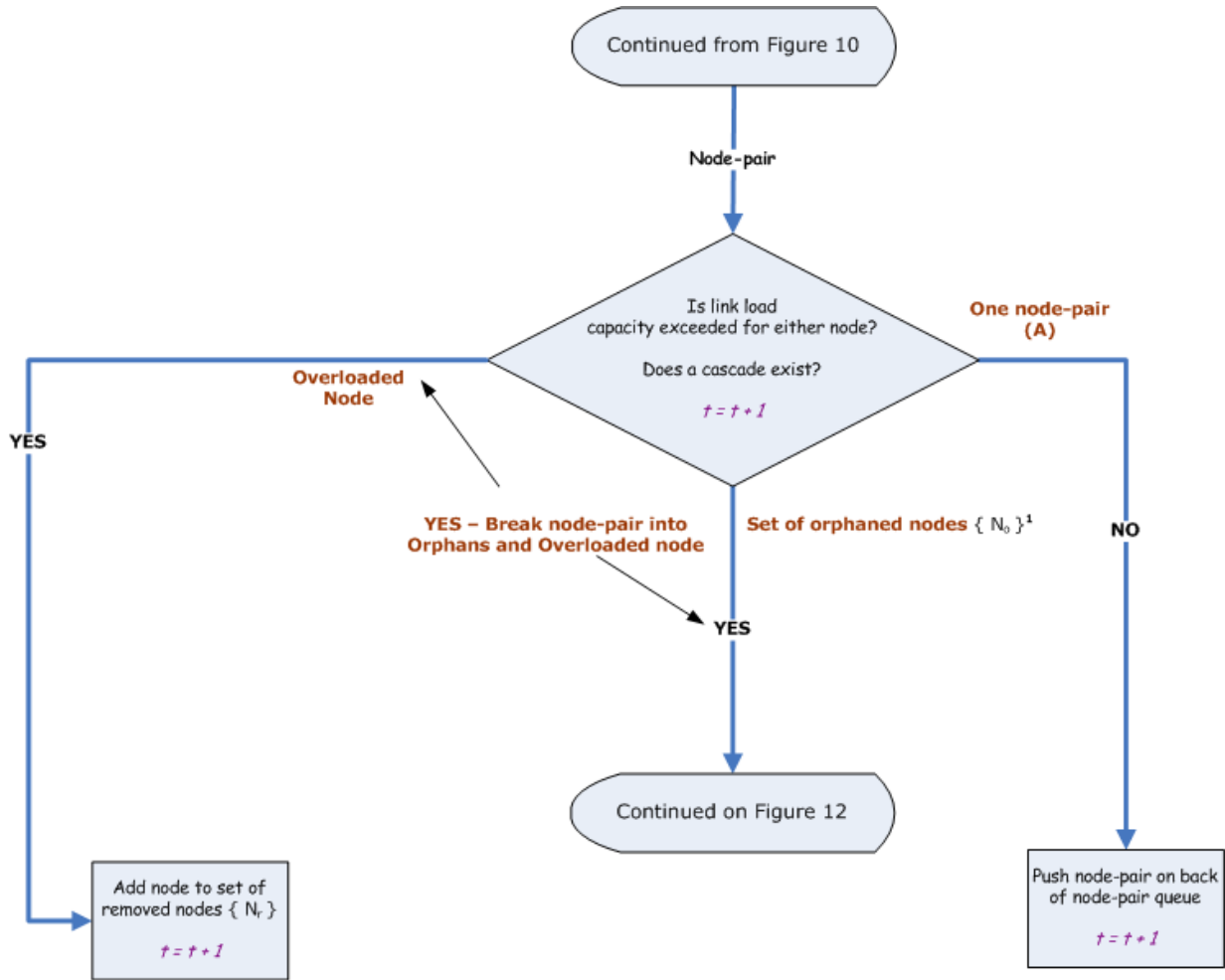


Figure 10. Attack-Flow, attack implementation – overview.



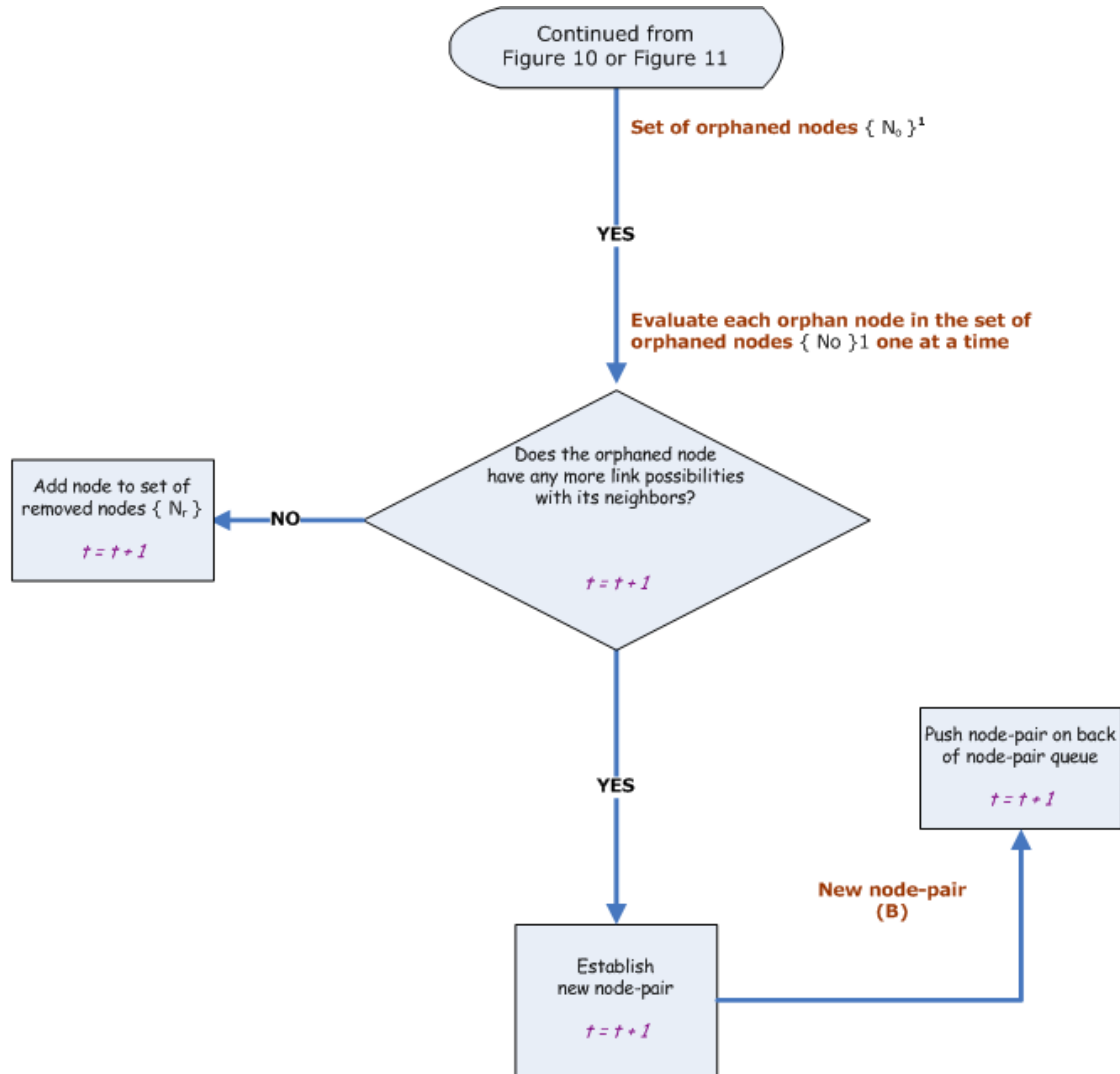
¹ $\{ N_o \} = \{ N_c \} + x_1$ where:
 $\{ N_c \}$ = set of all nodes orphaned due to overloaded node removal
 x_1 = orphaned node due to node-pair split

Figure 11. Attack-Flow, attack implementation – detail.



¹ $\{ N_o \} = \{ N_c \} + x_1$ where:
 $\{ N_c \}$ = set of all nodes orphaned due to overloaded node removal
 x_1 = orphaned node due to node-pair split

Figure 12. Attack-Flow, attack reaction – cascade affect.



¹ $\{ N_o \} = \{ N_c \} + x_1$ where:
 $\{ N_c \}$ = set of all nodes orphaned due to overloaded node removal
 x_1 = orphaned node due to node-pair split

Figure 13. Attack-Flow, attack reaction – potential link discovery.

Figure 14 presents a useful example for purposes of clarification. The following assumptions are in place for this example:

1. Node-B is a member of the set of removed nodes.
2. Node-X is an available neighbor node for node-A and is selected to form a new link with node-A.
3. Node-X becomes overloaded after forming new link with node-A.

As shown in figure 14, node-pair A-B is released from the node-pair queue, the node pair is evaluated and since node-B is a member of the removed nodes set, node-pair A-B is destroyed. If node-B wasn't a member of the removed nodes set then the node-pair A-B would be placed on the back of the node-pair queue for further processing (re-released when its turn comes). Since node-A is now orphaned; it forms a new node-pair with an available neighbor node (node-X) in a probabilistic manner based on the node degree of node-A's current set of neighbors. If there were no available neighbor nodes then node-A would become a permanently orphaned node and it would be added to the dynamic set of removed nodes. The current number of links of both node-A and node-X will then be evaluated against their link capacities to ascertain whether either node has become overloaded at the current state of the dynamic degree distribution. In this case, node-X link becomes overloaded due to the addition of its new link with node-A. Therefore node-X is added to the set of removed nodes and node A once again becomes an orphaned node. As shown previously Node-A will then search for another neighbor to link with. If neither node-A or node-X were overloaded then node-pair A-X would be added to the back to the node-pair queue for further processing (re-released when its turn comes). This same process continues for each node-pair until a halting condition occurs. At that time, this instance of the simulation is complete. A new attack scenario (initial set of critical nodes removed) will be implemented

against the original baseline node-pair connectivity state. The number of attack scenarios and their composition implemented during the simulated attacks will be determined as we proceed with the research. The aggregate node degree data representing altered relationships in the baseline connectivity model will be collected from each attack scenario and analyzed in the Data Analysis and Attack Inference module.

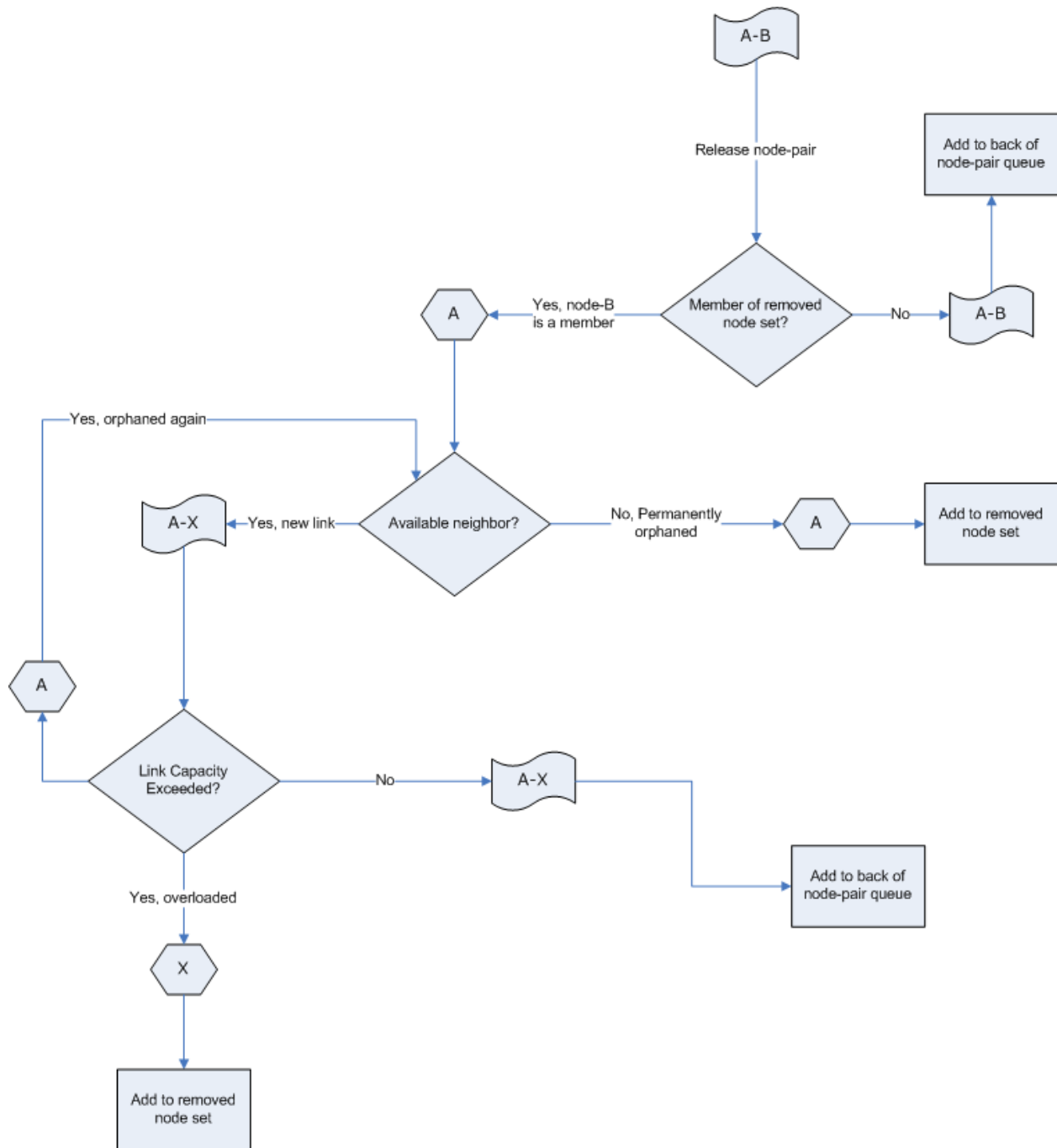


Figure 14. Attack-Flow, attack simulation example.

In summary, this simulation seeks to emulate cascaded node removal brought on by the original attack (original set of critical nodes selected for removal) through two mechanisms, when the node is removed due to link capacity overload and when a node becomes permanently

orphaned. Dynamic formation of new node-pairs and the destruction of the baseline node-pairs may represent the changing dynamics of a scale-free network under attack. Observation of the changing node degree distribution characteristics as represented by the evolving node-pair composition may present a quantifiable manner to infer an attack's existence. Global network connectivity metrics will be collected during pre-determined time intervals. The dynamic nature of link formation and destruction in the simulation is captured and processed in pseudo-real time that is at each clock tick. Therefore information such as the current node degree and available neighbors will be dynamic as the simulation proceeds. In the next section of this research proposal we will provide a general overview of our proposed data sampling methodology.

Data Collection and Analysis

The network connectivity state (node degree distribution) properties of the node-pairs before and after the simulated attack will be collected by the Data Collection and Process Monitoring modules. At pre-determined time intervals in the simulation we will collect data related to the content of the node-pair FIFO queue and the set of removed nodes, in addition to the collection of the relevant simulation state and node degree distribution data. The Process Monitoring module will ensure process integrity during all simulations through validation at pre-described breakpoints. Wells (2002) presented a useful and relevant CPN data collection and process monitoring implementation that may be extended for our purposes.

The observed connectivity state changes may provide adequate indication that an attack has commenced. The data collected from the baseline and attack profiles will then be funneled through an inference engine to determine the probability of an attack's existence under the various connectivity states. The data collected for initial state (baseline), intermediate states and the final stable state of the network connectivity model as represented by the set of ordered node-

pairs will be analyzed using the Data Analysis and Attack Probability module. Internet topology characteristics including average node degree, node degree distribution, joint degree distribution and node clustering proposed in the literature (Mahadevan et al., 2005) show promise for our attack inference engine. We plan to verify the attack results through a yet to be determined comparison with simulation results found in earlier empirical studies (Albert et al., 2000; Crucitti, Latora, Marchiori et al., 2004). The Data Analysis and Attack Probability module will accept data from the process monitoring and data collection modules.

The framework for the Data Analysis and Attack Probability module will be determined in the course of our research. Numerous approaches are cited in the literature related to attack models and attempts for early detection of anomalous behavior over large complex networks, such as the Internet (Cheetancheri et al., 2006; Jan & Markus, 2002; Liu, Zang, & Yu, 2005; Moore, Ellison, & Linger, 2001; Vanit-Anunchai & Billington, 2004). Bayesian Networks have been successful in probabilistic inference for complex systems (Vanit-Anunchai & Billington, 2004). The Bayesian Network-CPN approach (Vanit-Anunchai & Billington, 2004) appears to be promising in that it presents a potentially seamless and automated method to pass connectivity simulation data generated during our proposed simulations and infer the attack probability through a well established inference approach (Bayesian Networks).

Another technique that may be employed in our analysis to infer the existence of an attack from the network's connectivity state is binary logistic regression. Binary logistic regression is a multivariate regression technique designed specifically for inferring the probability that a binary outcome is correct (Hair, Black, Babin, Anderson, & Tatham, 2006a, 2006b). So for example, we might use binary logistic regression to infer the probability that an attack exists or it does not exist.

Timeline

As are depicted below, table 2 presents the proposed major milestones for this research.

Task	Targeted Timeline
Approval of research proposal	October, 2007
Develop detailed project plan	October, 2007
Develop functional specifications and module integration	October, 2007
Module development	October, 2007 to December, 2007
Module testing	January, 2008 to June, 2008
Module integration	July, 2008 to September, 2008
Module integration testing	October, 2008 to December, 2008
Data reduction	January, 2009
Run simulations and collect data	February, 2009 to April, 2009
Data analysis	May 2009
Write remaining chapters of dissertation	June, 2009 to August, 2009
Defend dissertation	September, 2009

Table 2. Proposed dissertation research timeline.

Conclusion

The previous discourse on the relevant proposed dissertation problem considers the currently inadequate options available to policy makers for legal, political, strategic and tactical attack retribution as well as the difficulties of determining a cyber attack's existence and origin. Subsequently, we endeavor to enhance the relevant body of knowledge through techniques that are consistent with the relevant approaches and theories such as: scale-free computer networks, preferential attachment, the "Theory of Cyber Attack Mechanics" and its novel approach for as a inference of cyber attacks using Internet traffic disruptions, and the relevant empirical observations related to attack modeling.

We have proposed a novel approach to modeling attacks over scale-free computer networks using Colored Petri Net modeling and simulation. We have proposed to represent a scale-free computer network's connectivity through simulation of the Internet's router connectivity and message transmission. Through a series of simulated attack scenarios over the Internet we will violate the simulation's preferential attachment rules through important node removal leading to message path re-routing in an anomalous manner. We will then observe the result of varying attack scenarios and their affects on a multitude of message path variants as represented through the traceroute ordered node-pair relationships. Subsequently, from these results we hope to be able to determine the probability that the network simulation is under attack within certain acceptable error parameters.

The relevance and important nature of determining the probability of the existence of an attack over the Internet has been demonstrated. Colored Petri Net attack models that are relevant to large complex networks are non-existent in the literature. In conclusion, the proposed research dissertation and the Attack-Flow model and simulation approach is relevant, significant

and presents a novel approach to the problems previously mentioned. Through my navigation of the relevant literature and my proposed academic contribution to the attack modeling body of knowledge, I believe that I have presented a strong case for the successful completion of this candidacy exam and the continuation of the proposed research in this proposal.

References

- Aiello, W., Chung, F., & Lu, L. (2000). A random graph model for massive graphs. *Proceedings of the Thirty-Second Annual ACM symposium on Theory of Computing*.
- Albert, R., & Barabasi, A. L. (2000). Topology of evolving networks: Local events and universality. *Physical Review Letters*, 85(24), 5234-5237.
- Albert, R., Jeong, H., & Barabasi, A. L. (1999). Diameter of the World-Wide Web. *Nature*, 401(6749), 130-131.
- Albert, R., Jeong, H., & Barabasi, A. L. (2000). Error and attack tolerance of complex networks. *Nature*, 406(6794), 378-382.
- Ash, J., & Newth, D. (2004). *Evolving cascading failure resistance in complex systems*. Paper presented at The 8th Asia Pacific Symposium on Intelligent and Evolutionary Systems, Cairns, Australia.
- Barabasi, A. L., & Albert, R. (1999). Emergence of scaling in random networks. *Science*, 286(5439), 509-512.
- Barabasi, A. L., & Albert, R. (2002). Statistical mechanics of complex networks. *Reviews of Modern Physics*, 74(1), 47.
- Barabasi, A. L., Albert, R., & Jeong, H. (2000). Scale-free characteristics of random networks: The topology of the World-Wide Web. *Physica A*, 281(1-4), 69-77.
- Barabasi, A. L., Ravasz, E., & Vicsek, T. (2001). Deterministic scale-free networks. *Physica A: Statistical Mechanics and its Applications*, 299(3-4), 559-564.
- Borchgrave, A., Cilluffo, F. J., Cardash, S. L., & Ledgerwood, M. M. (2001). *Cyber threats and information security meeting the 21st century challenge*. Washington, D.C.: Center for Strategic and International Studies.

- Caldarelli, G., Marchetti, R., & Pietronero, L. (2000). The fractal properties of Internet. *Europhysics letters*, 52, 386-992.
- Carrier, B. (2003). Defining digital forensic examination and analysis tools using abstraction layers [Electronic Version]. *International Journal of Digital Evidence*, 1 from <http://www.utica.edu/academic/institutes/ecii/ijde/articles.cfm>
- Casey, E. (2002). Error, Uncertainty, and Loss in Digital Evidence. *International Journal of Digital Evidence* [Electronic Version], 1. Retrieved February 10, 2006 from http://www.ijde.org/archives/docs/02_summer_art1.pdf
- Casey, E. (2004). Network traffic as a source of evidence: tool strengths, weaknesses, and future needs. *Digital Investigation*, 1(1), 28-43.
- Chakraborty, D., Ashir, A., Suganuma, T., Keeni, G. M., Roy, T. K., & Shiratori, N. (2004). Self-similar and fractal nature of Internet traffic. *International Journal of Network Management*, 14, 119-129.
- Cheetancheri, S. G., Agosta, J. M., Dash, D. H., Levitt, K. N., Rowe, J., & Schooler, E. M. (2006). *A distributed host-based worm detection system*. Paper presented at the Proceedings of the 2006 SIGCOMM workshop on Large-scale attack defense, Pisa, Italy.
- Cheung, S. (2006). Denial of service against the Domain Name System. *Ieee Security & Privacy*, 4(1), 40-45.
- Coloured Petri Nets at the University of Aarhus. (n.d.). Retrieved March, 2007, from <http://www.daimi.au.dk/CPnets/>
- Cooperative Association for Internet Data Analysis, CAIDA - Active Measurement Project. (n.d.). Retrieved October, 2007, from <http://amp.nlanr.net/>

- Crucitti, P., Latora, V., & Marchiori, M. (2004). Model for cascading failures in complex networks. *Physical Review E*, 69(4).
- Crucitti, P., Latora, V., Marchiori, M., & Rapisarda, A. (2003). Efficiency of scale-free networks: error and attack tolerance. *Physica A: Statistical Mechanics and its Applications*, 320, 622-642.
- Crucitti, P., Latora, V., Marchiori, M., & Rapisarda, A. (2004). Error and attack tolerance of complex networks. *Physica a-Statistical Mechanics and Its Applications*, 340(1-3), 388-394.
- Daniels, T. E. (2002). Reference models for the concealment and observation of origin identity in store-and-forward networks. *DAI*, 64(09B), 157.
- Daniels, T. E., & Spafford, E. H. (2000a). *Network traffic tracking systems: Folly in the large?* Paper presented at the Proceedings of the 2000 workshop on New security paradigms, Ballycotton, County Cork, Ireland.
- Daniels, T. E., & Spafford, E. H. (2000b). *Network traffic tracking systems: folly in the large?* Paper presented at the Proceedings of the 2000 workshop on New security paradigms.
- Erdos, P., & Renyi, A. (1960). On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Science*, 5, 17-61.
- Gordon, L. A., Loeb, M. P., Lucyshyn W., & Richardson, R. (2006). *2006 CSI/FBI Computer crime and security survey*.
- Guillaume, J. L., Latapy, M., & Magnien, C. (2005). Comparison of failures and attacks on random and scale-free networks. In *Principles of Distributed Systems* (Vol. 3544, pp. 186-196).

Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2006a). Factor analysis. In *Multivariate Data Analysis* (pp. 101-166). Upper Saddle River, New Jersey: Pearson Prentice Hall.

Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2006b). Multiple discriminant analysis and logistic regression. In *Multivariate Data Analysis* (pp. 269-382). Upper Saddle River, New Jersey: Pearson Prentice Hall.

Holme, P., Kim, B. J., Yoon, C. N., & Han, S. K. (2002). Attack vulnerability of complex networks. *Physical Review E*, 65(5).

Institute for Security Technology Studies (2002). Law enforcement tools and technologies for investigating cyber attacks: A national needs assessment [Electronic Version]. Retrieved October 22, 2004 from <http://www.ists.dartmouth.edu/TAG/lena.htm>

Institute for Security Technology Studies (2004a). Law enforcement tools and technologies for investigating cyber attacks: A national research and development agenda [Electronic Version]. Retrieved October 29, 2004 from <http://www.ists.dartmouth.edu/TAG/rand.htm>

Institute for Security Technology Studies (2004b). Law enforcement tools and technologies for investigating cyber attacks: Gap analysis report [Electronic Version]. Retrieved October 29, 2004 from http://www.ists.dartmouth.edu/TAG/gap_analysis.htm

International Corporation for Assigned Names and Numbers (2007). *Root server attack on 6 February 2007*. Retrieved October, 2007, from <http://www.icann.org/announcements/factsheet-dns-attack-08mar07.pdf>

- Jan, S., & Markus, S. (2002). *Collaborative attack modeling*. Paper presented at the Proceedings of the 2002 ACM symposium on Applied computing, Madrid, Spain.
- Jensen, K. (1997). *Coloured Petri Nets: Basic concepts, analysis methods and practical use* (Second ed. Vol. 1): Springer-Verlag.
- Jensen, K. (1998). An introduction to the practical use of Coloured Petri Nets. In W. R. a. G. Rozenberg (Ed.), *Lectures on Petri Nets II: Applications, Lecture Notes in Computer Science* (Vol. 1492, pp. 237-292): Springer-Verlag
- Jeong, H., Neda, Z., & Barabasi, A. L. (2003). Measuring preferential attachment in evolving networks. *Europhysics Letters*, *61*(4), 567-572.
- Jeong, H., Tombor, B., Albert, R., Oltvai, Z. N., & Barabasi, A. L. (2000). The large-scale organization of metabolic networks. *Nature*, *407*(6804), 651-654.
- Kleijnen, J. P. C. (2005). An overview of the design and analysis of simulation experiments for sensitivity analysis. *European Journal of Operational Research*, *164*(2), 287-300.
- Kleijnen, J. P. C. (1992). Sensitivity analysis of simulation experiments: regression analysis and statistical design. *Mathematics and Computers in Simulation*, *34*(3-4), 297-315.
- Kristensen, L. M., Christensen, S., & Jensen, K. (1998). The practitioner's guide to Coloured Petri Nets. *International Journal on Software Tools for Technology Transfer*, *2*, 98-132.
- Lakhina, A., Byers, J. W., Crovella, M., & Matta, I. (2002). *On the geographic location of internet resources*. Paper presented at the Proceedings of the 2nd ACM SIGCOMM Workshop on Internet measurement, Marseille, France.
- Latora, V., & Marchiori, M. (2004). How the science of complex networks can help developing strategies against terrorism. *Chaos, Solitons & Fractals*, *20*(1), 69-75.

- Leguay, J., Latapy, M., Friedman, T., & Salamatian, K. (2007). Describing and simulating internet routes. *Computer Networks*, 51(8), 2067.
- Lipson, H. F. (2002). *Tracking and tracing cyber-attacks: Technical challenges and global policy issues* (No. CMU/SEI-2002-SR-009): Carnegie Melon Software Engineering Institute, CERT Coordination Center.
- Liljenstam, M., Yuan, Y., Premore, B. J., & Nicol, D. (2002). *A mixed abstraction level simulation model of large-scale Internet worm infestations*. Paper presented at the 10th IEEE International Symposium on Modeling, Analysis, and Simulation of Computer and Telecommunications Systems (MASCOTS'02).
- Liu, P., Zang, W., & Yu, M. (2005). Incentive-based modeling and inference of attacker intent, objectives, and strategies. *ACM Trans. Inf. Syst. Secur.*, 8(1), 78-118.
- Mahadevan, P., Krioukov, D., Fomenkov, M., Dimitropoulos, X., claffy, k. c., & Vahdat, A. (2006). The Internet AS-level topology: Three data sources and one definitive metric. *SIGCOMM Comput. Commun. Rev.*, 36(1), 17-26.
- Mahadevan, P., Krioukov, D., Fomenkov, M., Huffaker, B., Dimitropoulos, X., claffy, k., et al. (2005). *Lessons from three views of the Internet topology: technical report* (No. tr-2005-02). San Diego Supercomputer Center, University of California, San Diego: Cooperative Association for Internet Data Analysis - CAIDA.
- Martins, O. A. (2005). *Affecting IP traceback with recent Internet topology maps*. Iowa State University, Ames, Iowa.
- Michalis, F., Petros, F., & Christos, F. (1999). *On power-law relationships of the Internet topology*. Paper presented at the Proceedings of the conference on Applications,

- technologies, architectures, and protocols for computer communication, Cambridge, Massachusetts, United States.
- Moore, A. P., Ellison, R. J., & Linger, R. C. (2001). *Attack modeling for information security and survivability*. Pittsburgh, PA: Carnegie Mellon University Software Engineering Institute.
- Moore, D., Shannon, C., Brown, D. J., Voelker, G. M., & Savage, S. (2006). Inferring Internet denial-of-service activity. *ACM Trans. Comput. Syst.*, 24(2), 115-139.
- Motter, A. E., & Lai, Y. C. (2002). Cascade-based attacks on complex networks. *Physical Review E*, 66(6).
- Peng, T., Leckie, C., & Ramamohanarao, K. (2007). Survey of network-based defense mechanisms countering the DoS and DDoS problems. *ACM Comput. Surv.*, 39(1), 3.
- Petri Nets World - Tools. (n. d.). Retrieved March, 2007, from <http://www.informatik.uni-hamburg.de/TGI/PetriNets/tools/>
- Pressman, R. S. (2004). *Software Engineering: A practitioner's approach* (6th ed.): McGraw-Hill Science/Engineering/Math.
- Rattray, G. J. (2001). The Cyber Threat. In (pp. 79-119). US Air Force Academy: USAF Institute for National Security Studies
- Ravi, K., Prabhakar, R., Sridhar, R., Sivakumar, D., Andrew, T., & Eli, U. (2000). *The Web as a graph*. Paper presented at the Proceedings of the nineteenth ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems, Dallas, Texas, United States.
- Redner, S. (1998). How popular is your paper? An empirical study of the citation distribution. *European Physical Journal B*, 4(2), 131-134.

- Saffre, F., Jovanovic, H., Hoile, C., & Nicolas, S. (2004). Scale-free topology for pervasive networks. *BT Technology Journal*, 22(3), 200-208.
- Salla, V. (2005). Error and attack tolerance of complex real networks. *MAI*, 44(04), 90.
- Saltelli, A., Ratto, M., Tarantola, S., & Campolongo, F. (2006). Sensitivity analysis practices: Strategies for model-based inference. *Reliability Engineering & System Safety*, 91(10-11), 1109-1125.
- Saltelli, A., Tarantola, S., Campolongo, F., & Ratto, M. (2004). *Sensitivity analysis in practice. A guide to assessing scientific models*. New York: Wiley.
- Stephenson, P. (2002a). Analysis and correlation. *Computer Fraud & Security*, 2002(12), 16-18.
- Stephenson, P. (2002b). The forensic investigation steps. *Computer Fraud & Security*, 2002(10), 17-19.
- Stephenson, P. (2003a). Modeling of post-Incident root cause analysis. *International Journal of Digital Evidence*, 2(2).
- Stephenson, P. (2003b). Normalization and deconfliction. *Computer Fraud & Security*, 2003(1), 17-19.
- Stephenson, P. (2006). Towards improving attribution confidence in cyber attacks. *Journal of Cyber Conflict Studies*, 1(1), 48-54.
- Stephenson, P. R., & Prueitt, P. S. (2005). *Towards a theory of cyber attack mechanics*. Paper presented at the IFIP wg 11.9 Digital Forensics, First Annual Conference.
- Sun, S., Liu, Z. X., Chen, Z. Q., & Yuan, Z. Z. (2007). Error and attack tolerance of evolving networks with local preferential attachment. *Physica a-Statistical Mechanics and Its Applications*, 373, 851-860.

- Tang, Y., & Daniels, T. E. (2005). *A simple framework for distributed forensics*. Paper presented at the Proceedings of the Second International Workshop on Security in Distributed Computing Systems (SDCS) (ICDCSW'05) - Volume 02.
- Vanit-Anunchai, S., & Billington, J. (2004). *Modelling probabilistic inference using Coloured Petri Nets and Factor Graphs*. Paper presented at the Proc. Fifth Workshop and Tutorial on Practical Use of Coloured Petri Nets and the CPN Tools, University of Aarhus, Denmark.
- Wang, C.-H., Yu, C.-W., Liang, C.-K., Yu, K.-M., Ouyang, W., Hsu, C.-H., et al. (2006). *Tracers placement for IP traceback against DDOS attacks*. Paper presented at the Proceeding of the 2006 international conference on Communications and mobile computing, Vancouver, British Columbia, Canada.
- Wells, L. (2002). *Performance analysis using Coloured Petri Nets*. Paper presented at the Proceedings of the 10th IEEE International Symposium on Modeling, Analysis, and Simulation of Computer and Telecommunications Systems (MASCOTS'02).
- Yegneswaran, V., Barford, P., & Ullrich, J. (2003). *Internet intrusions: Global characteristics and prevalence*. Paper presented at the Proceedings of the 2003 ACM SIGMETRICS international conference on Measurement and modeling of computer systems, San Diego, CA, USA.
- Yook, S.-H., Jeong, H., & Barabasi, A.-L. (2002). Modeling the Internet's large-scale topology. *Proceedings of the National Academy of Sciences of the United States of America*, 99(21), 13382-13386.