Preliminary System Dynamics Maps of the Insider Cyber-threat Problem

David Andersen Rockefeller College of Public Affairs and Policy University at Albany 1400 Washington Ave., Albany, NY 12222, USA david.andersen@albany.edu	Dawn M. Cappelli CERT Coordination Center, Software Engineering Institute 4500 Fifth Avenue Pittsburgh, PA 15213-3890, USA dmc@cert.org	Jose J. Gonzalez Faculty of Engineering and Science Agder University College Grooseveien 36 NO-4876 Grimstad, Norway Jose.J.Gonzalez@hia.no	Mohammad Mojtahedzadeh Attune Group, Inc. 16 Regina Court, Suite 1 Albany, NY 12054, USA mohammad@attunegroup.com
Andrew P. Moore CERT Coordination Center, Software Engineering Institute 4500 Fifth Avenue Pittsburgh, PA 15213-3890, USA apm@cert.org	Eliot Rich Department of Information Technology Management, School of Business, University at Albany , 1400 Washington Ave., Albany, NY 12222, USA e.rich@albany.edu	Jose Maria Sarriegui TECNUN, University of Navarra Paseo Manuel de Lardizabal 13 ES-20018 San Sebastian, Spain jmsarriegui@tecnun.es	Timothy J. Shimeall CERT Coordination Center, Software Engineering Institute 4500 Fifth Avenue Pittsburgh, PA 15213-3890, USA tjs@cert.org
Jeffrey M. Stanton School of Information Studies Syracuse University Syracuse, NY 13244, USA jmstanto@syr.edu	Elise A. Weaver Worcester Polytechnic Institute 100 Institute Road Worcester, MA 01609-2280, USA eweaver@wpi.edu	Aldo Zagonel Rockefeller College of Public Affairs and Policy University at Albany 11 Pheasant Ridge Dr. Albany, NY 12211, USA Zagonel@aol.com	

Abstract

Twenty five researchers from eight institutions and a variety of disciplines, viz. computer science, information security, knowledge management, law enforcement, psychology, organization science and system dynamics, found each other February 2004 in the "System Dynamics Modelling for Information Security: An Invitational Group Modeling Workshop" at Software Engineering Institute, Carnegie Mellon University.

The exercise produced preliminary system dynamics models of insider and outsider cyber attacks that motivated five institutions, viz. Syracuse University, TECNUN at University of Navarra, CERT/CC at Carnegie Mellon University, University at Albany and Agder University College, to launch an interdisciplinary research proposal (Improving Organizational Security and Survivability by Suppression of Dynamic Triggers).

This paper discusses the preliminary system dynamic maps of the insider cyber-threat and describes the main ideas behind the research proposal.

1. Introduction

Sometimes an unfinished product is worth presenting. This paper is – in a sense – a report of an unfinished product. Nevertheless, we suggest that its procedures and processes, the preliminary system dynamics maps we have sketched and the future perspectives we envision might be of general interest within the extended definition of the conference theme. The conference theme, *Collegiality*, is elaborated in the conference programme as *"…discussing any special role system dynamics has played in studies, or could play in the future, in the area of consensus building, conflict resolution, knowledge surfacing and sharing, and theory testing.*"

We will discuss our product's genesis (i.e., precursors that lead to an emergent SIG in security and their spin-offs), the variant of group modelling process we employed, the preliminary models and the role system dynamics could play in a promising research agenda. The intended presentation at the Twenty-Second International Conference of the System Dynamics Society will describe main points in a concise manner, followed by a structured discussion with the audience. We humbly hope that some of our outcomes – though admittedly more torsos than statues – might instigate akin developments – hopefully in an evolutionary path of progress. We hope to elicit valuable criticism and good ideas that help us approach our goal, which is quite ambitious: To improve organizational security and survivability by suppression of "dynamic triggers."

1.1 Precursors

About twenty researchers from eight institutions and a variety of disciplines, including computer science, information security, knowledge management, psychology, organization science and system dynamics, participated during February 2004 in the "System Dynamics Modelling for Information Security: An Invitational Group Modeling Workshop" at Software Engineering Institute, Carnegie Mellon University.[1] The workshop consisted of short plenary sessions each morning, followed by two parallel sessions dedicated to insider and outsider cyber attacks.

In retrospect, we decided to view the event as the Second Annual Workshop on System Dynamics Modelling for Information Security. The First Annual Workshop on System Dynamics Modelling for Information Security was a much smaller precursor event that occurred February 2003 at Agder University College in Grimstad, Norway. (This event received its portentous name in retrospect, too.)[2] The outcome of the February 2003 workshop was a number of intensively discussed papers – most of them co-authored by several participants and including one paper that was sketched in a group modelling process during the workshop. The revised and extended workshop papers were submitted to the Twenty-First International Conference of the System Dynamics Society. All were accepted and they got reviews that appears to confirm the usefulness of the "internal" reviewing and group modelling process in the workshop. One of the papers (Cooke 2003a, 2003b) led to the 2003 Dana Meadows Award, all papers appeared both in the

[1] See http://www.cert.org/research/sdmis/

[2] Strictly speaking the first workshop was dedicated to general security systems, although it did have a strong component of information security.

CD-ROM proceedings of the International Conference of the System Dynamics Society and were collected in a small book, entitled "From Modeling to Managing Security: A System Dynamics Approach" (Gonzalez 2003).

In retrospect, the internal quality assurance within a small group with common interests would have naturally belonged to the realm of the Security SIG – but this SIG was not founded and approved by the Systems Dynamics Society until quite recently.

While the book "From Modeling to Managing Security: A System Dynamics Approach" did not generate large demand, it did meet with interest in a crucial audience within the CERT® Coordination Center.[3] A recent study by two CERT/CC researchers had employed qualitative system dynamics (i.e. causal loop analysis) for controlling vulnerability (Ellison and Moore 2003, p. 38ff). Before embarking on that, the authors had remarked «…we are not aware of any work using system dynamics to explicitly study the threat environment or its impact on system operations.» After a general analysis of the capability of system dynamics, they concluded nevertheless: «…, we believe that system dynamics provides a foundation for developing methods and tools that help engineers understand, characterize, and communicate the impact of a malicious threat environment on organizational and system operations and their respective missions. Large-scale, inter-networked information systems are subject to volatility, nonlinearity, uncertainty, and time delays that add to their dynamic complexity and make assuring their security or survivability so difficult.» (ibid., p. 37-38). On hearing from Graham Winch[4] about the security session at the recent Twenty-First International Conference of the System Dynamics Society (which resulted in a system dynamics monograph dedicated to the modelling and management of security threats in organizations), a process took off that ultimately led to the "Second Annual Workshop on System Dynamics Modelling for Information Security" at Carnegie Mellon University in February 2004.

The bridging events between the First Annual Workshop on System Dynamics Modelling for Information Security (with the related security session at the Twenty-First International Conference of the System Dynamics Society) and the second annual workshop in February 2004 were a series of guest lectures with accompanying meetings in October 2003 that involved the authors of this paper (and other colleagues) at Albany, Carnegie Mellon, Syracuse and Worcester Polytechnic Institute.[5] After this round, a CERT/CC researcher expressed his impression thus: «System dynamics has the potential to significantly improve our capabilities and understanding in areas not well addressed by traditional security approaches.» (Lipson 2003)

---

[3] CERT/CC is a U.S.-based center of Internet security expertise, located at the Software Engineering Institute, a federally funded research and development center operated by Carnegie Mellon University.

[4] We are very grateful to Professor Graham Winch for triggering the connection between CERT/CC and the other authors of the monograph.

[5] Haskayne School of Business, University of Calgary was also involved initially but unfortunately various constraints prevented participation by David L. Cooke, who played an essential role at the First Annual Workshop on System Dynamics Modelling for Security.

The preliminary consensus reached in October 2003 motivated researchers at these institutions to pursue vigorously the goal of a collaborative workshop dedicated to preliminary group modelling of cyber security with the intention – if successful – to devise a research agenda and draft papers for this conference.

1.2 The Problem of Cyber Attacks

The growing reliance of government and commercial organizations on large-scale, widely interconnected information systems amplifies the consequences of malicious attacks on organizational information assets. Organizations continue to move key business functions to Internet-based operations, thereby increasing the exposure of critical information assets. In addition, the wide availability of increasingly sophisticated attacker tools is permitting large numbers of relatively inexperienced individuals to execute very sophisticated attacks. The number of security incidents reported by the CERT Coordination Center (CERT/CC 2004) has approximately doubled every year for the last six years, and was estimated to reach more than 137,000 in 2003. Each security incident may involve hundreds, or even thousands, of sites and may involve ongoing activity for long periods of time. Notice that the slope for the incident curve from 1997-2003 is significantly steeper than the corresponding slope for hosts (Figure 1),[6] which might indicate that the complexity of the Internet makes it increasingly difficult to manage its security.

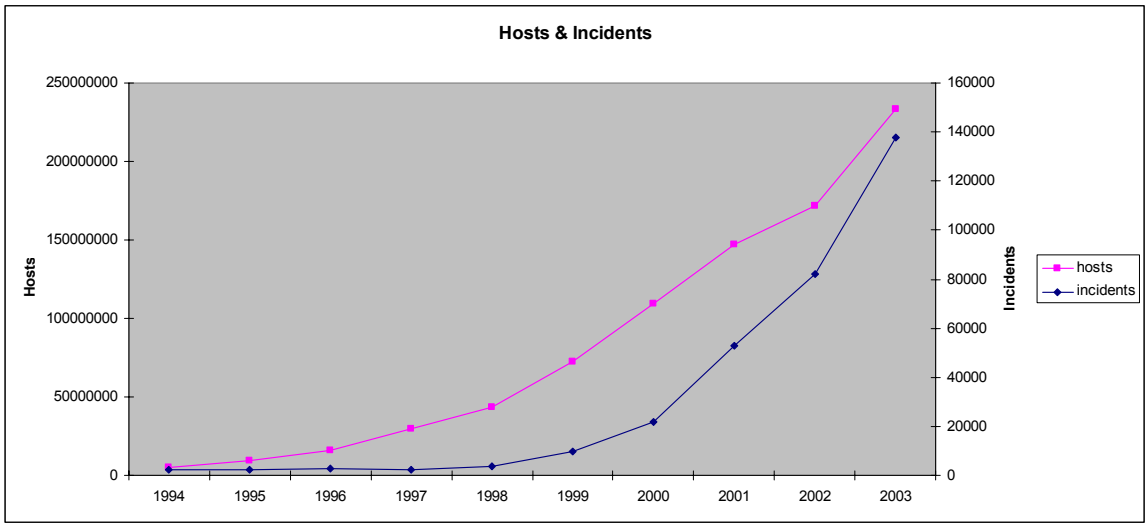

Figure 1 The number of security incidents and the number of hosts. Sources: Incidents: CERT/CC Statistics 1988-2003, (**http://www.cert.org/stats/cert_stats.html**); Hosts: Internet Systems Consortium (**http://www.isc.org/index.pl?/ops/ds/**)

While the prevalence and impact of attacks against organizational information is rapidly increasing, from the point of view of any single organization the probability of a catastrophic attack is still quite small. Characterized probabilistically, such low base rate events can be compared psychologically to other possible catastrophes such as earthquakes, lightning strikes, or plane crashes. In fact, the realization that catastrophic cyber-attacks are

[6] The quotient of "no. of incidents" / "no. of hosts" has approximately been doubling every *second* year since 1997.

of the class of low-base rate phenomena may explain, in part, an observed low degree of preparedness in organizations that have been the victim of cyber attacks. In studies of decision-making, research has shown that judgments of the risk of occurrence and the impact of outcomes of low base rate phenomena are sometimes strongly distorted (e.g., Johnson, Grazioli, Jamal, and Berryman 2001).

In practice, technology alone cannot completely prevent successful attacks on complex, unbounded systems. Network-based information systems increasingly have cross-organizational boundaries, possess no central administration, and lack unified security policies (Lipson and Fisher 1999). The distinction between insider and outsider may be dynamic in that a partner for one activity may be a competitor or adversary for another. These changes force an expansion of security issues from a narrow technical specialty to an organization-wide risk management concern that must deal with broad avenues of attack and the psychological motivations of both attackers and defenders. The technical and business process considerations may interact with key organizational variables including awareness of security issues, the supportiveness of organizational culture, the configuration of organizational procedures designed to protect information assets, and levels of trust and autonomy given to key staff members.

The transition from an optimistic stance that security is mainly a technological issue to a more sober attitude is reflected e.g. in a recent book by Bruce Schneier (2000), entitled "Secrets and Lies – Digital Security in a Networked World." Schneier – who is both a respected security scientist and a successful consultant – admits an early naïveté about the capabilities of technology but now characterizes security as a complex and continuous "process", rather than a "product." Schneier writes: "The Internet is probably *the most complex system ever developed*" (p. 6, emphasis added). Interestingly, Schneier describes information systems using similar language as a system dynamics expert: dynamic entities, interacting components, propagating consequences, unexpected ("emergent") properties and delayed effects (p. 5-9).

The Internet is perhaps the most complex artificial system ever developed; what is worst: it was not designed with security in mind. The CERT/CC researcher Howard Lipson (2002, p. 9) points out: «Perhaps the greatest threat to the Internet today is the abysmal state of security of so many of the systems connected to it.» He states: «…the Internet's fundamental technology was developed to serve the needs of a benign community of researchers and educators.» (ibid. p. 29), implying that the Internet was not designed to resist highly untrustworthy users. Preventing attacks is difficult – detecting them is not simple either. Lipson again: «Facilities for tracking and tracing the behavior of […] users were never a consideration, and a tracking infrastructure was never designed or implemented.» (ibid. p. 27).

It is not difficult to find pessimistic descriptions of the magnitude of the threat to cyber security in the literature that are founded on solid analysis. We restrict ourselves to quoting Lipson once more: «These high-stakes Internet applications pose enormously tempting targets of opportunity for criminals, terrorists, and hostile nations, and they overly stress the technology designed for a more innocent time.» (ibid. p. 29).

1.3 Cyber Data Restrictions

Good data is crucial – but unfortunately we cannot base our analysis on the presumption that relevant cyber data always exists, nor that existent data is available, nor that available data is good. Accordingly, the modelling of cyber systems must use a research strategy that can still deliver valuable insights despite the holes and deficiencies in the data material.

One of the difficulties in systematic modelling of cyber-attacks arises from the unavailability of data regarding these attacks. While such attacks are increasingly frequent on networked systems, systematically collected data on these attacks is not generally available. This lack of availability stems from three basic causes: Attackers generally act to conceal their attacks; defenders gather data on attacks for narrow purposes; organizations controlling information assets rarely share data on attacks.

First, successful information attacks depend to some degree on deception and surprise – networks that are prepared or forewarned for specific attacks generally suffer little or no damage from them. Thus, attackers must conceal as much information as possible on their attacks in order to preserve the utility of their methods – not a difficult task for them since the Internet was devised for "good guys" (cf. previous section). This situation results in incomplete data capture on the methods and objectives of attacks on information assets – with the notable exception arising from work on honeypots and honeynets (Spitzner 2003; The Honeynet Project 2004).

Second, defenders of information assets are often overburdened. As such, they have little motivation for large scale data collection activities. Data are generally collected only if useful for a specific defensive task, for forensic purposes or to document damage relevant for legal proceedings. A wide range of data formats is used in such data collection. The data are organized, not in a generically accessible database, but rather in formats specific to the use for which they are collected, making systematic survey of the data collected quite difficult and time intensive.

Third, attack data are often shared only in vague terms, if at all, by affected organizations. Sharing of information may be precluded by the rules of evidence in a criminal prosecution. In other cases, data on attacks may be withheld due to concerns over publicity, reputation, or worries about copycat activities. When detailed data are shared, they often become available only under restricted-use agreements or guarantees of confidentiality. As such, data that characterizes attacks across a broad range of organizations are rarely available to the research community.

Beyond these three aspects, cyber data that is reported to computer emergency response teams, such as CERT/CC, cannot be shared freely with other researchers – not even with collaborating researchers from other institutions. How to circumvent this problem is still not resolved, despite quite intense discussions during the Second Annual Workshop on System Dynamics Modelling for Information Security in February 2004.

The workshop participants representing the system dynamics methodology argued that system dynamics models would use aggregated data that cannot be traced to reporting institutions. As proxy to the data owners, CERT/CC researchers are in very delicate position: To report data not yet available in CERT/CC statistics– even at the aggregated level needed for system dynamics models – requires permission from data owners (i.e. CERT/CC "clients"). Data owners would not give such permission to researchers not covered by the secrecy agreements regulating the relationship between them and CERT/CC. Accordingly, a research collaboration between CERT/CC (which has not yet hired system dynamics modellers into their group) and outside system dynamics experts has to find ways for how to handle such severe restriction and produce models that nevertheless have some utility.

1.4 The Structure of this Paper

The way we have approached the data availability problem can be summarized like this:

- The Second Annual Workshop on System Dynamics Modelling for Information Security would attempt to create a preliminary system dynamics map of the insider cyber threat problem.[7] By a preliminary system dynamics map we mean a fairly detailed system dynamics model with the basic structure of the problem.
- In addition, the workshop would attempt to identify the most descriptive and significant dynamic stories implied by the preliminary system dynamics map.

We were expecting that the outcome of the workshop would be a decision to join forces, i.e. to design a collaborative research agenda and to write a joint paper. The present paper represents the initial outcome of that collaboration. How this collaboration might help – in the long run – to improve the availability of cyber data is discussed in the last section of this paper.

Section 2 discusses general aspects of the insider cyber threat and discusses the model of the Tim Lloyd/Omega insider attack developed by Melara, Sarriegui, Gonzalez, Sawicka, and Cooke (2003a; 2003b). Section 3 discusses the group modelling process employed at the Second Annual Workshop on System Dynamics Modelling for Information Security and its products, including preliminary system dynamics maps of the insider cyber threat problem. Section 4 discusses our tentative findings and sketches the research agenda of our collaborative group – including why we hope that our future research might contribute to improve the availability of cyber data.

2. The Insider Problem

The CERT® Coordination Center (CERT/CC) of Carnegie Mellon University's Software Engineering Institute has been collaborating with the United States Secret Service (USSS) since 2001 on research of insider threats. This research has been based on in-depth case analysis of actual insider threat crimes, as well as online surveys for gathering

[7] The workshop had two threads; the second one was dedicated to the outsider thread. The basic methodology was different, however. A full-fledged group modelling process was only applied to the insider problem. (Cf. Wiik, Gonzalez, Lipson, and Shimeall 2004, for a parallel paper dedicated to outsider attacks.)

supplemental information. The definition of an insider threat crime adopted in the USSS/CERT Coordination Center research will also be used as a basis for this paper:

> Any information system, network, or data compromise where the suspect has – or used to have – legitimate access to the network/data compromised. The definition includes suspects who are:
>
> 1) current or former employees of the company whose network was compromised;
> 2) current or former contractors of the company whose network was compromised;
>
> Information system, network, and data compromises include any incidents where there is any manipulation of, unauthorized access to, exceeding authorized access to, tampering with, or disabling any information system, network, or data. This includes any efforts to retrieve, change, destroy, or add information to an information system, network or database. These incidents can occur in ANY organization, public, private, or government, in ANY critical infrastructure sector.

The remainder of this paper is not part of the research described above.

2.1 Issues Surrounding the Insider Threat Problem

Risk management of the insider threat problem involves a complex combination of behavioural, technical, and organizational issues. Organizations can concentrate on physical and technical security measures such as authentication mechanisms, firewalls, and intrusion detection systems to defend against external cyber threats. However, insiders may be authorized to bypass all of those measures in order to perform their daily duties. Former employees are familiar with internal policies and procedures, which can also be exploited to facilitate attacks. External attackers can choose collusion with insiders as an attack mechanism. Although insider threats as defined above utilize technology to carry out their attack, a combination of technical, behavioural, and organizational issues must be considered in order to detect and prevent insider threats.

Because insiders are legitimate users of their organization's networks and systems, sophisticated technical capability is not necessarily required to carry out an insider attack. On the other hand, technically capable insiders are able, and have, carried out more sophisticated attacks, that can have more immediate, widespread impact. These technical insiders also sometimes have the capability to "cover their tracks" so that identification of the perpetrator is more difficult.

Insiders can be motivated by a variety of factors. Financial gain is a common motive in certain industries, while revenge can span industries. Theft of intellectual property is prevalent in some sectors, for various reasons: financial gain, grudge against current or former employer, or to enhance an employee's reputation with a new employer.

Organizational issues that factor into the insider threat problem range from the degree and impact of managerial trust in employees, to the organizational culture itself. The study of insider threats in the CERT/CC mentioned above is focusing on identification of behavioural and technical indicators of potential insider threats, suggesting best practices and other recommendations for prevention, detection, and identification, and formulating recommendations for organizational culture in which the risk of insider threats is minimized.

It is imperative that the insider threat research continue, because the impacts from crimes perpetrated by insiders are enormous, ranging from significant financial losses to severe impacts on reputation. A methodology is needed that can capture and analyze the complex interactions between behavioural, technical, and cultural issues, so that an integrated risk management approach can be developed for the problem.

2.2 Overview of Cases in the Public Domain

In addition to the general expertise and experiences of workshop participants, the second workshop used six well-publicized insider attacks as background for discussions. The cases were selected because they were well documented in the public domain and they had a number of properties in common. In particular, the cases involved organizations that provided a very trusted environment for certain classes of employees, including the insider. Recognition of and/or response to security threats (precursors) posed by the insider were minimal or non-existent. Insiders successively tested and/or lessened the organization's security controls. Reduction in security controls helped avoid detection of the attack and magnify its damaging impacts. The six cases exhibited all these properties in varying degrees. These cases also indicated the significance of the insider threat problem and helped with the identification and validation of reference modes of behaviour.

Figure 2 characterizes the six background cases (labelled A through F) along two dimensions – the degree of technical sophistication and the insider motivation. Two of the six cases were primarily motivated by financial gain (greed); the rest were primarily motivated by a grudge held by the employee against the organization or its members (e.g., an immediate supervisor). Several of the cases involved insiders whose attacks were motivated by a combination of greed and grudge. In particular, the cases in the middle along the horizontal axis involved insiders whose goal was to make money at the company's expense. The cases ranged according to technical sophistication as shown along the vertical axis. While this is a rough characterization, it is clear that attacks involving the construction of logic bombs (C and F) and the exploitation of network protocol race conditions (A) require more technical sophistication than attacks that involve the exploitation of vulnerabilities in procedures (B) or the manual deletion of critical software (D and E).

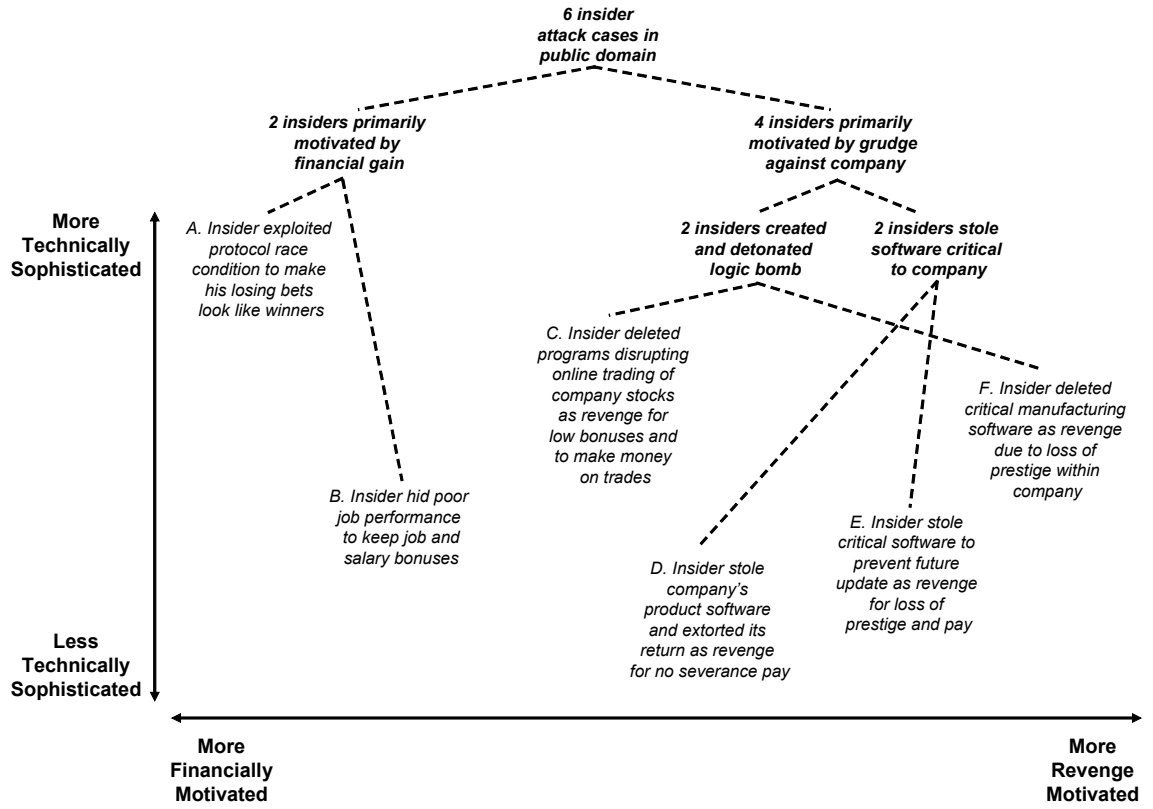

Figure 2 Characterization of Six Insider Threat Cases

The following provides a more detailed description of the above cases:

Case A

An employee (***A***) of a turf accountant (a company that processes horse races bets) committed several illegal actions, each more serious than the previous one, with the help of some accomplices that also began to work for the same company. ***A*** worked as an off-track bet processor and, initially, he started producing copies of unclaimed winning bets and passed them to an accomplice to claim the bets.

After ten months doing this, ***A*** placed a so called "pick four" bet via an accomplice's account, selecting two specific horses for the first two races, and betting on any horse in the last two races. After the first two races had been run, ***A*** manipulated the bet to reflect the actual winner of the first two races. He tried to do the same with a "pick six" bet, selecting four specific horses for the first four races and betting on any horse in the last two. The second time ***A*** tried this operation he got a win of $3M, but due to the unorthodox pattern of the bet, the win was withheld and an investigation began. As a result, ***A*** and his accomplices were arrested and all their winnings were forfeited.

Case B

An employee (***B***) working in the foreign currency department of a bank gained control over the data after performing several preparatory actions, such as eliminating some

monitoring tasks over his work or convincing back office personnel to take deliberately corrupted data from his own personal computer instead from the official Reuters terminal.

B entered some fictitious options into bank records that showed that he had apparently made beneficial operations for the bank. As a result, he was awarded substantial bonuses and was promoted to managing director of foreign currency trading. ***B*** periodically manipulated the data to increment the bank's non-existent profits and obtained increasing flexibility and independence to conduct his transactions. The bank officials became suspicious about the sums being demanded to cover the transactions and discovered that ***B*** had frequently exceeded the counterparty credit limits that the bank had established for foreign exchange trading and that he had recorded nonsensical transactions, such as options that supposedly expired unexercised the same day they were purchased.

Case C
An employee (***C***), who had root access and responsibility for the company's entire computer network, did not receive the bonus he was expecting and consequently decided to attack the firm. ***C*** bought many put option contracts (a put option contract is a type of security that increases in value as a company's stock price drops).

C constructed and executed a logic bomb causing the simultaneous deletion of the programs on servers distributed across the U.S. that allow online trading of the stocks of the company. The logic bomb caused problems during a short period of time. However, the system was recovered. No lowering of the stock values was evident as a result, but it cost over $3M to assess and repair the network.

Case D
A software project manager (***D***) verbally attacked co-workers and some of them left the project. ***D*** resigned after an argument about project progress: He was asked to write a project status report before leaving the firm. Although he apparently accepted this request, instead of complying, ***D*** asked the company for money in exchange for the only copy of the developed software. He had previously deleted all backup copies.

The company never recovered a complete copy of the software and had to spend a significant amount of time and money to reconstruct the product. The employee was arrested.

Case E
After the retirement of his previous supervisor, who had imposed only very lax controls, an employee (***E***) experienced more rigorous supervision of his work. ***E*** was asked to document his software but he failed to comply. The supervisor decided to revert ***E*** to his previous, lower pay status and relocate him. In response, ***E*** deletes the copy of the software from his laptop, quits his job, returns the laptop, and informs his supervisor that the only copy of the software has been permanently lost due to computer malfunction. Although the application could go on running, future updates were no longer feasible.

Months later encrypted copies of the code were recovered from the residence of ***E*** and he was indicted.

Case F
This case is explained more in depth in the next section 2.2.

A summary of the six cases can be seen in Table 1.

Table 1 Summary of the Six Insider Cases

CASE	ACTION	MOTIVE	PRECURSORS	IMPACT	TIME	OPPORTUNITY
A	Modify Data for Own Purposes	Financial Gain	From previous smaller illegal operations to bigger ones.	$210,000	1 year	Privilege access
B	Modify Data for Own Purposes	Financial Gain	-Change of management rules to gain control. -Took operations to the limit - Intimidation	$500 Million (trading losses)	4 years	- Absence of control - Management permission
C	Logic Bomb	Grudge	- Complaints about low bonus projections - Purchase of put option contracts	$3 Million	2 weeks	- Control over the system - Privileged access
D	Code Stealing	Grudge	-Intent to maintain control of the software. -Took control of the system. - Verbal attacks	Software (product) loss	1 year	Entire control of sole copy of software
E	Code Stealing	Grudge	Intent to maintain control of the software	Impossible future updates	2 years	Entire control of sole copy of software
F	Logic Bomb	Grudge	-Work environment discontent. - Took control of the system -Preparatory attacks.	Software loss $10 Million	6 months	Control over the system

2.3 Main Results of the Lloyd/Omega Model

Some of the authors of this paper have previously modelled an insider attack (Melara et al. 2003b). This model reproduced the attack of Tim Lloyd against the firm he worked for (Omega).

Tim Lloyd had worked for Omega for 11 years. As the company expanded into a global enterprise, his prominent position slipped from being one of technical authority into being just a team member. Feeling disrespected, he planted a logic bomb that caused prolonged system downtime, damages and lost contracts.

Before the preparation of the primary attack he had already carried out some incidents that affected proper operation of the information systems and also exhibited his discontent verbally and physically. Because of his behaviour he was reprimanded twice.

Months before the attack Lloyd removed programs from Omega's workstations and centralized them on one file server, telling workers not to store them locally any longer. He began to prepare for his attack, testing it several times before being terminated and finding a new job.

Three weeks after Lloyd had been fired from Omega, a logic bomb triggered by him deleted more than one thousand programs that "ran" the company. The backup tapes for these programs were never found.

Lloyd was convicted on computer sabotage based on evidence consisting of malicious code, three test programs and two old backup copies found on Lloyd's private PC.

The behaviour reference modes for this case are shown in Figure 3:

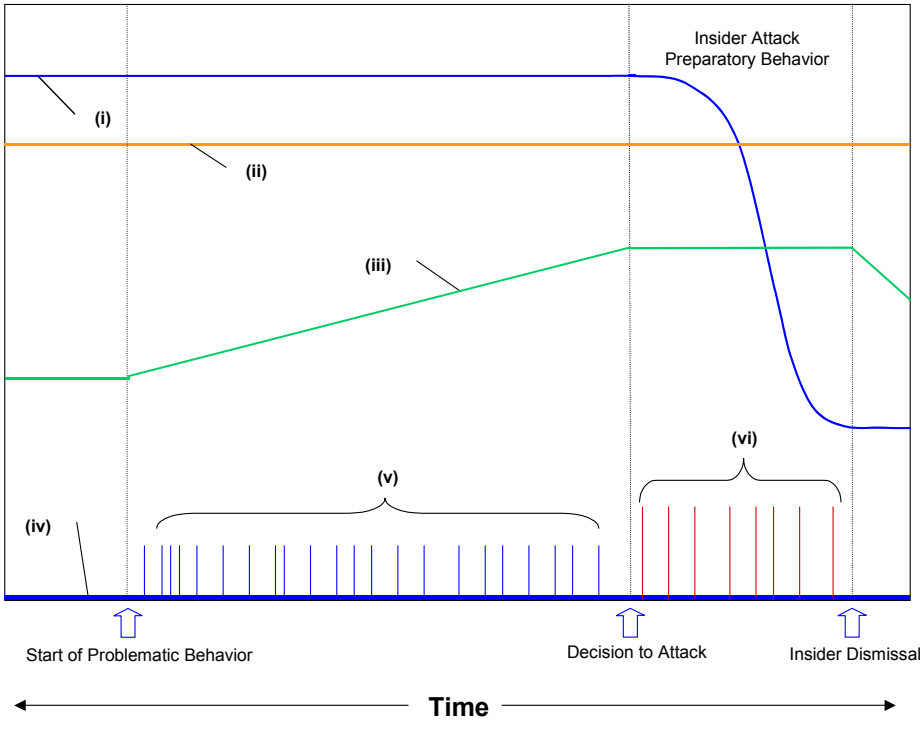

Figure 3 (i) Security Level; (ii) Pressure to Grow; (iii) Workplace Discontent; (iv) Formal Controls; (v) Disruptions of Workplace Climate and Precursor Incidents; (vi) Actions to Reduce Security Level.

Following his demotion from a star employee to an average worker, Tim Lloyd exhibited public actions of discontent. A crucial observation is that management perceived Lloyd's problematic behaviour as a disruption of workplace climate and not at all as a threat to the security of the company. Accordingly, the reference behaviour modes include Tim

Lloyd's disruptions of workplace climate as well as some security incidents that went unnoticed as security threats. Further, the reference behaviour includes management preoccupation with workplace climate and corresponding obliviousness toward the security threat posed by Lloyd. It is likely that the high pressure to grow, which had characterized Omega since 1985, made workplace climate the key aspect of concern for management.

There was an absence of formal policies such as designing correct segregation of security duties or maintaining an appropriate employee-supervisor relationship. Further, there was no evidence that any security audits were conducted.

Tim Lloyd made up his mind to strike some months in advance of the "big attack." His discontent may have triggered his actions to reduce the security level of the system. About a year before he committed the attack, he showed visible signs of discontent, and the failure of management to respond to this behaviour from a security perspective may have encouraged Lloyd to plan his attack. The lack of concern about security enabled Lloyd to act with impunity to make the system more vulnerable months before he committed the attack.

Interestingly, the security level was extremely low at the end of the considered time horizon, i.e. when the attack actually occurred. The security level had decreased significantly during the last months preceding the attack. The severe consequences of the attack support this conclusion.

It should be pointed out that the behaviour of the model is obtained from endogenous variables, with no external inputs. We mention two remarkable feedback loops (Melara et al. 2003b): One of them represents that a low management commitment to security implies low detection activities and that consequently there is a low risk perception, completing a positive feedback loop. The other reinforcing loop includes the insider's decision to attack, the precursor actions, the subsequent downtime, the deterioration of the work climate and the reinforcement of the insider's decision to attack.

3. Group Modelling of Generic Insider Cases

3.1 Introduction

The group modelling process occurred during the Second Annual Workshop on System Dynamics Modelling for Information Security from Tuesday to Thursday 17-19 February 2004. Monday 16 the "problem owners", i.e, the CERT/CC information security experts got a short introduction to causal loop diagramming as well as the basic aspects of stock-and-flow diagrams, the connection between causal structure and behaviour, delays and problem definition in system dynamics. A short version of this course was delivered again Tuesday morning for the benefit of some CERT/CC researchers who could not attend the course the previous day.

Note, however, that instead of a traditional process involving problem (and data owners) one had researchers from the collaborating CERT/CC study of insider threats as proxy for

the actual problem owners. Previous to the workshop, a detailed proposal with references to public available insider attack cases was developed.[8] A PowerPoint presentation of the proposal was shown in the first (plenary) session of the workshop Tuesday morning.

The team working on the generic insider threat problem completed its work using group model building practices (Vennix 1996). As described below, we used a fairly standard definition of modelling team roles (Richardson and Andersen 1995) and used a scripted group modelling process that has been previously described in the literature (Andersen and Richardson 1997). A more complete and "blow-by-blow" description of the details of the process that we used (applied to a similar, but different problem area) is currently in draft form (Luna-Reyes 2004).

3.2 Definition of Modelling Team Roles and Script Development

Although the group modelling process ran for three days, the effective time dedicated to actual modelling was closer to two days. Indeed, each morning there was a plenary session involving all participants of the workshop for cross-fertilization between the insider and outsider thread.

Within the insider threat team, Cappelli and Moore assumed the roles of meeting managers, taking responsibility for making any critical calls with respect to group process and steering the overall team over difficult areas of discussion. During the three days, five to eight members of the CERT/CC & USSS study on insider threat contributed with domain expertise. Stanton from Syracuse University added his expertise on organizational psychology and the role of trust.

Andersen fulfilled the role of group facilitator and held that role for most of the meeting time. The modelling team of Mojtahedzadeh, Weaver, Zagonel and Sarriegui split a number of roles and responsibilities. Weaver, with her background and training as a psychologist, focused on the emergence of psychological and behavioural mechanisms within the group discussion (she split her time between the insider and outsider threat groups). Zagonel had considerable experience working with Andersen creating group products on a rapid schedule; he had primary responsibility for capturing all group products in electronic form and has produced most of the figures and products shown in this section of the paper. Mojtahedzadeh concentrated on the emerging feedback structure being articulated by the group. Sarriegui, who had no previous experience from group modelling, but had previously collaborated in modelling an insider attack, acted more in the background and contributed to the quality assurance process.

An important feature of our modelling team was that we came into the exercise without a group process coach and indeed without a detailed script for the three days of work. Since we had not previously come together with the group or the meeting managers, it was not possible to have this advance work done. We tried to "turn this bug into a feature" by

[8] The proposal (see http://www.cert.org/research/sdmis/) was a collaborative involving SEI/CERT: Chris Bateman, Dawn Cappelli, Casey Dunlevy, Andrew Moore, Dave Mundie, Stephanie Rogers, Tim Shimeall; TECNUN, University of Navarra: Jose Maria Sarriegui; from Syracuse University: Jeffrey M. Stanton; and Agder University College: Jose J. Gonzalez.

inventing a new group modelling script. We convened the modelling team plus the meeting managers in a "fishbowl" exercise to plan the next steps in the modelling process. This provided the modelling team an opportunity to air its plans and concerns for the meeting and to involve the meeting managers more directly in the planning process. This process also helped the full team gain a sense of "what was coming next". We believe that this ad hoc script worked quite well and will try to refine and repeat it again.

3.3 Stakeholder Analysis

The formal group modelling session began with a series of problem defining exercises. The purposes of these exercises was to begin to define the problem that the group wanted to investigate and hence the boundary of the model to be constructed. The first of these was the creation of a stakeholder map as shown in Figure 4.

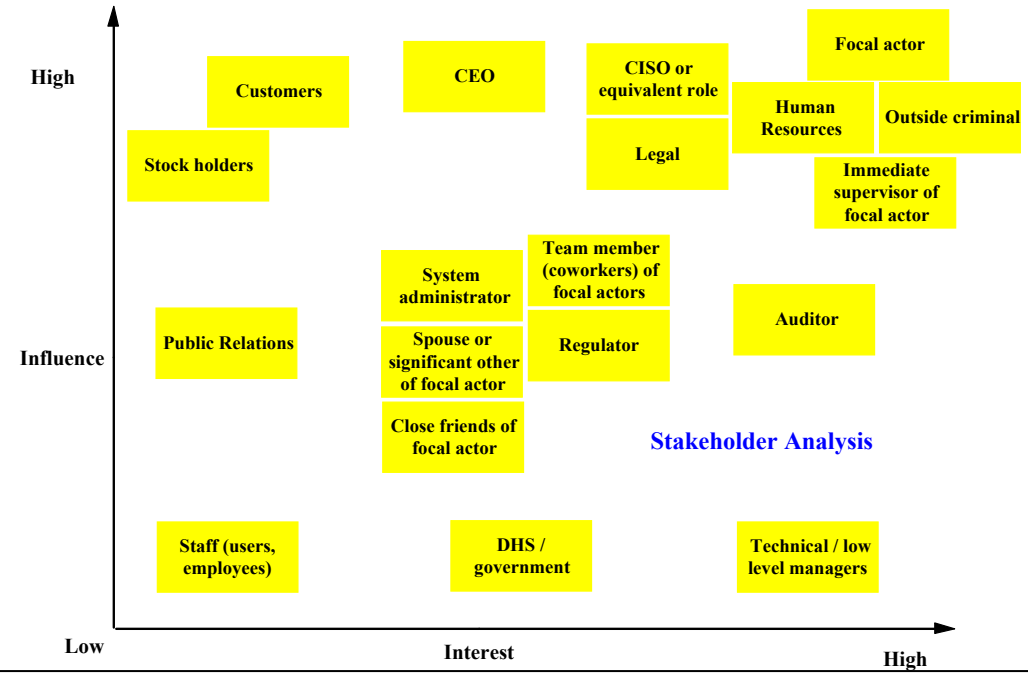

Figure 4 Stakeholder map created on Tuesday (the first of three days) afternoon

For this exercise, participants wrote down the name of key stakeholders whose actions would have a large influence on the behaviour of the model to be constructed. Participants were asked to place these named stakeholders on a two dimensional grid characterized by "Interest" in the problem at hand on the horizontal axis versus ability to "Influence" the problem on the vertical axis. Stakeholders located in the upper right hand corner of the diagram were the key stakeholders for this model, being persons both very interested in and very able to shape the outcome of the problem of interest.

The discussion was lively and lasted for about a half-hour. The group seemed most interested in making some actors (such as the focal actor-attacker) less influential and

making other actors more interested or influential. This quick and dirty form of a stakeholder analysis and management (Eden and Ackermann 1998) revealed that the group was keenly interested in the organizational, behavioural, and psychological states of mind of the focal actors and those line and technical managers responsible for the organization's cyber-defences. This initial half-hour discussion anchored the rest of the three-day meeting.

3.4 Policy Levers and Clusters

The second exercise, also completed on Tuesday afternoon, was a mapping and clustering of the policy options that the group believed to be available to solve the problem of interest. Note that the group had at this point not yet precisely defined what the problem was. Rather the group was beginning to define the problem indirectly by addressing a number of key questions related to the problem (e.g., who is involved in the problem—stakeholder map or what can be done about it—policy levers map). Members of the group posted the policy levers in similar clusters as they were put up for discussion.

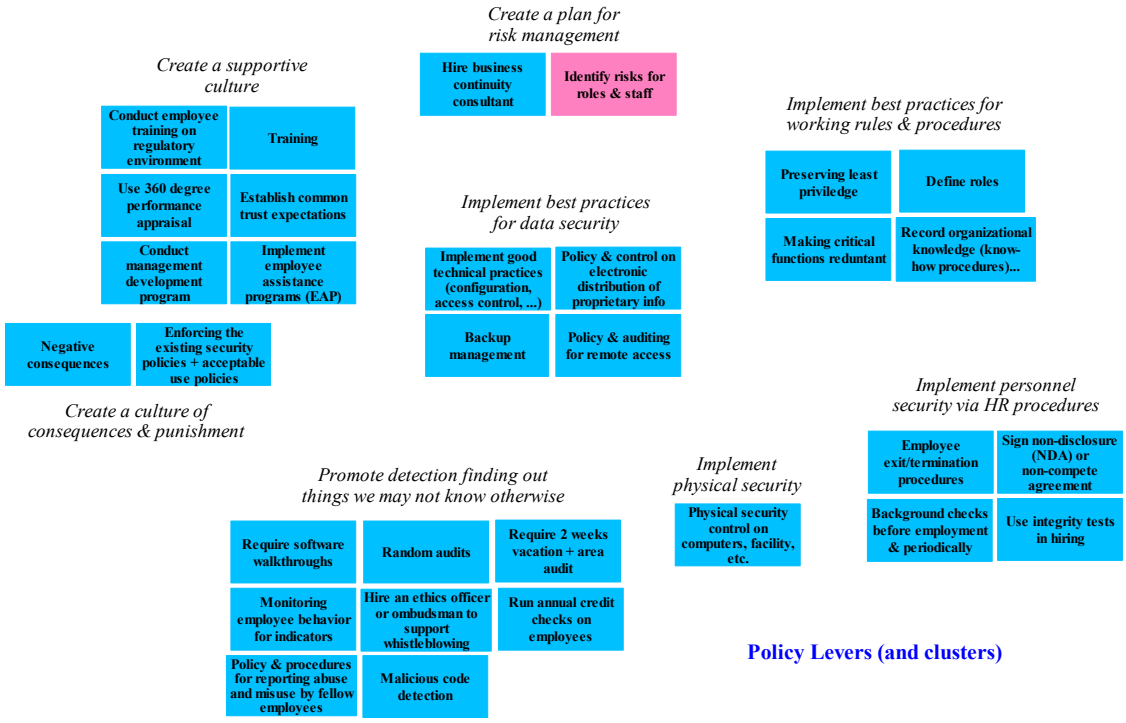

Figure 5 Map of policy levers organized by clusters

A careful analysis of the clusters identified in Figure 5 compared to the first causal structure shown in Figure I-1 in Appendix I reveals that information about the key stocks and structures of the eventual system map were implied by this exercise. This exercise stressed detection capability (ultimately a stock) and a supportive and trusting organizational culture (a key stock in the final map). In addition, the figure implies issues of formal risk management planning (which finally showed up as a key loop) and several types of "best practice" (which ultimately emerged as violations of best practice – things

that could have or should have been detected by the organization). Physical and technical security issues were also mentioned in this final map.

3.5 Reference Modes

The remainder of the afternoon devoted to problem and model boundary definition was devoted to a reference mode exercise. Working alone or in pairs, group participants were asked to sketch one or more key variables that seemed to them to illustrate the essential nature of the problem under study. For each variable or set of variables, participants were asked to specify the relevant time boundary and then to directly sketch the behaviour of interest. All sketches were posted on a board and discussed. In all, eighteen variable sketches were produced as listed in Figure 6 below.

■ Employee/management tension	■ Management attention
■ Trust in employee	■ Staff trust
■ Respect for insider	■ Awareness of risks
■ Behavioral oversight	■ Awareness
■ Position in company	■ Average job overload
■ Preventive HR procedures	■ Financial health of organization
■ Focal actor job satisfaction	■ Management of technology/data
■ Turnover of critical employees	■ Percent of shared organizational knowledge
■ Mistreating of fellow employees	■ Employee access upon termination

Figure 6 Titles of eighteen sketches elicited in the reference mode exercise

The reference mode exercise provided a wealth of dynamic insight, all of which has not yet been mined to the benefit of the full project. Our intention at the end of the day on Tuesday had been to return to these eighteen reference mode sketches and to cluster them and to discuss them more fully. A more complete analysis of the reference mode data is a task for future modelling. To illustrate some of the interesting dynamics emerging from this exercise, two sample sketches are presented below:

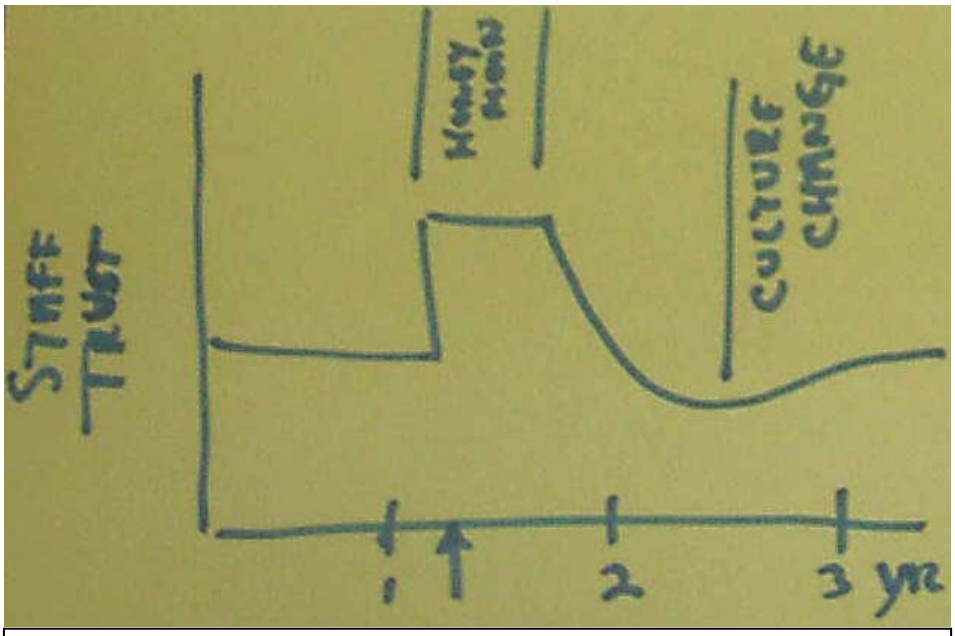

Figure 7 Sample reference mode sketch; Staff trust vs. time (3-year time boundary)

The sketch of staff trust versus time is one of the eighteen reference mode sketches, reproduced here as a photographic reproduction of the participant's own hand drawn version. The sketch illustrates this participant's view that a "honeymoon" period of trust would be followed by a sharp decline in trust accompanied by a cultural change. These dynamics are presumed to occur over a three-year time horizon. We selected this particular sketch to discuss because it illustrates key aspects of a shifting culture of trust, ultimately depicted in the "dynamic trigger hypothesis" as discussed below.

The sketch of preventive human resource procedures (Figure 8) is over a much longer time frame than the trust sketch shown in Figure 7. Figure 8 illustrates the participant's belief that organizations would over time systematically raise and then lower their organizational defences in the face of repeated cyber-attacks. The vertical dashed line in the hand drawn figure represents the time of a recognized attack, indicating that this participant believes that a culture of prevention will suddenly step up in a repeated saw tooth pattern following attacks. This is the classical reference mode of compliance erosion known in the literature as the "unrocked boat behaviour" (for a discussion of this reference mode and a theory of erosion of compliance cf. Gonzalez and Sawicka 2003).

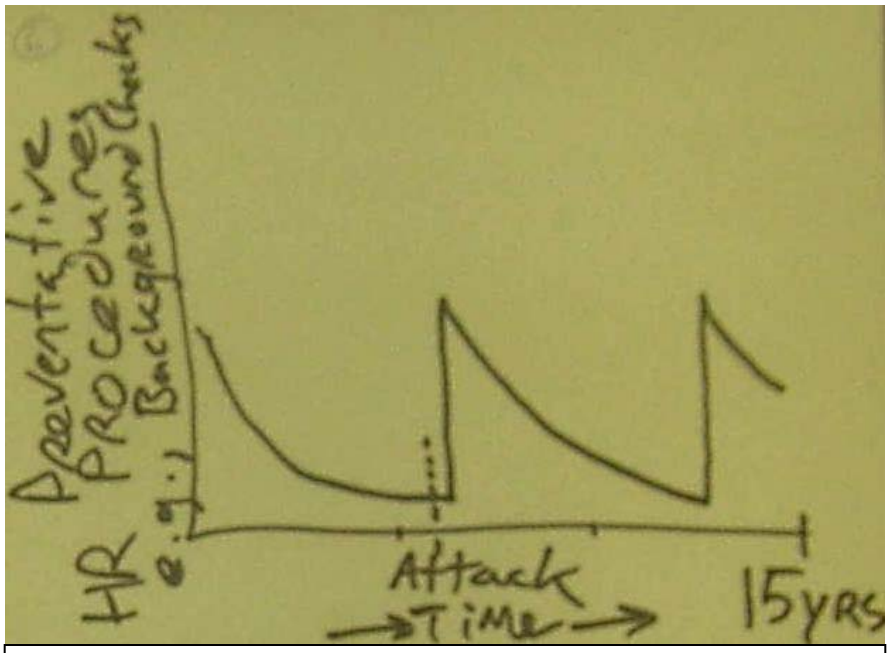

Figure 8 Sample reference model sketch: Preventive human resource procedures vs. time (15-year time boundary)

3.6 Causal-Loop Diagram of the Insider Threat Problem

By the end of Tuesday night, the group had accomplished a substantial amount of work. As we closed the day, our expectation was that we would return to the reference mode exercise and spend more time extracting "dynamic stories" from the group by clustering the raw reference mode sketches. Our actual path differed from this plan.

3.6.1 Laying Down the Basic Stock and Flow Structure (Tuesday night and Wednesday morning)

Returning from dinner on Tuesday night, the modelling team retired to a private session to review progress for that day and to plan for the next day. We were fortunate in that the recorder had been working hard (indeed he had skipped dinner) and we had available to us an unusually clear set of products from the day's work (essentially all of the products as they are presented in this paper were available after dinner).

The modelling team decided to take a risk the next morning with the group. After some vigorous discussion, the team proposed a set of stock variables and a geometry for their placement on a large board that would support a causal loop elicitation exercise the next morning. Key stocks that seemed to "leap off the page" from the stakeholder and policy cluster analysis were stocks related to management attention, an aggregate measure of security procedures in place, an accumulation associated with supportive culture, and something associated with attacker risk and willingness to attack. A key invention of the group was a sort of aging chain whereby violations of security procedures seemed to "grow into" precursor events, eventually maturing into full-fledged attack events. We

were not sure if these were independent events or linked into a chain (if it were a chain, what would be the conserved units associated with that chain?).

3.6.2 Mapping Learning from Experience, Audits, and Detection (Wednesday morning)

We began Wednesday morning (the second day of group modelling) by proposing the stock and flow structure first articulated by the modelling team the previous night. The first portion of the mapping exercise consisted of an extensive discussion of and agreement on the basic stock-and-flow structure proposed by the modelling team on Tuesday evening. The group had difficulty with the quasi stock-and-flow chain that eventually became violations of best practice, precursor events, and actual attacks. But the group did give the modelling team permission to proceed with the causal mapping exercise.

We decided to work with three views of the causal map, all linked through the anchor points of a common set of stocks. A large white board of cling sheets measuring approximately five feet by six feet was constructed on the front wall of the room. This "scratch wall" was used as a boundary object to capture initial discussions from the group. Our plan was to capture some portion of the causal structure and then to carefully transcribe that "layer" of structure to a larger cling sheet wall (about seven feet by twelve feet) on a different wall. Each layer would be carefully drawn by the modelling team, allowing the group to "transfer insight from and approve" the scratch to finished wall models before the scratch wall was erased for a second round of mapping. This exercise would work well if the basic anchoring stock structure did not change too much between layers, and in this we were fortunate. While the names of the anchoring stocks changed quite a bit, the basic geometry survived two days of intensive mapping. The third view of the model was a Vensim sketch that was based directly on the carefully modeller-drawn side wall. The figures shown in this paper are those Vensim sketches as they were presented to the full team on Thursday.

Figure I-1 in Appendix I shows the first layer of structure added to the positive map. This layer began with a series of negative loops illustrating how attacks could lead to increased managerial attention and hence investment in detection (that would reduce future attacks). A reinforcing loop that appeared to be working as a trap connected detection capability to management's perception of violations of best practice. This loop could act over time to suppress appropriate levels of investment in threat detection capability.

Finally, the group posited that organizations could set appropriate levels of threat deterrence and detection capability by conducting audits and assessments of organizational vulnerabilities. For example, "red teams" could simulate attacks on organizational information assets, helping management to learn about vulnerabilities without attacks and without having detection systems locate potential vulnerabilities or breaches of good practice.

3.6.3 Mapping Growth of Motive (Wednesday afternoon)

The second layer in the causal map focused on the growth of motive to attack by the focal actor. These effects are shown in red in Figure I-2 in Appendix I. Four stocks in Figure I-2 are shaded in, indicating to the group a "focal actor attack" sector of the model that contained variables and loops pertaining to the psychological and behavioural dynamics of an attacker. These dynamics are spanned by stocks that relate to motive, perceptions of risk, the creation and monitoring of precursor events, and finally attack behaviours themselves.

The motive structures are necessary to trigger any of the other attack behaviours. The group decided to view motive as a more or less undifferentiated event, not trying to sort out various types of motive. Aspects of a supportive organizational culture were presumed to mitigate attack motives.

Some of the most interesting loops in the mapping exercise arose from examinations of how the focal attacking actor dealt with perceived risk. The team seemed clear that attacking actors were quite careful and tended not to take undue risks. Rather, they launched so-called "precursor events" to probe organizational defences. Forensic analysis of attacks in the six cases uncovered after-the-fact evidence that the attacker had been probing organizational defences (without detection) for some before an actual attack was launched. Indeed, several participants mentioned that attacks themselves could be repeated once attackers were sufficiently emboldened (i.e., believing that risk of detection was quite low.) The dynamics shown in Figure I-2 laid the foundation for much of the "dynamic trigger hypothesis" as discussed below.

3.6.4 Mapping Trust and Deterrence (Thursday morning)

The final layer of the causal map explored the myriad linkages around the issues of trust and deterrence that knit the other pieces of the model together. This mapping was completed on Thursday morning and is shown in Figure I-3 in Appendix I as the final layer mapped in blue. This final layer focused on organizational trust as a central variable in the overall cyber-attack scenario. Organizational managers strive for a culture of support and trust for a number of excellent managerial reasons. Indeed, the participants believed that a supportive and trusting environment could serve to suppress motive to attack. However, in the presence of a motive to attack, trust can reinforce a tendency to under-invest in detection procedures, thereby opening up the organization to undetected precursor events and eventually full-scale cyber-attacks.

Hence organizational policies designed to engender organizational trust and a supportive culture can have the unintended effects of making an organizational more vulnerable to cyber attacks mediated by a string of undetected precursor events. This is exactly the pattern found in the forensic data available from publicly documented cyber attacks.

3.7 Hypothesizing Dynamic Mechanisms

By Thursday evening, the workshop's full causal map was available to the modelling team. Working "in the back room" (i.e., not with the full group), the modelling team extracted a set of three key reinforcing loops that seemed to be a key focus of the whole mapping exercise. These three loops are already present in the full map shown in Figure I-3 in Appendix 1 but their possible importance is masked by the over 4,000 other feedback loops that contribute to the visual complexity of that figure. Extracting a small number of loops simplified and focused the discussion.

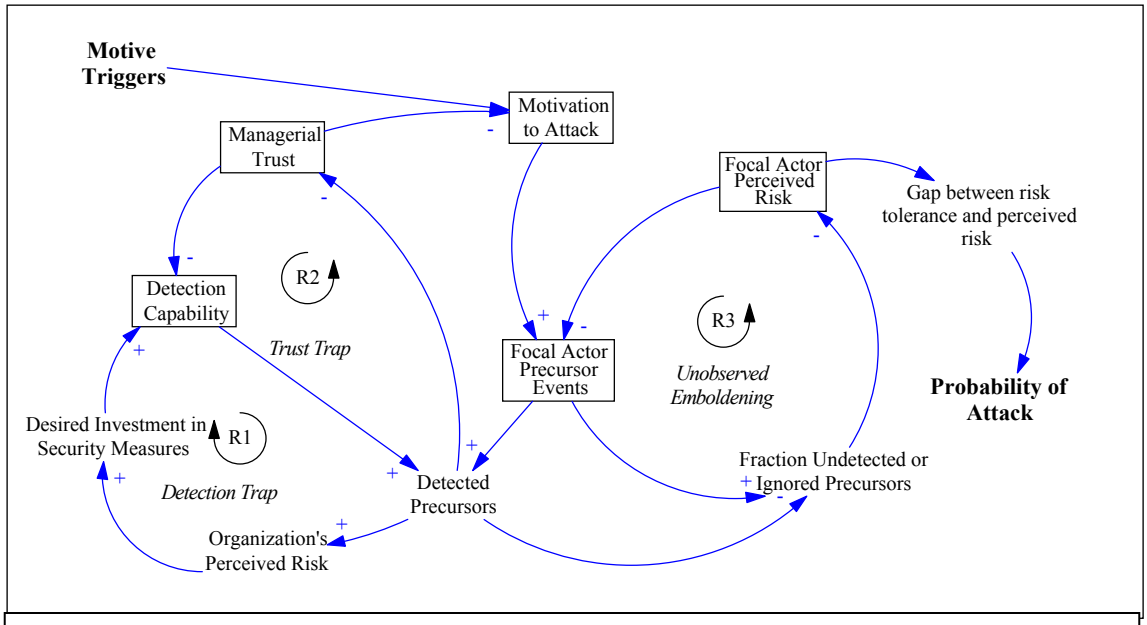

Figure 9 Overview of core mechanisms underpinning the Dynamic Trigger Hypothesis

Figure 9 presents the extract from the back room exercise that illustrates a reduced set of feedback mechanisms that combine to create the "dynamic trigger hypothesis". This dynamic trigger mediates between an exogenous motive trigger (upper left quadrant) to attack and a dynamically escalating probability of attack (lower right).

The interaction of three key feedback loops shown in Figure 9 show how the focal actor's behaviour prior to the attack (precursor events) combined with unintended consequences of explicit managerial actions may lead to the focal actor's dynamically decreasing perceptions of risk and increasing motivation to execute the ultimate attack.

A series of three connected verbal stories were crafted around Figure Figure 9 to express in verbal terms the overall dynamic hypothesis emerging from the workshop. These three stories-as-dynamic-hypotheses are presented below:

- *H1—Detection Trap.* ***The absence of detection capability suppresses detection of on-going violations and precursor events thereby***

suppressing desired investments in security measures (such as detection capability).

The detection trap contributes to an explanation of why organizations that have been the victims of cyber attacks have historically under-invested in accepted cyber security practices. Over periods when these organizations have not been victims of attacks, they have let down their defences either by not investing in the first place or by continuing to invest in detection capabilities.

- *H2—Trust Trap*. ***Over time, trust can disable an organization's compromise detection capability, leading to fewer detected precursor events and increasing violations of best practices. In turn, fewer detected events can reinforce the (perhaps erroneous) conclusion that compromise detection is not needed and can increase managerial perception that high trust in employees is warranted.***

Managers in all types of organizations strive to implement good management practices by creating supportive organizational cultures. Indeed such supportive cultures can reduce employees' motivation to attack and can diffuse such motives when they do arise through employee counselling and other support program. In addition, the existence of high levels of managerial trust can enhance overall productivity and reduce transaction costs across the organization. A possible unintended consequence of high levels of organizational trust appears in the figure above, however. The trust trap mechanism may help to explain how well intentioned organizational activities can erode an organization's defences to cyber attacks.

- *H3—Unobserved Emboldening.* ***Left undetected, precursor events reduce an actor's perception of risk. In turn, reduced perceptions of risk lead to additional precursor events. This reinforcing cycle of emboldening can remain unobserved by management (absent detection of precursor events-—see H1—Detection Trap and H2—Trust Trap).***

The third causal mechanism, unobserved emboldening, completes a feedback-rich causal pathway between initial motive (shown in the upper left corner of Figure 9) and a dynamically escalating probability of attack (shown in the lower right portion of Figure 9). This dynamic mechanism presents a causal structure that explains, in part, how focal actors often initiate a series of precursor events that probe organizational defences. Since these are reinforcing feedback effects, in the presence of appropriate motives focal actors may drive down their perceived risk until the gap between risk tolerance and perceived risk is low enough to create conditions highly conducive to an attack. In signal detection theoretic terms, these lower perceived risks may be linked to a dynamically changing threshold to act. Taken as a whole, these dynamic trigger hypotheses seek to explain how motives and conditions conducive to attack escalate into actual attack behaviour. The hypotheses describe feedback mechanisms that explain how attackers mitigate their personal level of perceived risk, holding off on final attack behaviours until perceived

risks fall to a level the attacker is willing to accept. While these hypotheses may be appealing for practical reasons they are nonetheless only hypotheses and must be operationalized within a framework of overall organizational and attacker behaviour, tested in an initial proof-of-concept study, and subjected to empirical testing and refinement using both qualitative and quantitative data sources. We believe that the appropriate starting point for these activities lies in a clear understanding of organizationally relevant theory and research. In the next section we present an overview of research on organizational trust as (partial) grounding for the proposed program of research.

3.8 Theoretical Bases of Organizational Trust

Organizational researchers have found that trust serves a foundation for good interpersonal relationships (Rempel, Holmes, and Zanna 1985), cooperative social behaviour (Barnard 1938; Blau 1964), and reduction of social transaction costs (Jones 1984). Trust also appears to play an important role in the functioning of hierarchical authority relationships such as the relationships between supervisors and employees (Tyler and Lind 1992). The extent to which trust exists in hierarchical relationships seems to influence numerous psychological and organizational issues such as the extent of formal controls, degree of cooperation, supervisory spans of control, and the quality of labour-management relations (Eisenhardt 1989; Jones 1984).

Because of the importance of trust, numerous efforts have explored the central role of trust on the attitudes and behaviours of individuals in organizations (e.g., Lewicki and Bunker 1996; Whitner, Brodt, Korsgaard, and Werner 1998). At least two major perspectives have been used to frame the analysis of trust in a hierarchical relationship: a social exchange perspective and an agency perspective. A social exchange perspective suggests that supervisors who wish to develop trust with those they supervise engage in prosocial behaviour to engender reciprocity and positive affect among their workers (Blau 1964; Whitner et al. 1998). From this perspective, supervisors attempt to ingratiate themselves with key workers in order to reap an expected reciprocal benefit in the form of enhanced effort or performance on the part of that worker. One can see the seeds of dynamic Hypothesis 2 in these actions: In a move to engender trust, supervisors may appease workers whose efforts are critical to a project. Methods for this include relaxing formal controls over the employee, expanding the employee's privileges or perquisites, or diminishing the closeness of supervision over that employee. In support of this point, research has suggested that reducing the closeness of supervision increases employees' feelings of control and reduces employees' perceptions of stress (Aiello and Kolb 1995; Carayon 1994; Stanton and Barnes-Farrell 1996). For the typical employee, increasing perceptions of control and decreasing the experience of stress are desirable goals, but with respect to an attacker, these same goals may help to lead to the emboldening described in dynamic Hypothesis 3. Further, any combination of relaxing controls, increasing privileges, or reducing closeness of supervision may also decrease supervisory capabilities for subsequent detection of problematic employee activities (Flamholtz 1979; Mintzberg 1973)

Using a different perspective on trust, agency theorists propose that trust is a required element of a contractual arrangement between a principal and an agent only in circumstances where close monitoring of the agent's behaviour is not possible or cost effective (Eisenhardt 1989). From the agency perspective, when it is economically and logistically feasible, close and accurate monitoring is always a preferable alternative to trust. Agency theory thus balances the expense and inconvenience of close monitoring against the risk of relying upon trust as an alternative to monitoring. Note how this balancing reflects, in part, the dynamic trigger described in Hypothesis 1. Monitoring mechanisms are not cost free, but the evidence of malfeasance that managers might accept as a basis for investing in monitoring capabilities is less likely to be available in the absence of existing monitoring capabilities.

The foregoing discussion provides a brief sample of the existing array of psychologically grounded organizational theory that our team can draw upon to help produce further advances in this research. We can also use the available empirical research examining those theories as a cross check with the system dynamics models we create.

4. Discussion

Five participating institutions at the Second Annual Workshop on System Dynamics Modelling for Information Security, viz. CERT/CC at Carnegie Mellon, University at Albany, Syracuse University (all US) and the European partners Agder University College and TECNUN/Universidad de Navarra, decided to formulate a research agenda and project proposals based on such agenda. The first outcome is a project application entitled "Improving Organizational Security and Survivability by Suppression of Dynamic Triggers" that has been submitted to the National Science Foundation.

4.1 Can System Dynamics Help Improve Cyber Data Availability?

Experience teaches that security and safety failures virtually always have numerous precursor incidents: For every flight crash there are tens or hundreds of near-crashes; the famous software time bomb at Omega was preceded by many indications that the malicious insider intended to attack (Gaudin 2000; Melara et al. 2003b); the 9-11 terrorist strike in 2001 had a forerunner in the bomb attack in the World Trade Center in 1993 and many other precursor incidents that were not perceived for what they were (Emerson 2002).

Schneier (2000, p. 392) argues passionately that cyber attacks need to be publicized – implying collecting and sharing data: «We need to publicize attacks. We need to publicly understand why systems fail. We need to share information about security breaches: causes, vulnerabilities, effects, methodologies. Secrecy only aids the attackers.»

The logic is compelling: «When a DC-10 falls out of the sky, everyone knows it. There are investigations and reports, and eventually people learn from these accidents. You can go to the Air Safety Reporting System and read the detailed reports of tens of thousands of accidents and *near-accidents* [our emphasis] since 1975.» (Schneier 2000, p. 391). Turning around the argument: Cyber security would be much better if a better reporting system were in place.

But for cyber security the task is not easy (cf. the discussion in **§1.3 Cyber Data Restrictions**). Companies that go public are not rewarded. Quoting Schneier again (2000, p. 391-392): «When Citibank lost $12 million to a Russian hacker in 1995, it announced that the bank had been hacked into and instituted new and more profound security measures to prevent such attacks from occurring in the future. Even so, millions of dollars were withdrawn by people who believed their funds were vulnerable immediately after Citibank's announcement. Ultimately, Citibank recovered, but the lesson to Citibank was clear and unambiguous: "Don't publicize."»

How can system dynamics help improve cyber data availability? For one thing, system dynamics modelling does not require incident-specific data about security breaches, but rather aggregated data and stocks and flows of quite an abstract nature (see e.g. the variables in Figure I-3). Therefore, we hope to build partnerships with potential owners of cyber data willing to share such "innocuous" (i.e. untraceable and non-sensitive) data. In other words: Data owners might increasingly trust system dynamics modellers if emergent collaborations demonstrate that such exercise is feasible – and useful

Indeed, one of the nice things about system dynamics modelling is the common experience that even models based on "poor" data can be helpful if expert judgement combined with whatever data that is available contributes to identify important causal structures. The resulting models are useful in the sense that they provide reasonable explanations for system behaviour. Further, experience shows that such preliminary system dynamics models provide valuable suggestions for additional data mining. Expecting this to the case for our emerging collaboration, we gamble that demonstrable usefulness (plus model-based suggestions for what kind of data is most urgently needed) might trigger an iterative process of improved data collection and modelling.

4.2 The Dynamic Trigger Hypothesis Revisited

Keeping in mind that numerous precursor incidents anticipate the actual attack, it is not unreasonable to postulate a causal link between the chain of determinants and outcomes of the precursor incidents and the attack. Such link is hard, if not impossible to recognize, if precursor incidents remain "incidents" – i.e. scattered and rare events against a noisy backdrop. Accordingly, one would expect that the commonality in the incidents becomes salient once their pattern is seen in the light of feedback structures. Thus far, we are simply claiming that identifying feedbacks responsible for the pattern of precursors mounting up before the actual strike would make precursors more conspicuous and, hence, improve the chance to prevent the strike, or at least to mitigate its consequences.

The "dynamic trigger hypothesis" goes farther, in that it postulates that there are dynamic mechanisms unfolding along with the precursor chain. Such dynamic trigger is assumed to mediate between an exogenous motive trigger to attack and a dynamically escalating probability of attack. In addition to the preliminary analysis given above (**§3.7 Hypothesizing Dynamic Mechanisms**) we mention the system dynamics analysis of the Lloyd/Omega case (Melara et al. 2003b). For short, we refer the reader to the discussion of its main results in **§2.3 Main Results of the Lloyd/Omega Model**, especially to the

concluding paragraph describing two unchecked[9] reinforcing feedback loops that fuelled precursor actions and presumably were decisive for Lloyd's final attack.

4.3 Perspectives for Further Research

We are developing a long-term research perspective articulated by the Project on System Dynamics Modeling for Information Security.[10] As illustrated in Figure 10, the project's long term research agenda envisions a range of activities beginning with system level mapping (an exercise begun at the February 2004 workshop) and ultimately creating better security practices in governmental and private organizations. Between these modest beginnings and the grand overall goals lies a series of focused model building exercises, extensive model calibration, qualitative and quantitative data collection efforts, and the design and creation of microworlds, interactive learning environments and decision support products based on the models and policy insights gained from the modeling efforts.

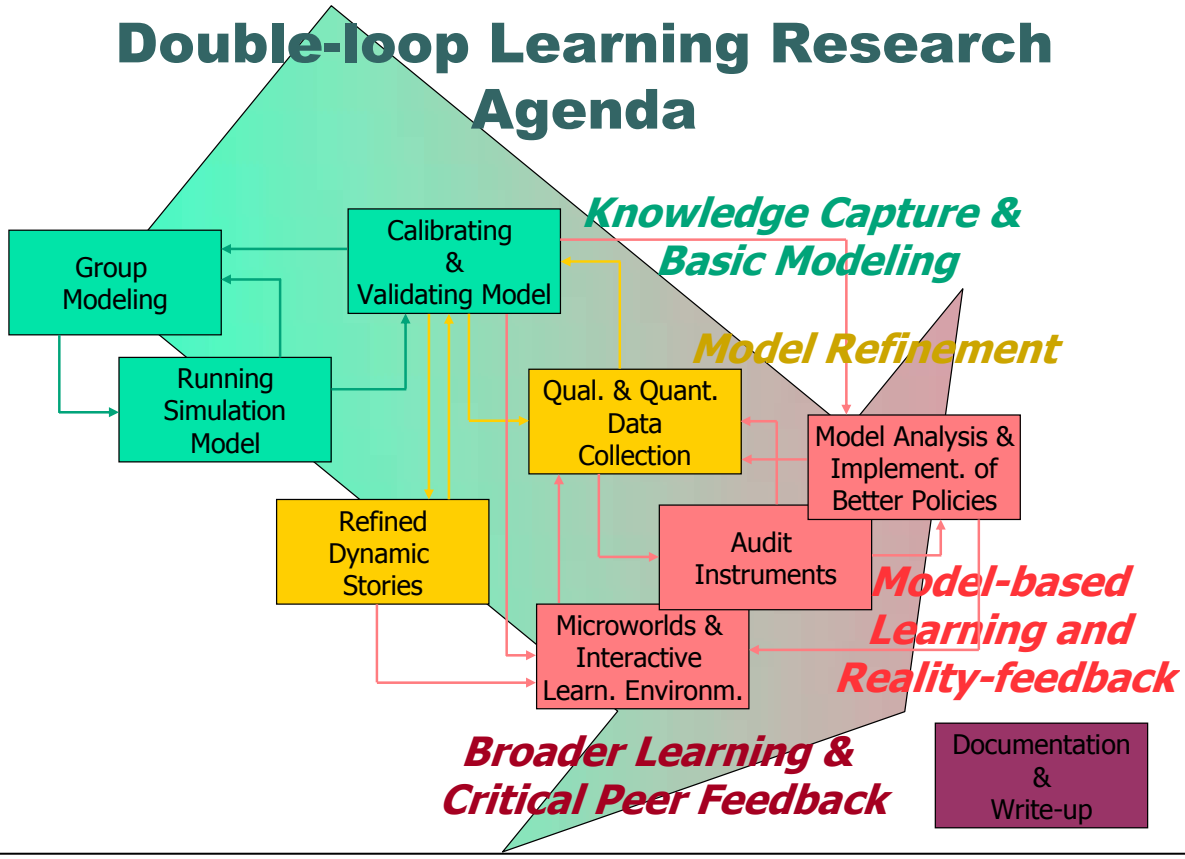

Figure 10 : Overview of Research Products and Tasks Intended to Link System Modeling Efforts to Improved Cyber-Security in Government and Private Organizations

[9] Unchecked by the *management*, who partly overlooked, partly misinterpreted what was going on.
[10] This is still a working title.

The proposed "double-loop learning research agenda" should be seen as three stages that become active as the project proceeds. There are iterations *within* each stage and added iterations *across* stages. The lowest stage (Knowledge Capturing & Basic Modeling) is identical to Phase One of the project. In Phase Two, both the lowest stage and the second stage (Model Refinement) are active, with interactions and interactions within and across the two stages. The highest stage (Model-base Learning and Reality-Feedback) is reached in Phase Three – now there are iterations within the three stages and across all stages.

Phase One, "Knowledge Capturing & Basic Modeling", involves Group Modeling, Running Simulation Model and Calibrating & Validating Model. Group model building, a well-established discipline within System Dynamics (Andersen and Richardson 1997; Vennix 1996; Vennix, Akkermans, and Rouwette 1996), elicits relevant knowledge from domain experts through group processes. Proceeding through several stages, going from qualitative system maps to "first cut" calibrated and validated models, a basic simulation model is developed.

In Phase Two, the basic simulation model from Phase One is further enriched with structure and data (Qualitative and Quantitative Data Collection) derived from psychological and social sciences (such as dynamic decision theory and signal decision theory). Insights from the model are articulated (Refined Dynamic Stories). The result is a more mature simulation model that will serve as platform for the next phase.

Phase Three applies the obtained model and model insights to derive Microworlds & Interactive Learning Environments for organizational learning (mainly – but also for long-term improvement of model structure and model calibration), Audit Instruments (for increased organizational consciousness and awareness– but also for long-term improvement of model structure and model calibration) and Model Analysis and Implementation of Better Policies.

The proposed approach recognizes that cyber attack data is highly fragmented, quite incomplete and subject to severe restriction between incident registration and analysis agencies (e.g., CERT/CC) and reporting organizations. Group model building operates on expert judgment to create models reflecting expert's understanding of problem structure as well as valid parameter ranges. As the research project unfolds, more domain experts contribute and audiences for audit processes and interactive learning environments play a dual role as users of tools and feedback agents (adding more verification and validation to the system dynamics model).

4.4 Rounding Up

There is growing consensus that, in order to be cost-effective, organizational priorities around information security concerns must emphasize those that enable maintaining the organization's essential services (at least at some degraded level) despite malicious attacks (Ellison et al. 1999; Anderson 2001; Knight, Strunk, and Sullivan 2003). This emphasis forces an expansion of security issues from a narrow technical specialty to an organization-wide risk management concern that must deal with broad avenues of attack and the motivations of both attackers and defenders. A focus on maintenance of essential

services requires a systemic approach that takes into account the whole range of organizational policies, practices, procedures, and technologies that may contribute to the occurrence of security incidents. Even the organization's workplace culture needs to be considered, including the values, beliefs, and behaviours of employees that shape the way that they conduct their jobs.

As an example of the limitations of existing technology, intrusion detection systems can address only a small part of the problem, at least in its current form. Existing intrusion detection technology targets the identification of only computer- and network-based attacks. Security compromises by users that are abusing their legitimate authority – a characteristic of insider attacks by definition – do not involve events designed to be detected by the vast majority of intrusion detection tools available. Anomaly detection tools that monitor individual applications for user activity that deviates significantly from a predefined profile may be useful, but these tools are known to be expensive to operate, only minimally effective, and not widely available. In addition, attacks that "fly over the radar" of intrusion detection technology - such as exploitations of vulnerabilities in procedures, physical structures, or personnel - need to be taken as seriously as technological attacks (Anderson 2001). Disregarding these factors could be very misleading and result in large gaps in our system defences.

Without a view to maintaining essential services, an organization may waste much time and resources attempting to detect and analyze attacks that have no impact on their ability to succeed. A report on the state of the practice of intrusion detection technologies recommends that, among other things, future technologies should integrate a more diverse source of attack data to ameliorate inaccuracies, defend against attacks that are more sophisticated than those of the average hacker, and integrate human analysis as part of event diagnosis (Allen et al. 1999).

We agree with these recommendations, but suggest taking them a step further to deal directly with the inherent limitations of a strictly technological approach. Organizations should focus on intrusion detection and response holistically by integrating a comprehensive intrusion detection and response capability with an organization's policies and procedures, as well as with the technology. The system dynamics model described in this paper identifies deterministic, continuous feedback processes for intrusion detection in the large. We expect this approach to improve the measurability of an organization's survival over time, in comparison with an approach that uses stochastic models of risk based on random event logic.

References

Aiello, J.R., and K.J. Kolb. 1995. Electronic performance monitoring and social context: impact on productivity and stress. *Journal of Applied Psychology* 80:339-353.

Allen, J., A. Christie, W. Fithen, J. McHugh, J. Pickel, and E. Stoner. 1999. State of the Practice of Intrusion Detection Technologies. Pittsburgh, PA: Software Engineering Institute, Carnegie Mellon University.

Andersen, David F., and George P. Richardson. 1997. Scripts for group model building. *System Dynamics Review* 13 (2):107-129.

Anderson, Ross. 2001. *Security Engineering: A Comprehensive Guide to Building Dependable Distributed Systems*: John Wiley & Sons.
Barnard, C. 1938. *The functions of the executive*. Cambridge, MA: Harvard University Press.
Blau, P. M. 1964. *Exchange and power in social life*. New York: Wiley.
Carayon, P. 1994. Effects of electronic performance monitoring on job design and worker stress: Results of two studies. *International Journal of Human-Computer Interaction* 6:177-190.
CERT/CC. 2004. *CERT/CC Statistics 1988-2003* 2004 [cited February 6 2004].
Cooke, David L. 2003a. Learning from Incidents. Proceedings of the 21st International Conference of the System Dynamics Society, at New York, NY, USA.
———. 2003b. Learning from Incidents. In *From Modeling to Managing Security: A System Dynamics Approach*. Kristiansand, Norway: Norwegian Academic Press, http://www.hoyskoleforlaget.no/hia035/.
Eden, Colin, and Fran Ackermann. 1998. *Making Strategy*. Thousand Oaks, CA: Sage Publication.
Eisenhardt, K. M. 1989. Agency theory: An assessment and review. *Academy of Management Review* 14:57-74.
Ellison, R. J. , D. A. Fisher, R. C. Linger, H. J. Lipson, T. A. Longstaff, and N. R. Mead. 1999. Survivable Network Systems: An Emerging Discipline. Pittsburgh, PA: Software Engineering Institute, Carnegie Mellon University.
Ellison, Robert J., and Andrew Moore. 2003. *Trustworthy Refinement Through Intrusion-Aware Design (TRIAD)*. CMU/SEI 2003 [cited September 10 2003]. Available from http://www.cert.org/archive/pdf/03tr002.pdf.
Emerson, Steven. 2002. *American Jihad: The Terrorists Living Among Us*. New York: The Free Press.
Flamholtz, E. 1979. Organizational control systems as a management tool. *California Management Review* 22:50-59.
Gaudin, Sharon. 2002. *Case Study of Insider Sabotage: The Tim Lloyd/Omega Case*. Computer Security Journal 2000 [cited 20 October 2002]. Available from http://www.gocsi.com/pdfs/insider.pdf.
Gonzalez, Jose J, ed. 2003. *From Modeling to Managing Security: A System Dynamics Approach*. Vol. 35, *Research Series*. Kristiansand, Norway: Norwegian Academic Press, http://www.hoyskoleforlaget.no/hia035/.
Gonzalez, Jose J, and Agata Sawicka. 2003. The Role of Learning and Risk Perception in Compliance. In *From Modeling to Managing Security: A System Dynamics Approach*, edited by J. J. Gonzalez. Kristiansand, Norway: Norwegian Academic Press, http://www.hoyskoleforlaget.no/hia035/.
Johnson, P. E., S. Grazioli, K. Jamal, and G. Berryman. 2001. Detecting Deception: Adversarial Problem Solving in a Low Base Rate World. *Cognitive Science* 25 (3):355-392.
Jones, G. R. 1984. Task visibility, free riding, and shirking: Explaining the effect of structure and technology on employee behavior. *Academy of Management Review* 9:684-695.
Knight, J., E.A. Strunk, and K.J. Sullivan. 2003. Towards a Rigorous Definition of Information System Survivability. Paper read at DISCEX 2003, at Washington, D.C.

Lewicki, R. J., and B. B. Bunker. 1996. Developing and maintaining trust in work relationships. In *Trust in Organizations: Frontiers of Theory and Research.*, edited by R. M. Kramer and T. R. Tyler. Thousand Oaks: Sage.
Lipson, H. F. *Tracking and Tracing Cyber-Attacks: Technical Challenges and Global Policy Issues* 2002 [cited March 27, 2004. Available from http://www.cert.org/archive/pdf/02sr009.pdf.
Lipson, H. F., and D. A. Fisher. 1999. Survivability - A new technical and business perspective on security. Proceedings of the 1999 New Security Paradigms Workshop, September 21-24, at Caledon Hill, Ontario.
Lipson, Howard F. 2003. Personal communication, October 21.
Luna-Reyes, Luis F. 2004. Collaboration, trust and knowledge sharing in information technology intensive projects in the public sector. Ph.D. thesis (in press), School of Information Science and Policy, University at Albany, State University of New York, Albany, NY.
Melara, Carlos, Jose Maria Sarriegui, Jose J Gonzalez, Agata Sawicka, and David L Cooke. 2003a. A system dynamics model of an insider attack on an information system. Proceedings of the 21st International Conference of the System Dynamics Society July 20-24., at New York, NY, USA.
———. 2003b. A system dynamics model of an insider attack on an information system. In *From Modeling to Managing Security: A System Dynamics Approach*, edited by J. J. Gonzalez. Kristiansand, Norway: Norwegian Academic Press, http://www.hoyskoleforlaget.no/hia035/.
Mintzberg, H.. 1973. *The nature of managerial work*. New York: Harper & Row.
Rempel, J. K, J.G. Holmes, and M.P. Zanna. 1985. Trust in close relationships. *Journal of Personality and Social Psychology* 49:95-112.
Richardson, George P., and David F. Andersen. 1995. Teamwork in group model building. *System Dynamics Review* 11 (2):113-138.
Schneier, Bruce. 2000. *Secrets and Lies: Digital Security in a Networked World*. New York: John Wiley & Sons, Inc.
Spitzner, Lance. 2003. *Honeypots: Tracking Hackers*. Boston: Addison-Wesley Publishing Company.
Stanton, J. M., and J. L Barnes-Farrell. 1996. Effects of computer monitoring on personal control, satisfaction and performance. *Journal of Applied Psychology* 81:738-745.
The Honeynet Project. 2004. *Know Your Enemy: Learning About Security Threats*. 2 ed. Boston: Addison-Wesley Publishing Company.
Tyler, T. R., and E. A. Lind. 1992. A relational model of authority in groups. In *Advances in Experimental Social Psychology*, edited by M. P. Zanna. San Diego: Academic Press.
Vennix, Jac A. M. 1996. *Group model building: Facilitating team learning using system dynamics*. Chichester: John Wiley & Sons.
Vennix, Jac A. M., Henk A. Akkermans, and Etienne A. Rouwette. 1996. Group model-building to facilitate organizational change: an exploratory study. *System Dynamics Review* 12 (1):39-58.
Whitner, E. M, S. E. Brodt, M. A. Korsgaard, and J. M. Werner. 1998. Managers as initiators of trust: An exchange relationship framework for understanding managerial trustworthy behavior. *Academy of management review* 23 (3):513-530.

Wiik, Johannes, Jose J Gonzalez, Howard F. Lipson, and Timothy J. Shimeall. 2004. Modeling the Lifecycle of Software-based Vulnerabilities. Proceedings of the 22nd International Conference of the System Dynamics Society July 20-24., at Oxford, UK.

Appendix I

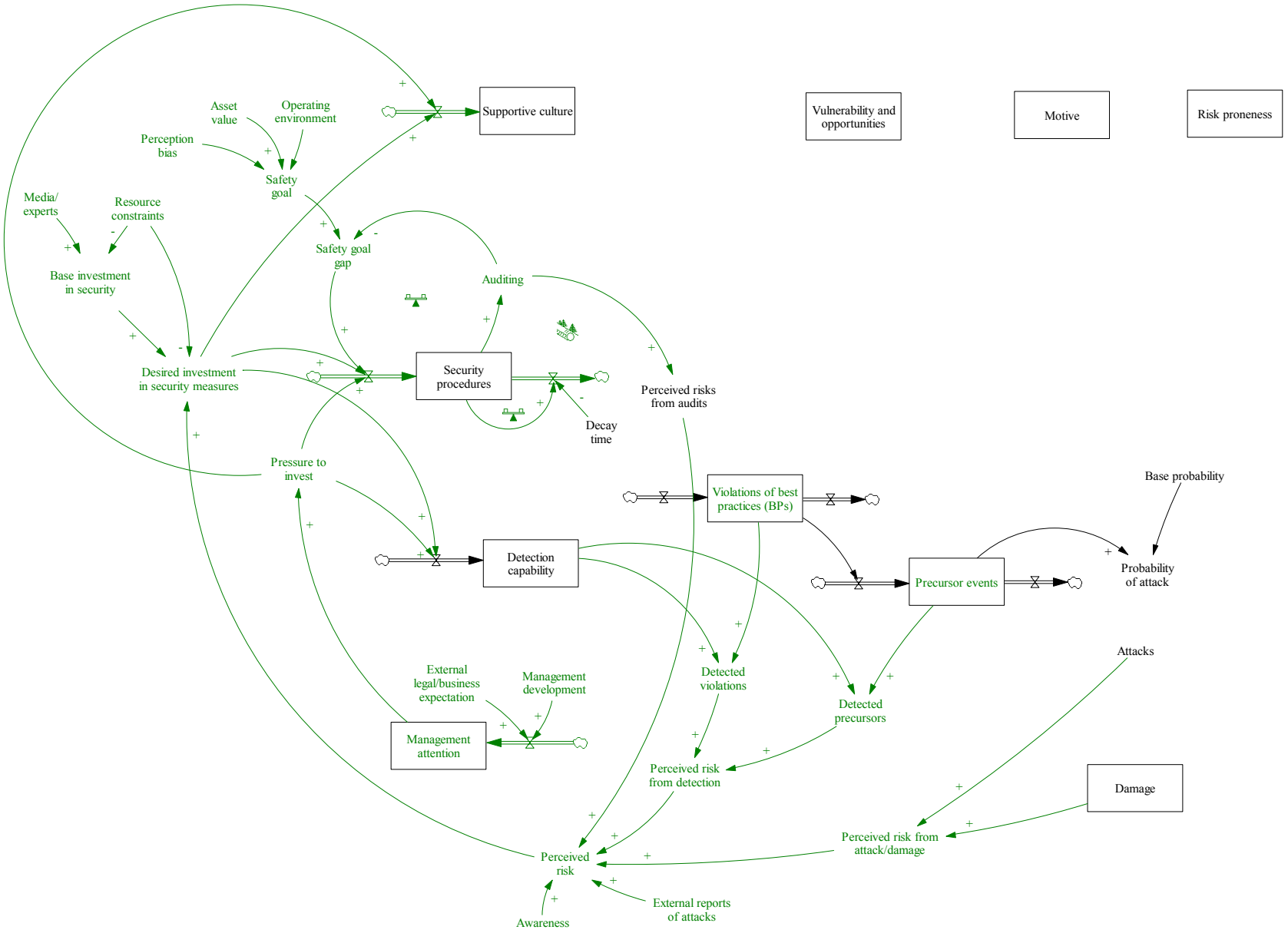

Figure I-1 Causal map of learning from experience, audits, and detection (Wednesday morning)

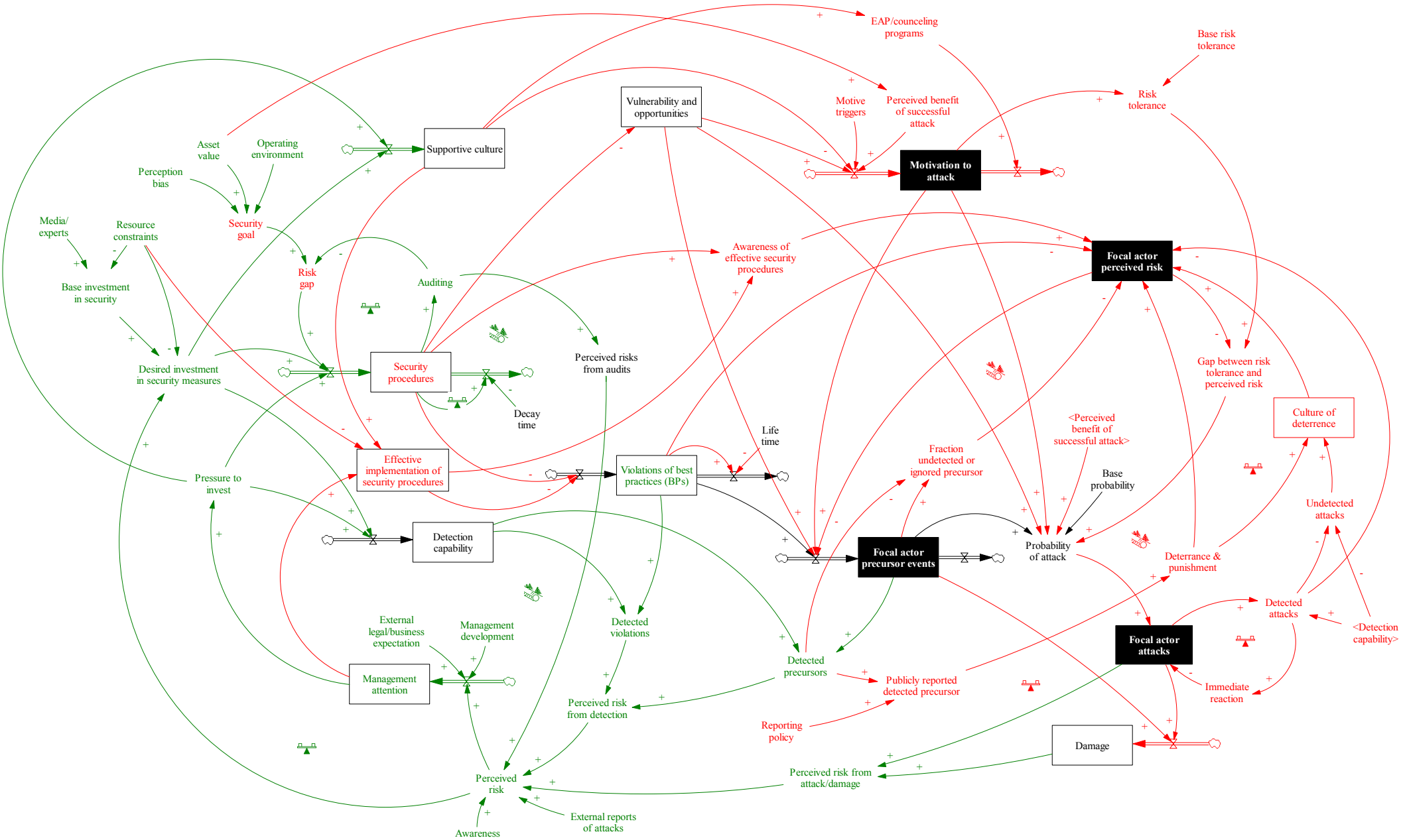

Figure I-2 Causal map of growth of motive (Wednesday afternoon)

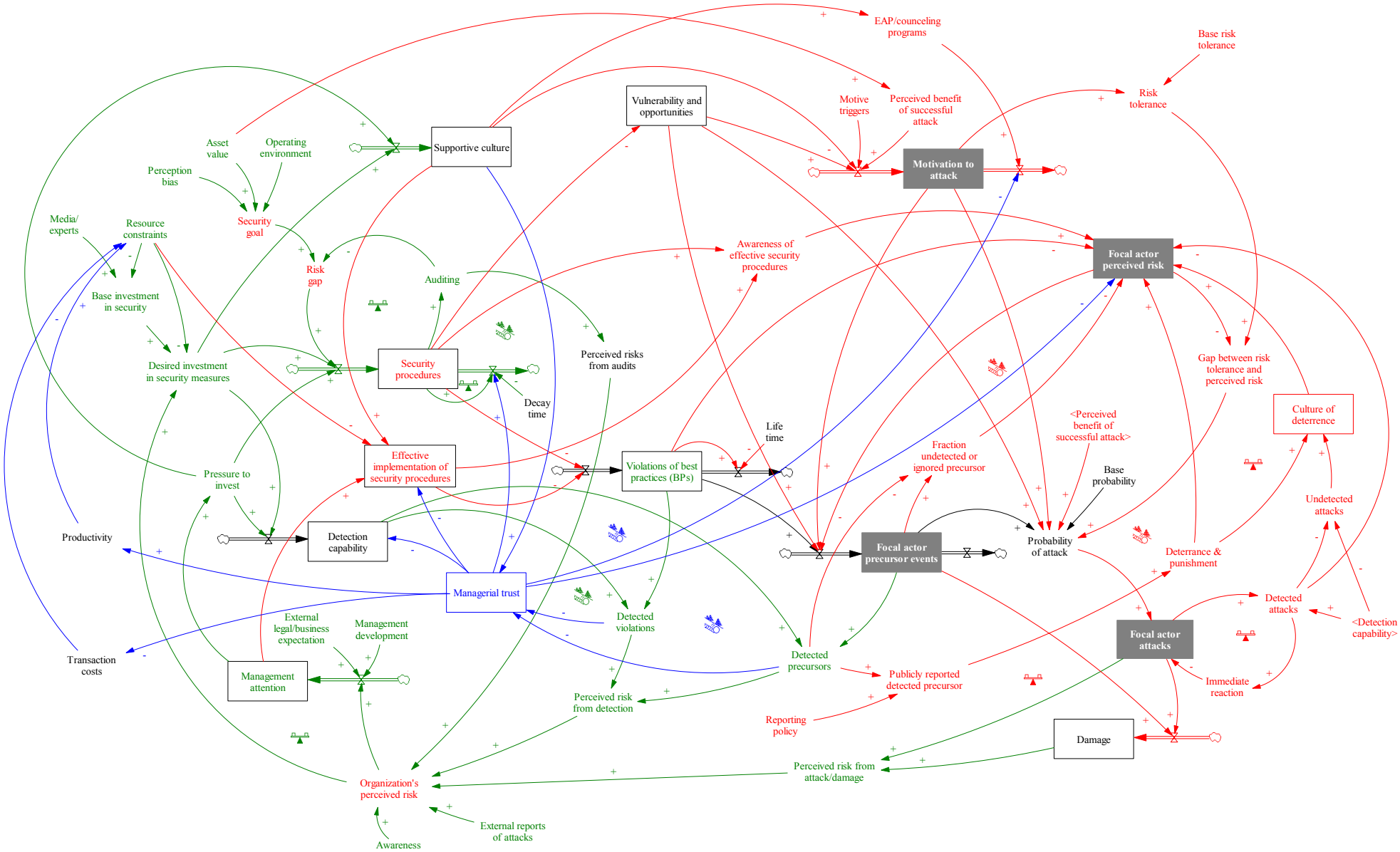

Figure I-3 Workshop's full causal map including trust and deterrence structures (Thursday morning)

Back to the Top