On nth Order Attacks

Daniel BILAR^{a,1}

^a Department of Computer Science, University of New Orleans, USA

Abstract. An nth order attack seeks to degrade, disable or subvert an end system indirectly by targeting one or more end mission-sustaining ancillary systems. We discuss the vulnerability etiology enabling such attacks. We illustrate the notion of these attacks with concrete historical, current and forward-looking examples; also in the context of cyberwar against advanced computerized societies. We sketch the challenges and requirements to detect and mitigate the effects of nth order attacks.

Keywords. nth order attack, Highly Optimized Tolerance, ancillary system, assumption violation, economic warfare, critical infrastructure

1. Introduction

The goal of nth order cyber-warfare is to induce instabilities in mission-sustaining ancillary systems that ultimately degrade, disable or subvert an end system. Such systems may be technical/algorithmic; however, societal, psychological, ideological, economic, biological and natural systems may be targets, as well. Ancillary systems include pars pro toto memory resource allocation, throughput control, hardware/software manufacturing, visualization environments, social welfare systems, human networks, power generation/transmission/distribution, voting systems, data and goods supply lines, reputation management, entropy externalization, business models and economic systems.

For example, a denial of service attack against a web server can be seen as a case of a 2nd order attack against the resource allocation subsystem of the TCP transport subsystem. Thompson's trojaned compiler in "Reflection on Trusting Trust" may be seen as a 3rd order attack against software manufacturing tools [1].

This paper defines and discusses this class of attacks and tries to explain their etiology via reference to Highly Optimized Tolerance (HOT) processes. HOT processes induce structured systems through optimization mechanisms that incorporate tradeoffs between objective functions and resource constraints in probabilistic environments. Pertinent to our discussion is the property that such optimization-generated systems are *robust towards common perturbations, but especially fragile towards rare events*, such as unanticipated changes in the environment. Inducing such 'rare events' in mission-sustaining ancillary systems is thus the goal of nth order attacks.

The rest of this paper is organized as follows: Sec. 2 explains the main concepts that motivate our subsequent discussion. Sec. 3 reviews related work. We give concrete examples of n^{th} order attack in Sec. 4. Sec. 5 discusses analytical aspects of n^{th} order

¹Corresponding Author: Daniel Bilar, Department of Computer Science, University of New Orleans, 2000 Lakeshore Drive, New Orleans LA 70148, USA; Email: daniel@cs.uno.edu

attacks. Sec. 6 briefly sketches theoretical and practical remediation approaches. Sec. 7 gives final thoughts on the urgency of addressing the theme of the paper.

2. Overview

The following section serves to flesh out the nomenclature and concepts used throughout the paper. We shall start with the abstract notion of a 'system'; the definition of which varies across time and domains. For the purposes of this discussion, we adopt a recursive variant of biologist von Bertalanffy's seminal work on General Systems Theory [2]:

A system is a whole that functions by virtue of the interaction between constitutive components. As such, it is defined by these relationships. Components may be other systems.

For our purposes, the attractiveness of the definition lies in its emphasis on *openness* and the allowance for *structural similarities* across different domains with concomitant correspondence of governing behavior. For an short, readable, largely non-technical overview of competing system theories, the reader is referred to [3, ch. 2].

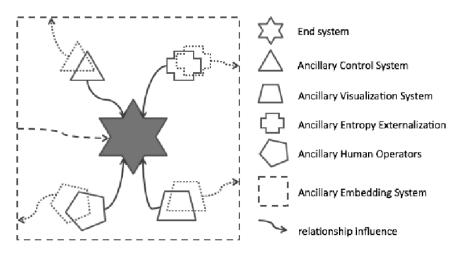


Figure 1. System view end system

Ancillary systems are responsible for control mechanisms, fault detection/resilience/recovery, energy/data flow, economic viability, human usability, data processing/structures, graceful startup/shutdown, reputation management, governance, social order and more. Such systems may be technical/scientific/algorithmic; however, societal, psychological, ideological, economic, biological and natural systems are included, as well.

Ancillary systems span different scales and varying orders of complexity. They may be embedded in or encompass the end system, and may in turn be composed of and influenced by other ancillary systems. Figs. 2(a) and 2(b) list an embedding (say a business model) and embedded ancillary system (in this example human operators) with reference to an end system (denoted by the center star) from Fig. 1.

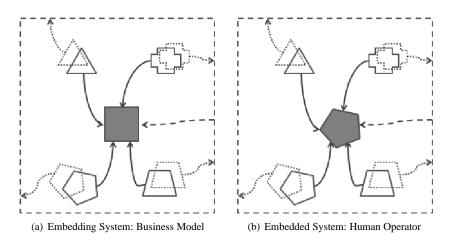


Figure 2. Examples components of embedding system (a) embedded system(b) of an end system

A Network Intrusion Detection System (NIDS) may serve as an illustrative end system example. Its ancillary control system negotiates the data and instruction interplay between sensors, analysis, database and decision/response engine. The ancillary visualization system displays the events and possible remediation options. The human operator subsystem must interpret the happenings and subsequently make the decisions that are not automated to the best of its reasoning ability. The entropy externalization subsystem is (among other things) responsible for cleaning out accumulation of dynamic data through sliding windows/logging and filtering out sensor noise. The end system itself is embedded in a business model that governs aspects of its design, implementation and activity: profit model, signature update cycles, customer support and more.

The ancillary systems of the NIDS end system have themselves subsystems: Human operators (Fig. 2(b)) field a visual subsystem subject to parameters (no UV sight, certain percentage of color-blindness, angular resolution etc). Their control system may be thought of as their reasoning strength and limitations (cognitive dissonance, herd instinct, unconscious intelligence [4] etc), as well as their physiological mechanisms (hormone secretions of the hypothalamus that regulate sleep, hunger, temperature etc). The human subsystem of human operators may be coworkers, friends, the fellow polity, family. Entropy externalization systems manifest themselves in physical (as in human waste product expulsion), as well as mental and psychological mechanisms (stress relief through exercise, keeping a diary, art, talking on the phone etc).

The business model (Fig. 2(a)) is embedded itself in an economic environment, say a free market economy, which influences its setup (tax codes, corporate structure, sales channels, liquidity parameters such as interest rates which determine acceptable debt ratios etc). The control subsystem may consists of corporate governance, union influence, mission statement, and legislative regulations. Its visualization subsystem may include accounting publication systems (standardized formats like IFRS with its own assumptions), dress codes, as well as marketing approaches (corporate image, advertisements etc). Human operator system may be stockholders, consultants writing the business plan, company workers, product consumers, company management, and competitors. Finally, the entropy externalization ancillary system of the business model may include mechanisms to off-set losses to subsidiaries, third-tier rebranding of products for steep sales discount, 'poison pills' to counter hostile takeovers, corporate fusion plans, and more.

2.1. nth order attacks

An nth order attack tries to indirectly degrade, disable or subvert an end system by targeting one or more ancillary systems.

With this qualitative definition in hand (which we shall pick up in Sec. 5), let us revisit the NIDS example in Fig. 1 with its control, human, entropy externalization, embedding and visualization ancillary systems. How would one go about perpetuating an nth order attack against an NIDS? One could take on the control system via a DoS attack against the response/decision engine, or try to supply fake/poisoned data to the analysis engine. Given biological, cognitive and psychological human parameters, enough false positives at 3am will make human operators tone down the sensitivity of the analysis component. One form of entropy electronic equipment produces is heat. The vast majority of Intel and AMD CPUs, for instance, reach critical heat at about 55-85°C[5, p. 5-13], which may cause the BIOS to shutdown to prevent damage: Hence, one attack against this entropy externalization system raises the ambient temperature of the building in which the NIDS components are deployed (say by low-tech clogging the climate intake vents). PNNL's Starlight [6] offers a comprehensive NIDS visualization system, replete with 2-D and 3-D multimedia visualizations supporting comparisons and emphasizing interrelationships. As can be intuited by the Starlight Network Intrusion Detection Graph² in Fig. 3, once data flow reaches a critical mass (by virtue of screen resolution and human limitations) visuals will degenerate into saturated pixel blobs, obviating their usefulness. The susceptibility of security visualization methods to intentional noise remains a serious concern, as described by [7].

Why do these attack work? Why does any attack, cyber- or otherwise, work? The answer we propose is surprisingly simple: *Attacks work because they violate assumptions*. Any finite system by design must incorporate implicit and explicit assumptions into its structure, functionality, and language. These systems are formulated with 'expected', 'typical' cases in mind and the assumptions reflect these expected use cases: A man-in-the-middle attack violates the assumption that you are talking to the party you expected; a race condition attack violates ordering assumptions; a buffer overflow attack violates an explicit resource assumption; BGP routing and DNS case poisoning attacks violate implicit trust assumption of non-malicious open architecture participants. Likewise, terroristic activities in open societies are easy to pull off because spaces are open, population freedom of movement not controlled - hence they violate implicit societal trust assumptions. Lastly, many democratic voting schemes assume 'honest' voters, and hence can be undermined by strategic voting [8]. There are scores of examples, in every domain.

We shall revisit the trust assumption in open societies in Sec. 7. Our next goal, however, is to gain some intuition about the etiology of the problem: We present a putative generative mechanism which crucially depends on assumptions to highlight the consequences of violating said assumptions.

²In the interest of fairness, it should be noted that this image is originally in color, not gray shades.

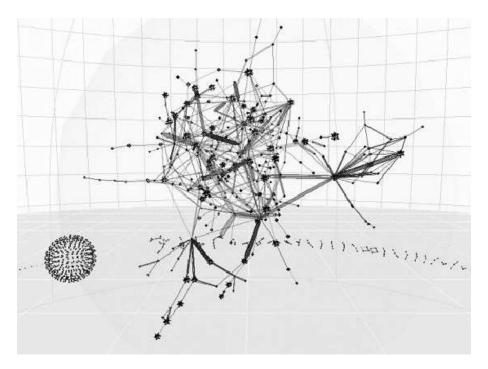


Figure 3. Starlight NIDS Graph

2.2. Highly Optimized Tolerance

Highly Optimized Tolerance (HOT) is a generative mechanism that seeks to explain the structure, statistics and resiliency of interconnected systems. Originally proposed to account for the ubiquity of so-called power laws in natural and engineered systems, it has been fruitfully applied to the study of forest ecosystems, router network robustness, internet traffic, power systems and immune systems. The strength of HOT models is four-fold: First, by virtue of its emphasis on evolved and engineered complexity through feedback, tradeoffs between objective functions and resource constraints in a probabilistic environment, it models the majority of real-life systems which are subjected to such pressures. Secondly, its features include high efficiency, performance, and robustness to designed-for uncertainties, i.e. 'average' cases. Thirdly, it conversely exhibits hypersensitivity to unanticipated perturbations, i.e. 'rare' cases. This too, is a feature of most systems, as we will see. Lastly, unlike rival generative mechanisms, the resulting structural configurations are domain-specific and non-generic [9]. For a discussion of power laws, a primer on HOT and a survey of generative mechanisms (including HOT), the reader is referred to [10,11,12], respectively.

2.3. HOT example: Buffer overflow

We shall proceed to present a first example to highlight a HOT process-induced vulnerability that can be subject to a 0th order attack.

Below we find an instantiation of a HOT model: A Probability, Loss, Resource (PLR) optimization problem [13]; a generalized restatement of Shannon source coding

for data compression yielding the Shannon-Kolmogorov entropy for the objective function J. The reader is referred to [14] for details and more examples.

$$\min J$$
 (1)

subject to

$$\sum r_i \le R \tag{2}$$

where

$$J = \sum p_i l_i \tag{3}$$

$$l_i = f(r_i) \tag{4}$$

$$1 \le i \le M \tag{5}$$

We have a set of M events (Eq. 5) occurring iid with probability p_i incurring loss l_i (Eq. 3), the sum-product of which is our objective function to be minimized (Eq. 1). Resources r_i are hedged against losses l_i , with normalizing $f(r_i) = -\log r_i$ (Eq. 4), subject to resource bounds R (Eq. 2). We will demonstrate a mapping between this abstracted PLR model and the following short C program (adapted from [15]) which will be subjected to a buffer overflow.

```
#include <stdlib .h>
#include <stdio.h>
#include <string.h>
int provePequalsNP()
/* Next paper .. */
}
int bof()
{
char buffer [8]; /* an 8 byte character buffer */
strcpy(buffer,gets()); /* get input from the user*/
/* may not return if buffer overflowed */
return 42;
}
int main(int argc, char ** argv)
bof(); /* call bof() function */
/* execution may never reach
next function because of overflow */
provePequalsNP();
return 1000000; /* exit with Clay prize */
}
```

We shall assume here that the probabilistic environment is adequately represented by the user. She is asked for input in gets(), this represents the event. In the C code, the human designer specified an 8 byte buffer (char buffer[8]) and the compiler would

dutifully allocate the minimum buffer needed for 8 bytes (this is the resource r). Hence, the constrained resources r is the variable buffer. The loss associated with the user input event is really a step function; as long as the user satisfies the assumption of the designer, the 'loss' is constant, and can be seen (simplified) as just the 'normal' loss incurred in proper continuation of control flow. Put differently, as long as user input is ≤ 8 bytes, the resource r is minimally sufficient to ensure normal control flow continuation. If, however, the user decides to input 'Honorificabilitudinitatibus' (this lengthy wink to the Bard was implicitly assumed to be an unlikely/impossible event by the human designer in the code declaration), the loss l functions takes a huge step jump: a catastrophic failure ensues since strcpy(buffer,gets()) overflows buffer. The improbable event breaches the resource and now, control flow may be rerouted, the process crashed, shellcode executed via a stack overflow - or in our example, fame remains elusive.

How did this vulnerability come about? In keeping with our hypothesis, we may discern two distinct, domain-specific HOT (Highly Optimized Tolerance) optimization processes at play - one involving human designers and the other, code compilers - that had a hand in allocating the resource that was breached. The first domain-specific mechanism that induces a cost-optimized, resource-constrained structure on the executable program is the human element. Humans using best-practice software development techniques have to juggle at various stage of the design and coding stages: Evolvability vs specificity of the system, functionality vs code size, source readability vs development time, debugging time vs time-to-market, just to name a few conflicting objective function and resource constraints. The second domain-specific mechanism that induces a cost-optimized, resource-constrained structure on the executable is the compiler. The compiler functions as a HOT process. Cost function here include memory footprint, execution cycles, and power consumption minimization, whereas the constraints typically involve register and cache line allocation, opcode sequence selection, number/stages of pipelines, ALU and FPU utilization.

3. Background and Related Work

The issues of vulnerabilities in ancillary systems and their impact on end systems have been discussed in the popular press. Makansi issues a clarion call to action - part historical, current and future US survey, part Cassandra-cry [16] - on the sorry state of the US electricity grid. Pertinent to our discussion is his focus on the grid's transmission subsystem: Maintenance neglect of transmission lines, pylons and most importantly, the nearlyunguarded substations. It is the opinion of the author that the neglect of the ancillary transmission system viz. the grid system constitutes a prima facie example of constraintbased value optimization as suggested by HOT, given that the former accounts for less than 10% of the electricity asset value chain.

Within a more general framework of catastrophic societal scenarios, Clarke [17] raises awareness of seldom-mentioned ancillary systems. He stresses hidden but pervasive technological and social interdependence and subsequently calls for a more expansive definition of critical infrastructure. In the context of nth order attacks, he mentions the essentially defenseless railway system and abounding chemical plants (a devilish target, since chemicals are very often shipped on railways through population centers). His emphasizing near-blind spot subsystems like kindergarten teachers (in the US, around

20% of the population is in K-12 schools for about half the day) and morticians/undertakers³) remains a rare and meretricious exception.⁴

The modeling tools provided by complex network theory have been used to evaluate the susceptibility of critical infrastructure to both failure and attack. Network theory lends itself to the main concepts of this paper, in that network graphs can be used to represent influence diagrams, and system decomposition. In addition, through statistical link-node distribution analysis, one is able to define a variety of centrality (vulgo 'importance') metrics (see Newman [19] for a book-length academic primer). Static social network analysis was applied by Celebi [20, pp. 127-141] to network graphs of websites affiliated with the terrorist PKK. Using graph metrics such as geodesic distance, connectivity and principal component analysis, the goal was to identify the most influential websites (so-called hubs) order to break information connectivity; in other words, pinpointing neuralgic nodes for removal to impede the functioning of the network.

Saddling the horse from the other end- and as a cautionary tale of what can be learned in open societies built on trust - is the nigh unbelievable story⁵ of the PhD thesis White House cybersecurity czar Richard Clarke wanted 'burned' in 2002. Sean Gorman, a geography PhD student at George Mason University, gathered data on the US's fiber optic cable network entirely from open sources. He managed to layer the fiber-optic infrastructure - the information backbone supporting much of the US's military, civilian, financial, air traffic, water, power and control critical infrastructure - onto business and industrial sectors. The resulting map, which he could mine algorithmically with network analysis methods for neuralgic points, was termed a 'terrorist treasure map'. In the end, he was allowed to publish a neutered version of his thesis [21].

From a dynamic modeling perspective in the context of TCP network/web server request adaptation mechanisms, the paper series by Guirguis and Bestravos serve as a good starting point [22,23]: They systematically investigate so-called Reduction of Quality (RoQ) attacks. RoQ attacks target adaptation mechanisms used in network protocols. They achieve their effectiveness by non-DoS, low-bandwidth traffic maliciously optimized against the admission controllers and load balancers, thereby continuously forcing the adaptive mechanism to oscillate between over-load and under-load conditions. Conceptually speaking, RoQ attacks may be viewed as a class of nth order attacks (1st or 2nd order degradation attacks). Fig. 4 shall help us understand the generalizable modus operandus of RoQ attacks.

Assume the system services requests at a high steady state rate x^* , thanks to its adaptation subsystem that seeks to optimize service rates. Malicious traffic in form of an RoQ attack (burst time t shaded) push the system from its steady state equilibrium; the system, through its adaptation mechanism, slowly convergences at rate ν to the new, lower steady state y^* . Since attacks have ceased, after some time, the system's adaptation mechanism is able to converge at a higher rate μ back to the the high steady state x^* . Optimized RoQ attacks would then begin anew, forcing the system to oscillate between x^* and y^* just when it has settled, thereby degrading performance of the end system.

³From [18]: "... the most terrifying aspect of the epidemic was the piling up of bodies" and from historian Alfred Crosby as quoted in [17, p.166]: ".. the accumulation of corpses will, more than anything else, sap and even break the morale of a population"

⁴Clarke's epistemological mindset of possibilistic vs probabilistic thinking heeds poet's William Carlos Williams' admonition: *What would happen in a world, lit by the imagination?* If on nothing else, decision

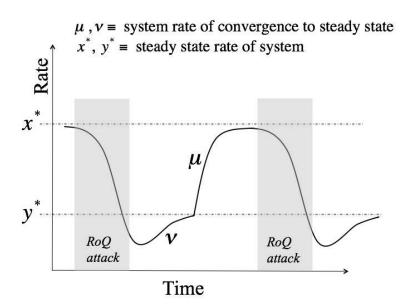


Figure 4. RoQ attacks force the adaptation mechanism with malicious traffic into dropping from a high system steady state rate x^* into to lower system steady state y^* . Picture adapted from [22, p. 3]

Putting it in the nomenclature used in this paper: The RoQ attack's δ requests per second for burst time t (grey shaded) repeated over period T constitutes the 'rare event' which the adaptation system was not expected to handle efficiently. Hence, the adaptation mechanism - as a HOT process designed for common perturbations, but fragile towards rare events - finds its internal assumptions (designed for normal traffic) violated. We now move on to concrete examples of nth order attacks.

4. Example of nth order attacks

4.1. Embedding Ancillary System

Estonia, after regaining independence in 1991, decided on a massive nation-wide 'cyberfication' program: Comprehensive Internet access together with a population registry for authentication/ID purposes would enable the Baltic nation to 'Tiger-Leap' into the 21th century. The result of this push was an extraordinarily far-reaching state information system consisting of (among other things) a PKI infrastructure, over 70 state information systems, financial institutions, state/private portals and associated data exchange layer subcomponents.

In April/May 2007, Estonia suffered a two-phased denial of service attack (predominantly ICMP and TCP SYN⁶). The first phase (04/27/07 - 04/29/07) knocked out government web servers and news sites, and included some semantic hacking such as web de-

makers are strongly urged to follow up on Clarke's works.

⁵See a 2003 Washington Post article at http://tinyurl.com/zuyrv

⁶Nazario offers insightful traffic analysis of Estonia (http://tinyurl.com/2359fq) and the more intense 2008 South Ossetia attack (http://tinyurl.com/6psa6r)

Attacks	Destination	Owner	Description
35	195.80.105.107/32	pol.ee (now politsei.ee)	Estonian police
7	195.80.106.72/32	www.riigikogu.ee	Estonian Parliament
36	195.80.109.158/32	www.riik.ee, www.valitsus.ee	State communication entry portal, Estonian Gov- ernment
2	195.80.124.53/32	m53.envir.ee	Ministry of the Environment
4	213.184.50.6/32	Estonian CERT	
6	213.184.49.194/32	www.agri.ee	Ministry of Agriculture
35	213.184.50.69/32	www.fin.ee	Ministry of Finance
1	62.65.192.24/32	starman.ee	Private telecom provider

 Table 1. Second phase, 128 DDoS attacks: ICMP (115), TCP SYN (4), generic (9). Most serious 10 attacks:

 10+ hours at 90 Mb/s. Peak on May 9: Attack shut down 58 sites at once. Data from Nazario (Arbor Networks)

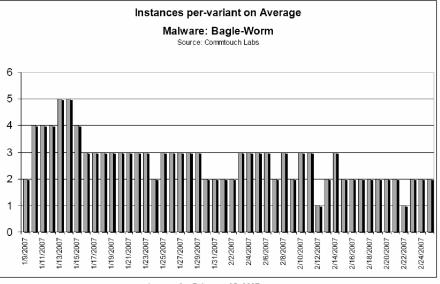
facements. The second phase (04/30/07-05/17/07), coordinating a botnet encompassing some 178 countries, was aimed at critical infrastructures: The two largest banks, neural-gic routers at the ISP level and some governmental portals which were unavailable for a couple of hours. As can be gleaned from Table 1, during the second phase of attacks, the police, government and state communication portals, as well as the Ministry of Finance bore the brunt of the traffic.

This case also highlights the question of perspective in classifying the level of indirection of an nth order attack. On one technical level, the attack could be classified as 2nd order degradation attack, since it consisted of relative primitive DoS traffic aimed at resource allocation mechanisms underlying electronic services. From the point of the individual, it may be classified as a 3rd or 4th order destabilization attack, since it, say, undermined the information infrastructure needed for data exchange between the supermarket and the banks that enable him/her to use credit cards to buy groceries. For a short description of Estonian development, a timeline of the two-phased cyber-attack that took place and subsequent reactions, the reader is referred to [20, pp. 93-103]. We would like to stress these cyberattacks went hand-in-hand with planned physical disruptions: SMScoordinated flash mobs causing traffic jams, trade and tourism interruption by train and road blockades, physical attacks against parliament, and more. This synergistic levèe en masse of the Russian ethnic minority to foment unrest on the ground, in conjunction with the cyberattacks against societal critical infrastructure (see Table 1) were aimed at destabilizing Estonian society. In its comprehensiveness and goals, these efforts constituted the rare event in our model; in terms of modern conflict, it heralds a new class of 'total war' (see Sec. 7).

4.2. Business Model Ancillary System

The email-born Bagle worm first appeared in January 2004 and still ranks - 5 years later - among the top 15 malware families found in the wild, with a prevalence of roughly 2%. It reached its apex in 2006/2007, ranking among the top four, with a prevalence of roughly 15%. For an incisive write-up, the reader is referred to [25].

What makes this worm noteworthy in our context is its 4th order attack m.o.: Through a clever blend of so-called server-side polymorphism and 'high variant-low instance' release, it managed to circumvent conventional pattern-based antivirus (AV) signature detection by *attacking the economic cost structure of the AV companies* itself. With serverside polymorphic malware, the mutation and encryption code transform engine that produce variants is not incorporated into the individual instances, but resides remotely on a server. This outsourcing make the job of traditional signature-based AV companies (who



January 9 – February 25, 2007

Figure 5. Small batches per variant. Picture from [24].

analyze the specimens) harder, since their analysts have less of a code base to work with. This in and of itself could have been dealt with: Bagle's true innovation was to sabotage the economic incentives of AV companies to distill such a signature by generating enormous number of variants in very small batches.

Fig. 5 illustrates the simple but highly effective distribution approach: It lists the average number of instances of the same code, per variant each day of the report period. We see that very small batches of the same code were released but a huge number of variants thereof (30'000 distinct variants server-side supplied in 2007 alone, an average of 625 new variants a day[24]). This constitutes arguably a 4th order attack, since this mechanism neither targeted a vulnerable program on the end system (0th order), nor disabled host or server-based AV services (1st order), nor targeted (say through denial of service or DNS rerouting) either the start or end points of the AV signature distribution system (2nd and 3rd order), but cleverly vitiated the economic incentives of the AV companies to develop signatures (4th order). With modern malware, it is simply not cost effective to invest even one day's worth of highly skilled analyst's time to develop signatures for rapidly mutating, low-count instances - exactly the type of rare event for which the business model was not designed.

4.3. Human Operator Ancillary System

Bond and Danezis invite the reader to entertain following Gedankenspiel [26], inspired by Faust's pact with Mephistoteles: Person W sends a program to person Z, accompanied by an email singing said program's praises. For it promises powers: The power to remotely browse X's hard disk, the power to read the emails between X and Y. Curiosity and maybe malice piqued, Z installs the program and lo, it does not deceive: It delivers on its promises, certainly, but surreptitiously keeps a log of Z's activities and rummages through Z's files. After a critical mass of incriminating evidence is gathered, the program now uses a combination of threats and bribes to get Z to propagate itself: From Data Destruction ("I'll delete all your files") to Revelation ("I'll tell Y you were spying on X and Y')' to Reporting ("I'll report your illegal downloads to the RIAA") to Access Denial ("I'll encrypt all your files") to Freebies ("You'll get tons of free software") and the promise of more powers ("You'll get the power to watch webcams").

The truly devious innovation of this SATAN virus consists of very elegantly leveraging the *psychological ancillary system of the human operator*: It appeals first to a mix of neutral (curiosity, risk) to base (greed, lust for power) psychological instincts. After a time of reward to re-enforce the risky behavior, it then brings the full gamut of shame, fear, cowardice and cognitive dissonance to bear in order to harness two additional subsystems of the human operator: His own human operator subsystem (select the next victim) and his rational subsystem (convince him/her to install me). The induced calculated betrayal of interpersonal trust (the rare event) seems particularly odious. You can almost see the friend exclaiming: "How could Z do this to me, as a friend?" As far as 1st or 2nd order subversion attacks against human operators are concerned, the conceptual SATAN virus is extraordinarily clever.⁷

5. Analysis

With reference to the schematic network graph given in Fig. 6, the US national end 'super' system of interdependent critical infrastructure ancillary systems, we outline some characteristics for a theoretical nth order attack analysis framework.

- 1. We require first a notion of *evolving system state*, since we are dealing with dynamical systems.
- 2. Any model has to furthermore incorporate notions of *cross-dependencies*, since systems are open and coupled.
- 3. These dependencies must include *ties to assumption violations* (as denoted in Eq. 2 of the HOT model in Sec. 2.2) to propagate effects between systems.
- 4. These propagated dependencies must have an *impact* on the system state that is quantitatively measurable.
- 5. Lastly, the modeling formalism has to be high-level enough that there be a reasonably direct correspondence between the system elements modeled and the formalism of the approach.⁸

We explain the rationale for these requirements with the help of Fig. 6. For instance, the communications infrastructure is powered primarily through the power infrastructure. If power delivery is disrupted, telecommunications may switch to backup generators which rely on fuel from the energy distribution infrastructure, delivered via the transportation infrastructure paid for through the financial infrastructure. Conversely, the communica-

⁷It is the author's opinion that this conceptual SATAN virus offers one more astounding innovation, namely symbiotic human-viral code. Even more extraordinary from the point of view of information complexity, the probably simpler viral code manages to induce the 'production' of the more complex human code (propagation module) *dynamically* by invoking evolutionarily and socially generated 'factory routines'.

⁸As an wished-for bonus (maybe there is a Santa Claus), model analysis should be tractable, i.e. any modeling approach used must try to avoid combinatorial state space explosions

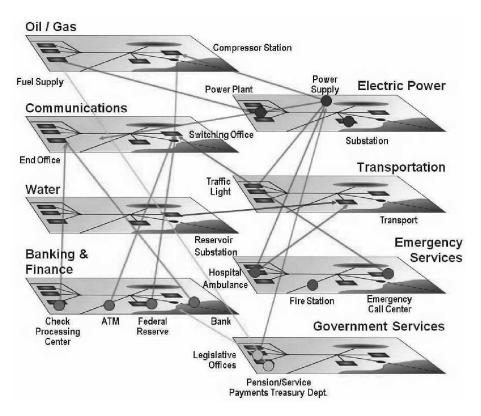


Figure 6. Network of Critical Infrastructure. Picture from Sandia as shown in [27, p.12]

tions infrastructure provides control to the power and transportation infrastructure and underlies much of the financial infrastructure. A fair question that a candidate framework should be able to answer: How much power will we lose for how long if we degrade the communications infrastructure's performance by 20%?

5.1. Theoretical framework

There is a wealth of research on static network graph analysis (see [28] for a practical overview); its main drawbacks remain the inadequate handling of evolving dynamic behavior and cross-dependencies/feedback loops. Since we are concerned with system failure/degradation/subversion, reliability theory formalisms and models suggest themselves.

A first stab system decomposition into constitutive subsystems can lend itself to a simple Fault Tree Analysis. FTA has been used for decades to model failure in multicomponent systems. Invented in 1961 by Bell Labs to improve the reliability of the Minuteman Launch Control System, it has since then been extensively used for evaluating system safety in engineering disciplines as diverse as power, nuclear, electric, and source code analysis [29]. FTA investigates independent pathways between failures of components that lead to the fault tree's top-event. In our parlance, this would be affecting the the end system. Its representation takes the form of a logical diagram in which the topevent's occurrence depends on a specific combination of basic events, which are com-

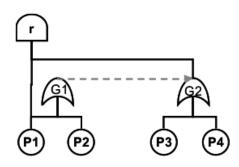


Figure 7. A BLDMP (\mathcal{F} , **r**, **T**, (P_i)) consists multi-top coherent fault tree \mathcal{F} , main top event r, set of triggers T, set of 'triggered' Markov processes P_i associated with leaves of \mathcal{F} (denoted by the red dashed line), and two categories of state for P_i , normal and failure that are triggered via appropriate transfer functions. Picture from [31]

bined with two primary types of gates, AND and OR. Canonical FTAs have no notion of component dependencies, or conditional event sequence timing. As such, they do not meet our requirements; however, extensions such as ones offered in Dynamic FTA [30] do incorporate some, but not all, requirements delineated above.

We offer one modern approach (itself a generalization of Dynamic Fault Trees) that may be fruitfully applied for our purposes, subject to our requirements: Boolean Logic Driven Markov Processes (BLDMP) [32]. BLDMP combines low-level global Markovian state space evolution with a higher level FTA modeling approach. Each leaf is associated with a Markov process which can model the dynamic behavior of a system. Forms of cross-dependencies can be modeled by triggered Markov processes. This fault tree represents the 'structure function' of the system. This structure imposed on the Markov graph can be used to prune the state space, thereby avoiding combinatorial explosions and making analysis more tractable. It remains an open question, though, whether the quantitative impact of these propagated dependencies can be determined analytically, given the non-linear complex dynamics of the setup, or whether one has to resort to an mix of expert judgment, simulation results and historical empirical data.

There exists an alphabet soup's worth of alternative formalisms describing dynamical systems, each with their strength and weaknesses. We briefly mention so-called Dynamic Reliability Block Diagrams (DRBD), as developed by [33]. In contrast to BLMDP's hybrid state space/combinatorial formalism, DRBD is based on the single, high level formalism of RDB [34, Sec. 3.10]. Of interest to us is primarily its dynamic expressiveness, which derives from a technique to model at a low level simple dependencies. This basic dependency 'building block' can be combined with others to model any dynamic behavior (see [35, sec. 3]). The topic space is by no means exhausted: For an extensive reference list of methodologies/formalisms and a reference work, the reader is referred to [33, Sec. 6][34], respectively.

In terms of developed models, the Vulnerability of Information System (VIS) project [36] very recently took a stab at answering questions similar to the critical infrastructure one posed at the beginning of Sec. 5. VIS attempts to quantitatively measure the impact of unexpected Information Communication Technology (ICT) breakdown on economic sectors deemed potentially most ICT-vulnerable. Fondazione Formit (VIS' project lead) singled out five countries (Ireland, Italy, Luxemburg, Romania and Spain) and identified

key ICT-impacted economic sectors within each: Among these, we find public administration, sewage disposal, telecommunication, finance, water and electricity supply - sectors which qualify as critical infrastructure. They subsequently selected representative companies within these sectors for micro-analysis to study ICT breakdown effects, existing recovery strategies and costs. On a macro-level (and more interesting in the context of our discussion), a sophisticated econometric partial equilibrium model taking EU sectorial interdependences as well as cascade effects into account was developed. The model, which allowed for free variables such as ICT breakdown intensity, breakdown and recovery time length, measured the effects of reduced ICT performance on output and value loss, employment and price change, as well as social welfare loss with a time horizon of one day to three months. The validity of some model output was also assessed by means of expert judgment impact analysis in the case study companies and subsequent country-specific micro-simulations.

As of the time of this writing (July 2009), the final report is in the works; perusing available material, the numbers seem overly optimistic: Cumulative Spanish output loss after one month (assuming 10% ICT instantaneous loss, with 50% recovery in five days) hovers below 1% across all economic sectors. One might take issue with the strong equilibria assumptions in the econometric model, yet the crux lies with the recovery assumption: 10% ICT loss with 50% recovery within 5 days may be realistic in terms of accidents or technological glitches, but very likely unrealizable in the face of intentional attacks (in fairness, intentional attacks were puzzlingly not in the project scope's risk space). Thus we stress again, albeit in a different context, the pitfalls of strong assumptions, as well as the dangerous allure of fantasy recovery ('error handling') documents, a topic which we shall return to below.

5.2. Practical framework

Since analytical modeling proves to be non-trivial in its requirements, perhaps an approach along the lines of a simulation offers an alternative. Indeed; given a controlled, instrumented environment in which the end system can be situated, actual nth order attacks against ancillary systems and their concomitant effects can be observed and then evaluated. Such is the case with software application running on a single machine, where destabilization efforts can be effected through an embedding ancillary system acting as mediating OS middleware. We list Holodeck⁹; a fault injection framework that allows Windows programs to run in simulated hostile environments [37]. Its functionality includes the ability to create resource starvation situations affecting ancillary systems such as memory, hard disk, network bandwidth; as well as error handling ancillary system in the form of data poisoning such as corrupted resource files/network streams, unexpected API return values, and a gamut of explicit fault injections.

Empirical evidence collected over two decades support Holodeck's emphasis on *error handling ancillary systems*. Miller subjected Unix, Windows and OS X utilities in the simplest case to random (not malicious) keyboard input, and reported end system crash failure rates of 25%-40%, 24%, and 7%, respectively [38][39]. Sociological and organizational case studies by Clarke [40], analyzing what he terms 'fantasy documents'

⁹Commercially available at http://www.securityinnovation.com/holodeck/

(disaster contingency plans¹⁰), corroborate the brittleness of error handling subsystems, as well.

6. Remediation

In our view, remediation efforts must either address the assumption violations underlying the vulnerabilities, or devise a control mechanism to keep the system in a stable state, should it come under attack. We crystallized thusly: Since we posited that the etiology of nth order attacks (any attack) lay in the HOT-induced violations of assumptions, is there a way of dynamically mutating those assumptions? If not, can we prevent malicious parties from learning of these assumptions? Lastly, if we cannot prevent a violation, can we return a system back to a stable state?

An effective, protocol compliant, but rarely used TCP feature in Linux kernels exists which prevents some forms of degradation attacks against the TCP resource allocation mechanism: SYN Cookies. The server outsources the state of a half-open connection (kept normally on the server) in the form of a cryptographic challenge (the cookie) back to the client[41]. This is an example of an assumption mutation. Internet cognoscenti have heard of the 'Slashdot' effect - when legitimate connection requests overwhelm the server because of popularity of content. This problem was tackled early on in 2001 by Akamai [42] in the form of dynamic load balancing, which constitutes a runtime assumption mutation.

Keeping parties from learning about exact resource boundaries (and subsequent exploitation) may be able to borrow methods from thwarting so-called side channel attacks. Side channel attacks try to infer a system process' state by means of (sometime inadvertently, sometimes unavoidably) leaked observables like time to completion, EM radiation, sound, protocol return values generated in course of the system's evolution. These attacks range the gamut from ingenious timing analysis on B-tree lookup operations and data structure rebalancing (which lead to the release of database privileged information [43]), to differential power analysis where current used in switching reveal activities that can be mapped to processes [44], to CPU operation inferences through characteristic acoustic spectral signatures [45]. In all these instances, processes leaked information. It may be possible to design and operate systems in such a way that the leaking of resource boundaries (the assumptions an attacker wants to violate) is minimized. We hypothesize that designs that incorporate the insights of Maximum Entropy Principles (for an introduction see [46]) are a step in the right direction.

For state control, Ott's [47] work on controlling chaotic systems may yield some fruitful insights, since the interdependent, nested systems under consideration in this paper are more than likely to exhibit non-linear, complex, chaotic behavior due to feedback relationships. In a nutshell, Ott's OGY method injects tiny perturbations into the system when it threatens to veer off towards an unstable state. These perturbations (a control vector based on the system state's Jacobian eigenvectors) 'nudge' the chaotic system back towards a fixed point and into a stable state. For a beginner's primer on non-linear systems, the reader is referred to [48].

¹⁰A classic remains LILCO's ill-fated February 13th 1986 Shoreham evacuation plan. The aim of this exercise was to demonstrate the evacuation plan feasibility. It failed at step 1: The bus drivers (a logistical and psychological vital link; tasked among other things to evacuate children) failed [40, pp. 26-30]

7. Epilogue

In a worthwhile comparative study [49], Fukuyama of 'End of History' fame discusses the notion of societal trust as a gateway to prosperity. He maintains that members of 'high-trust' societies (like the United States) can leverage wide-circle (beyond family ties) trust to form efficient, optimized civic and economic organizations. It is hard to overstate how deeply this trust subsystem permeates every facet of open societies, how much it lowers tangible and intangible transaction costs between individuals, corporations and the state, and how easy an assumption it is to violate for malicious actors - with disastrous effects on the end system.

This realization was not lost on Bin Laden and his fellow strategists. In a 2004 broadcast, he boasted (quoting research from Chatham House [50]) that the 9/11 attacks had cost al-Qaeda only \$500,000 while inflicting at least \$500 billion of economic losses on America. Accordingly, the Islamist supremacists' playbook calls for beating the US by systematically attacking the US economy's vulnerabilities. The most accessible vulnerabilities in open societies are induced by deeply ingrained trust assumptions these societies have developed over decades and take for granted: that freedom of movement, freedom of speech, freedom of religious assembly, assistance from the social welfare state, immigration policy will not be used to subvert society from within; that a participant in mass transit, a shopper at the mall, a fertilizer buyer, a student reading nuclear engineering, a worshipper at a house of prayer will not commit mass murder. The chilling passage (excerpted from [51]) is worth quoting at length (italics are ours):

The Islamic nation has entered through al-Qa'ida's war with America a new period that is different from all the other periods experienced by Muslims against their enemies. This period is based on economic war due to the peculiar nature of the adversary in this ferocious battle. Usually, wars are based on military strength and victory belongs to those who are militarily superior on the battlefield...But our war with America is fundamentally different, for the first priority is defeating it economically. For that, anything that negatively affects its economy is considered for us a step in the right direction on the path to victory. Military defeats do not greatly effect how we measure total victory, but these defeats indirectly affect the economy which can be demonstrated by the breaching of the confidence of capitalists and investors in this nation's ability to safeguard their various trade and dealings [..] Any operation targeting an area of infrastructure in a new country that does not have a history of countering these operations is considered as bleeding (exhausting) to the greater enemy America and the targeted nation itself. It is so because these nations will be required to protect all similar potential targets which results in economic exhaustion (bleeding)... For example, if a hotel that caters to western tourists in Indonesia is targeted, the enemy will be required to protect all hotels that cater to western tourists in all countries which may become a target of similar attacks. You can say the same thing about living residences, economic establishments, embassies [..]

Similarly, the PRC People Liberation Army's emphasis on asymmetric warfare and ongoing push to develop modern "Assassin's Mace" weapons within the doctrine of "The Inferior Defeats the Superiors"¹¹ should give some pause. The Director of Foreign Military Studies at the Academy of Military Sciences in Beijing, Major General Pan Junfeng, offered following tidbits reminiscent of nth order warfare (presumably against the US) in a 1996 issue of China Military Science (as cited in [53, p.12]):

¹¹Philosophic outlines of said doctrine are already found in Sun Tzu, the modern incarnation can be traced to Mao, implementation to the 1980s, and open discussions among specialized scholars abound since the early 1990s [52]

We can make the enemy's command centers not work by changing their data system. We can cause the enemy's headquarters to make incorrect judgments by sending disinformation. We can dominate the enemy's banking system and even its entire social order.

The interested reader is invited to peruse the some of the PLA's official and unofficial takes on future warfare in [54,53].

We would be remiss in our discussion if we were not to mention an nth order attack against the ultimate ancillary system: Electromagnetic pulse attacks against the electricity grid. An April 2008 report to the US House Armed Services Committee [27] outlined the effects on critical civilian infrastructure, should a nuclear weapon¹² be detonated 200-400 miles over Kansas (italics are ours):

The functioning of society and the economy is critically dependent upon the availability of electricity. Essentially every aspect of American society requires electrical power to function. Contemporary U.S. society is not structured, nor does it have the means, to provide for the needs of nearly 300 million Americans without electricity. Continued electrical supply is necessary for sustaining water supplies, production and distribution of food, fuel, communications, and everything else that is a part of our economy. [..] *No infrastructure other than electric power has the potential for nearly complete collapse in the event of a sufficiently robust EMP attack* [..] Large-scale load losses in excess of 10 percent are likely at EMP threat levels. Instantaneous unanticipated loss of load, by itself, can cause system collapse. This is possible at 1 percent loss, and is very likely above 10 percent [..] Should the electrical power system be lost for any substantial period of time, the Commission believes that the consequences are likely to be catastrophic to civilian society. Machines will stop; transportation and communication will be severely restricted; heating, cooling, and lighting will cease; food and water supplies will be interrupted; and many people may die.

We therefore close on a somber note: The issues touched upon in this paper are not merely of academic or scientific interest. In practical terms, they go to the very heart of how future conflicts between open societies and their enemies will be waged - and are waged as we speak.

References

- [1] K. Thompson, "Reflections on Trusting Trust," CACM, vol. 27, pp. 761–764, August 1984.
- [2] L. Von Bertalanffy, "An Outline of General System Theory," *British Journal for the Philosophy of Science*, pp. 134–165, 1950.
- [3] P. Érdi, Complexity Explained. Springer, November 2007.
- [4] G. Gigerenzer, Gut feelings: The Intelligence of the Unconscious. Viking Books, 2007.
- [5] G. Torres and C. Lima, "Maximum CPU Temperature." http://tinyurl.com/oytsnv, October 2007.
- [6] US DOE, "Starlight Information System." Pacific Northwest National Lab, 2003.
- [7] G. Conti, "Attacking Information Visualization System Usability Overloading and Deceiving the Human," in *Proceedings of the 2005 Symposium on Usable Privacy and Security*, pp. 89–100, ACM, 2005.
- [8] W. Poundstone, *Gaming the Vote: Why Elections Aren't Fair*. Hill and Wang, 2008.
- [9] J. M. Carlson and J. Doyle, "Highly Optimized Tolerance: A Mechanism for Power Laws in Designed Systems," *Physical Review E*, vol. 60, no. 2, pp. 1412+, 1999.
- [10] A. Clauset, C. R. Shalizi, and M. Newman, "Power-Law Distributions in Empirical Data," SIAM Reviews, June 2007.
- [11] J. Carlson and J. Doyle, "Highly Optimized Tolerance: Robustness and Design in Complex Systems," *Physical Review Letters*, vol. 84, pp. 2529+, March 2000.

¹²The exact yield necessary is presumably classified and/or undeterminable through simulation, expert physicist estimates range from 10-1500 kilotons.

- [12] M. Newman, "Power Laws, Pareto Distributions and Zipf's Law," *Contemporary Physics*, vol. 46, pp. 323–351, September 2005.
- [13] L. Manning, J. Carlson, and J. Doyle, "Highly Optimized Tolerance and Power Laws in Dense and Sparse Resource Regimes," *Physical Review E*, vol. 72, pp. 16108+, July 2005.
- [14] J. Doyle and J. Carlson, "Power Laws, Highly Optimized Tolerance, and Generalized Source Coding," *Physical Review Letters*, vol. 84, p. 5656:5659, June 2000.
- [15] J. C. Foster, V. Osipov, N. Bhalla, and N. Heinen, Buffer Overflow Attacks. Syngress, 2005.
- [16] J. Makansi, *Lights Out: The Electricity Crisis, the Global Economy, and What it Means to You.* Wiley, 2007.
- [17] L. Clarke, Worst Cases: Terror and Catastrophe in the Popular Imagination. University of Chicago, 2006.
- [18] J. Barry, The Great Influenza: The Epic Story of the Deadliest Plague in History. Penguin, 2005.
- [19] M. Newman, A.-L. Barabasi, and D. J. Watts, *The Structure and Dynamics of Networks: (Princeton Studies in Complexity)*. Princeton University Press, April 2006.
- [20] Centre of Excellence Defence Against Terrorism, ed., Responses to Cyber Terrorism, vol. 34 of NATO Science for Peace and Security Series E. IOS Press, 2008.
- [21] S. Gorman, Networks, Security and Complexity: The Role of Public Policy in Critical Infrastructure Protection. Edward Elgar Publishing, 2005.
- [22] M. Guirguis and A. Bestavros, "Reduction of Quality (RoQ) Attacks on Internet End-Systems," in 2005 Proceedings IEEE INFOCOM, vol. 2, March 2005.
- [23] M. Guirguis and A. Bestavros, "Adversarial Exploits of End-Systems Adaptation Dynamics," *Journal of Parallel and Distributed Computing*, vol. 67, no. 3, pp. 318–335, 2007.
- [24] Commtouch, "Server-Side Polymorphic Viruses Surge Past AV Defenses." http://tinyurl.com/2vewz8, May 2007.
- [25] Commtouch, "Malware Outbreak Trend Report: Bagle-Worm." http://tinyurl.com/39gnz4, March 2007.
- [26] M. Bond and G. Danezis, "A Pact with the Devil," in *Proceedings of the 2006 Workshop on New Security Paradigms*, pp. 77–82, ACM, 2006.
- [27] W. Graham, "Report of the Commission to Assess the Threat to the United States from EMP Attack: Critical National Infrastructures," tech. rep., Congressional Report, April 2008.
- [28] W. De Nooy, Exploratory Social Network Analysis with Pajek. Cambridge University, 2004.
- [29] C. Ericson, "Fault Tree Analysis A History," in Proceedings of the 17th International System Safety Conference, 1999.
- [30] J. Dugan and S. Bavuso, "Dynamic Fault-Tree Models for Fault-Tolerant Computer Systems," *IEEE Transactions on Reliability*, vol. 41, no. 3, pp. 363–377, 1992.
- [31] M. Bouissou, "A Generalization of Dynamic Fault Trees through Boolean logic Driven Markov Processes (BDMP)®," in *Proceedings of the safety and reliability conference (ESREL07)*, 2007.
- [32] M. Bouissou and J. Bon, "A New Formalism that Combines Advantages of Fault-Trees and Markov Models: Boolean Logic Driven Markov Processes," *Reliability Engineering and System Safety*, vol. 82, no. 2, pp. 149–163, 2003.
- [33] S. Distefano and A. Puliafito, "Dependability Evaluation with Dynamic Reliability Block Diagrams and Dynamic Fault Trees," *IEEE Transactions on Dependable and Secure Computing*, vol. 6, no. 1, pp. 4–17, 2009.
- [34] M. Rausand and A. Hoyland, System Reliability Theory: Models, Statistical Methods, and Applications. Wiley-Interscience, 2004.
- [35] S. DiStefano, "How to Capture Dynamic Behaviours of Dependable Systems," *International Journal of Parallel Emergent Distributed Systems*, vol. 24, no. 2, pp. 127–150, 2009.
- [36] Formit, "VIS the Vulnerability of Information System and its Inter-Sectorial, Economic and Social Impacts." http://www.formit.org/vis/, May 2009.
- [37] J. Whittaker, How to Break Software Security: Effective Techniques for Security Testing. Addison Wesley, 2004.
- [38] B. Miller, L. Fredriksen, and B. So, "An Empirical Study of the Reliability of UNIX Utilities," CACM, vol. 33, no. 12, pp. 32–44, 1990.
- [39] B. Miller, G. Cooksey, and F. Moore, "An Empirical Study of the Robustness of MacOS Applications Using Random Testing," in *Proceedings of the 1st International Workshop on Random Testing*, pp. 46– 54, ACM, 2006.
- [40] L. Clarke, *Mission Improbable*. University of Chicago, 1999.

- [41] A. Zuquete, "Improving the Functionality of SYN Cookies," in Proceedings of 6th IFIP Communications and Multimedia Security Conference, pp. 57–77, 2002.
- [42] T. Leighton, "The Akamai Approach to Achieving Performance and Reliability on the Internet," in Proceedings of the 26th ACM Symposium on Principles of Distributed Computing, ACM, 2007.
- [43] A. Futoransky, D. Saura, and A. Waissbein, "The ND2DB Attack: Database Content Extraction Using Timing Attacks on the Indexing Algorithms," in *Proceedings of the 1st USENIX Workshop on Offensive Technologies*, pp. 1–9, USENIX, 2007.
- [44] T. Popp, S. Mangard, and E. Oswald, "Power Analysis Attacks and Countermeasures," *IEEE Design & Test of Computers Design and Test of ICs for Secure Embedded Computing*, vol. 24, no. 6, pp. 535–543, 2007.
- [45] A. Shamir and E. Tromer, "Acoustic Cryptanalysis. On Nosy People and Noisy Machines," tech. rep., RSA and MIT CSAIL, 2004.
- [46] H. Kesavan and J. Kapur, "The Generalized Maximum Entropy Principle," Systems, Man and Cybernetics, IEEE Transactions on, vol. 19, pp. 1042–1052, Sep/Oct 1989.
- [47] E. Ott and M. Spano, "Controlling Chaos," Physics Today, vol. 48, no. 5, pp. 34-40, 1995.
- [48] L. Lam, Nonlinear Physics for Beginners: Fractals, Chaos, Solitons, Pattern Formation, Cellular Automata and Complex Systems. World Scientific Press, 1998.
- [49] F. Fukuyama, Trust: The Social Virtues and the Creation of Prosperity. Free Press, 1996.
- [50] B. Hoffman and G. Weimann, "Econo-jihad," May 2009.
- [51] S. Salama, "Unraveling Al-Qa'ida's Target Selection Calculus," *Terrorism and Political Islam*, p. 41:44, 2007.
- [52] J. R. Lilley and D. L. Shambaugh, eds., *China's Military Faces the Future*, ch. 3-4. Studies on Contemporary China, M.E. Sharpe, 1999.
- [53] M. Pillsbury, "China's Military Strategy Towards the United States: A View from Open Sources," tech. rep., US-China Economic and Security Review Commission, November 2001.
- [54] M. Pillsbury, Chinese Views of Future Warfare. National Defense University Press, September 1998.